Whither teachers' assessment: trust in teachers' assessment since the Norwood Review of 1943

Michael Gordon Walker

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

The University of Leeds
School of Education

September 2020

The candidate confirms that the work submitted is his/her own and that appropriate credit has been given where reference has been made to the work of others.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Acknowledgements

Abstract

Educational assessment is pivotal to the process of teaching and learning, and has become a key element of 'high-stakes' school accountability measures in England. The role of teachers in 'high-stakes' assessments has changed over time in response to concerns over the efficacy of their judgements in what have become measures of their own performance. Using the work of Foucault as a theoretical framework, this study presents an analysis of these changes from the mid-twentieth century to the present day drawing on data collected before the Covid-19 pandemic. The study examines the perceptions of teachers, initial teacher educators and key-influencers on the levels of trust assigned to teachers' assessments and the degree to which initial teacher training (ITT) prepares teachers in the theory and practice of educational assessment.

Using a pragmatic mixed methods approach, academic and other documentary evidence are reviewed to develop an analysis of the role of teachers over time. Quantitative and qualitative data were generated through semi-structured interviews and questionnaire surveys with teachers, key-influencers and ITT providers. The findings show that the role of teachers in high-stakes assessments has fluctuated between positions of high trust to one of current marginalisation. The study suggests a power struggle over who controls the assessment system in that contemporaneous social, political and technical forces drive change rather than any defensible notion of ideal educational assessment practice. The study finds important discrepancies between the views of ITT providers and teachers around levels of expertise found amongst ITT providers and schools, but agreement from all stakeholders of the need to improve knowledge and practice in this field.

The study recommends the design of a national curriculum on educational assessment to support initial and on-going teacher education. Further research into workable standard setting and moderation procedures for teacher-based assessments is also recommended.

Table of Contents

List of Tables

List of Figures

Abbreviations

ACSL	Association of School and College Leaders
AERA	The American Educational Research Association
AfL	Assessment for Learning
APA	American Psychological Association
APP	Assessing Pupils' Progress
AQA	Assessment and Qualifications Alliance
AST	Advanced Skills Teacher
ATL	Association of Teachers and Lecturers
BBC	The British Broadcasting Company
BOS	The Bristol On-line Survey Tool
BTEC	Business and Technology Education Council
CCEA	Council for the Curriculum, Examinations and Assessment
CCT	Chartered College of Teaching
CEP	Centre for Economic Performance
CIEA	Chartered Institute of Educational Assessors
CPD	Continuous Professional Development
CPVE	Certificate in Pre-Vocational Education
CSE	Certificate of Secondary Education
DES	Department for Education and Science, England
DfE	Department for Education, England
DfEE	Department for Education and Employment, England
EB	English Baccalaureate
ESRC	Economic and Social Research Council
ETS	Educational Testing Service
FFT	Fischer Family Trust
GCE	General Certificate of Education
GCSE	General Certificate of Secondary Education
GNVQ	General National Vocational Qualification
GTP	Graduate Teaching Programme
HMC	Headmasters' and Headmistresses' Conference
HMCI	Her Majesty's Chief Inspector of Schools, England
IoD	Institute of Directors
ITE	Initial Teacher Education
ITT	Initial Teacher Training

JCQ	Joint Council for Qualifications
JUSCO	The Junior School Collaboration
KS1	Key Stage One
KS2	Key Stage Two
KS3	Key Stage Three
MAT	Multi-Academy Trust
MoE	Ministry of Education
NAA	National Assessment Agency
NAHT	National Association of Head Teachers
NASUWT	National Association of Headmasters Union of Women Teachers
NCME	National Council on Measurement in Education
NCVQ	National Council for Vocational Qualifications
NEAB	Northern Examinations and Assessment Board
NEU	National Education Union
NFER	National Foundation for Educational Research
NCC	National Curriculum Council
NCSL	National College for School Leadership
NUT	National Union of Teachers
NVQ	National Vocational Qualification
OECD	Organisation for Economic Co-operation and Development
Ofqual	The Office of Qualifications and Examinations Regulation
Ofsted	Office for Standards in Education, Children's Services and Skills
ONS	Office of National Statistics
PGCE	Postgraduate Certificate in Education
PIRLS	Progress in International Reading Literacy Study
PISA	Programme for International Student Assessment
Progress 8	a measure of progress a pupil makes from the end of primary school to the end of secondary school used in accountability measures
QCA	Qualifications and Curriculum Authority
QCDA	Qualifications and Curriculum Development Agency
QTS	Qualified Teacher Status
RSA	Royal Society for the encouragement of Arts, Manufactures and Commerce
SATS	Standardised Assessment Tasks
SBA	School Based Assessment
SCITT	School Centred Initial Teachers Training
SEAC	Secondary Examinations and Assessment Council
SEC	Secondary Examinations Council

SLE	Specialist Leader of Education
SSEC	Secondary Schools Examination Council
SPSS	IBM Statistical Package for the Social Sciences
STA	Standards and Testing Agency
TALLIS	The OECD Teaching and Learning International Survey
TDA	Training and Development Agency for Schools
TGAT	Task Group on Assessment and Testing
TIMMS	Trends in International Mathematics and Science Study
T Levels	T Levels are new courses which follow GCSEs and are equivalent to 3 A levels.
TSC	Teaching Schools Council
UCET	Universities' Council for the Education of Teachers
UKRLP	United Kingdom Register of Learning Providers
UN	United Nations
UNESCO	The United Nations Educational, Scientific and Cultural Organisation
VET	Vocational and Educational Training
11+	An examination for selection purposes taken by children aged eleven years in some local education authorities in England. Also expressed as 'Eleven plus'.

Chapter 1: Introduction to this research project

1.1 Researcher background

At the end of my first term as a teacher, I was asked by my head of department to provide four questions as my contribution to the end of term test for a group of pupils in what is now described as 'Year 10'. I had no idea how to go about this, but I was advised by a colleague to select questions from previous GCE O level papers. No one questioned or commented on my selection and it was only when I began to mark the responses that as all my pupils gave incorrect but consistent answers that I realised two things. 1, in my preparation to become a teacher, no-one mentioned assessment: and 2, the wrong but consistent answers submitted by my students told me as much about my teaching as it did about their capacity to learn. It was only two terms after this that a student in a mathematics examination I was invigilating engraved his answers into his desktop in response to a particular question that instructed candidates that when answering questions on logarithms, they should use the tables provided. From that moment on, my interest in educational assessment was formed.

At this seminal point now some 47 years ago, I realised the pivotal role of assessment in the teaching and learning process and the importance of assessment design – be that in the classroom or for the purpose of qualifications. At that time, teachers played a key role in the development of examinations for the Certificate of Secondar Education (CSE) and it wasn't long before I became a chief examiner and moderator before being nominated as a 'regional expert' to support the introduction of the GCSE in 1986. Shortly after this, I took up a post at the National Curriculum Council (NCC) to assist in the development of the National Curriculum introduced by the 1988 Education Reform Act and the associated regime of standardised tasks and tests – one of many changes to educational assessment made during my time in education. Over subsequent years, I worked in various reincarnations of the NCC charged with monitoring the National Curriculum and the English qualification system and providing advice to ministers from differing political persuasions. I eventually retired from my post as Executive Director of Education at the Qualifications and Curriculum Development Agency (QCDA) during which time I held responsibility for monitoring the performance of examination boards offering general and vocational qualifications, and between 2008 and 2011, accountability for National Curriculum assessments.

So why is my background of relevance to the reader of this project thesis? Even from a cursory over-view of my background in education, it becomes clear that I have many years of experience in a range of roles in the English system of education ranging from the practical experiences gained as a teacher, through to the political dimensions of a state-run system gained through years as a policy adviser and system administrator. This is a position that permits access to and knowledge of the everyday practice of teachers and those having influence on national policy. I believe it is therefore of importance that the reader is aware of my positionality in relation to this study.

1.1.1 Researcher positionality

In order to take an ethical approach, this study has to acknowledge the position of the researcher in terms of neutrality, ideology or political stance (Wilson, 2017); this can be defined as researcher positionality:

> Positionality refers to the stance or positioning of the researcher in relation to the social and political context of the study – the community, the organisation or the participant group (Rowe, 2014. p. 628).

According to Holmes (2020), positionality impacts on the; "…views, values and beliefs about the research design, conduct and output(s)" (p. 2) of the research project: these considerations have a bearing on my philosophical position or 'world view' which is discussed in Section 3.2. Therefore, my experience as a teacher, an education policy adviser, and in developing and over-seeing National Curriculum assessments and the regulation of awarding bodies may have a bearing on my philosophical position or 'world view'. My various roles in the English education system have also given me privileged access to some of the key actors in the English education system, for example policy makers, regulators and teacher professional associations. And although all of these contacts have been at a professional level, it has enabled me to include such people in my research along with practicing teachers who volunteered their participation following their responses to the questionnaire surveys (Chapters 7 and 8). However, as a researcher, my aim is to provide objective and value free analysis around the research questions (Chapter 3) through drawing on the views of experts and practitioners in the field and giving preference to their perspectives so as to avoid as far as possible any personal bias. But these various relationships with research participants generate consideration of

what is described in the literature as the 'insider' and 'outsider' researcher. This is addressed in the following section.

1.1.2 The Insider – outsider researcher

In simplistic terms, an insider is a researcher who engages in topics; "…which are related to him/her or on topics related to a group he/she is associated with" (Saidin and Yaacob, 2016. Accessed online) whereas an outsider engages in research in topics outside of their experience or background (ibid). Clearly, based on my personal background, I am aligned with the position of insider as described by Saidin and Yaacob. As such, my insider knowledge provides access to key players and also provides a level of personal expertise in the field to help understand and question the views of others. It is also suggested that participants may therefore be more willing to share and provide richer data (Berger, 2013, cited in Woods, 2019, accessed online). Conversely, being an insider raises questions over the level of researcher objectivity with the possibility of selecting; "…participants who are most like them" (Woods, 2019). Both positions of insider and outsider have their relative advantages and disadvantages and to some degree, and any concerns over my insider status may be minimised by the fact that I retired six years before starting this research project reducing some of the more immediate familiarity together with selecting participants for interview who would present a range of academic, political and practitioner perspectives.

But throughout the process, I have used my knowledge and experience to draw on and understand the views of others whilst at the same time being aware of how my own bias and preconceptions may influence that understanding (Maykut and Morehouse, 1994). My intention throughout this project therefore has been to enhance; "… the chances of findings being valid…by a judicious combination of involvement and estrangement" (Hammersley, 1993, p. 219 cited in Hellawell, 2006).   Hellawell (ibid) elaborates this position further:

> I would contend that ideally the researcher should be both inside *and* outside the perception of the 'researcher'. That is to say that, as Hammersley (1993) implied, …both empathy *and* alienation are useful qualities for a researcher (original emphasis in italics, p. 487).

1.1.3 Researcher value position

In more recent times, I have held the position of Vice-chair of the Chartered Institute of Educational Assessors (CIEA), an organisation with the stated goal of improving the quality of educational assessment wherever it takes place free from any political or ideological allegiance. So as a value position, I hold the view that teachers should be more trusted as a profession when it comes to making assessments of their pupils. Equally I am of the view that in general, teachers are not currently well equipped in a technical sense to make valid and reliable assessments and that the focus on external summative assessment as a key accountability measure has presented perverse incentives, for example teaching to the test or gaming the system, that raise questions of trust in their veracity. I therefore believe that teachers have to demonstrate their professional ability to make fair and valid assessments.

1.2 Whither teachers' assessment

Having spent over forty-five years as a teacher, regulator and government adviser in the English education sector, I have seen how the role of teachers in what have become 'high-stakes' assessments has changed over time. The term 'high-stakes' used throughout this study can be in part defined as:

> Assessments with important consequences for test takers, on the basis of their performance. Passing has important benefits, such as progressing to a higher grade, a high school diploma, a scholarship, entrance to the labor (sic) market or getting a license to practice a profession. Failing also has consequences, such as being forced to take a remedial class or not being able to practice a profession. Examples of high-stakes tests include college entrance examinations, high/secondary school exit examinations, and professional licensing examinations (UNESCO Institute for Statistics, 2019).

However, since the outcomes of such assessments were utilised as measures of school performance in the early 1990s, the outcomes of tests and examinations have also held consequences for teachers, schools and politicians: hence this impact is also included in the term 'high-stakes'.

It would appear to me that changes to the role played by teachers in high-stakes assessments has more often been driven by political dogma and short term fixes to

broader issues in the education system rather than steps towards an ideal model of educational assessment. As much of my time was spent regulating and monitoring the work of awarding bodies and for a period as accountable officer for National Curriculum assessments, I have witnessed first hand how education policy in England has been influenced by the debates over the benefits and drawbacks of using teacher based assessment judgments in an accountability system in which the outcomes of tests and public examinations have become a central measure of the performance of the education system. Part of the debate has centred on the technical issues that would impact any assessment regime primarily around the concepts of validity and reliability. For the purpose of this study, validity is viewed as the degree to which the inferences drawn from the outcomes of an assessment can be justified; and for reliability, the degree of consistency found in repeated assessments (see for example AERA, 1999; Opposs and He, 2011; Baird et al., 2011; Shaw and Crisp, 2011; Isaacs et al., 2013; Popham, 2010). These two central concepts are widely discussed and debated in the academic literature and are often the focus of more detailed considerations such as the reliability of marking, conscious and un-conscious bias and the construct validity of assessment instruments (see for example Wiliam, 2001; Harlen, 2005a; Meadows and Billington, 2005; Jellis, 2018). The concepts of validity and reliability are discussed further in Chapter 2. But other broader considerations have entered the "…everyday discussion among citizens, educational professionals and politicians" (Pellegrino, 1999, p. 5) and have equally influenced the debate and ultimately government policy on the format and use of educational assessments. These include the efficacy of educational assessment outcomes as accountability measures, perceptions of trust and the capability of teachers in assessment matters, and the power relations between educators, administrators and politicians (Foucault, 1976; Major, 1991; Hargreaves, 1996; Kendall & Wickham, 2003; NAHT, 2014: Lightman, 2015).

This study draws on the extensive body of literature on educational assessment with claims and counter claims over the benefits and disadvantages of the inclusion of teacher based assessments in high-stakes assessments. Yet as this study will show, relatively little is known about the reliability of teachers' assessments and the impact of large-scale moderation procedures such as those used in England for National Curriculum assessments and GCSE examinations. Further, and despite recent reports (Carter, 2015; McIntosh, 2015) citing deficiencies around teacher and teacher educator's assessment knowledge, detail around the content and perceptions of quality of the theory and practice of assessment in teacher preparation courses in England is also sparse. However, as this research study will show, the inclusion of and trust in teachers' judgments in high-stakes

assessment has fluctuated but largely diminished over time (see Chapters 2 and 3) and as the Foucaultian analysis in Chapter 5 demonstrates, control of tests and examinations brings with it considerable power over the direction of education policy. Yet as Chapter 5 illustrates, teacher and teacher associations have focused their concerns on the negative impact of school performance tables on pupil and teachers' wellbeing and teacher workload rather than on the development of a workable regime of educational assessment in which they exert at least some level of control. Further, there is little exploration of the views held by teachers and those holding positions of influence on educational policy in the English education system on the purpose and limitations of educational assessment and the extent to which initial teacher education (ITE) prepares teachers in this particular field of expertise.[1]

1.2 The focus of this study

Working within the context of the education system in England, this study examines how the role played by teachers in high-stakes assessments has changed since the mid-twentieth century to the present day. The publication of the Norwood Report in 1943 has been selected as the datum point for this research on the grounds that in respect of the role played by teachers in high-stakes assessments, Norwood had a significant influence on the 1944 Education Act, an Act that "…provided the framework for the post-war education service, which indeed remained intact for almost fifty years" (Barber, 1944. p. ix) and a vision for the role of teachers in public examinations and tests that were to become a feature of the 1950s and beyond. In considering such change, I also seek to understand the perceptions of educational assessment amongst practicing teachers and individuals in positions of influence, and together with teacher educators, the extent to which the knowledge and understanding of assessment theory and practice found in initial teacher education courses prepares teachers for entry into the teaching profession. The aim of this research study is to provide greater insight into the utilisation of teachers' assessments over time and the implications for assessment validity and reliability in high-stakes assessments. This is presented against the back-drop of a decision taken in March 2020 by the Secretary of State for Education (Williamson, 2020) to cancel all public examinations in England in summer 2020 and issue statistically standardised results based on teacher generated grades and rank-ordering of students; this approach is unprecedented in the English education system. It is hoped that this research will

---

[1] In terms of terminology, the use of ITE is preferred to that of initial teacher training (ITT), but the latter is more generally used as an alternative and is used by the Department of Education in England in regards to routes into teaching; both terms are used throughout this thesis.

contribute to the post-Covid-19 pandemic review of the outcomes of this decision. The impact of the pandemic on test and examination arrangements for 2020 is discussed in Section 1.8.

1.3 Overview of this chapter

Having introduced the area of research interest and focus of this study, the chapter goes on to briefly explain the position of educational assessment in the English education system over the period of interest. This is followed by sections that examine the role played by teachers in what have become a high-stakes tests and examinations and concerns over the reliability of teachers' assessments in a system of accountability in which such outcomes are used as central indicators of performance. Further sections look at the attempts made to increase the reliability of teacher based assessments amid concerns over the levels of assessment expertise found in schools and within initial teacher training (ITT) institutions. The chapter next presents the rationale and intended contribution of this research with reference to the Covid-19 pandemic. The key research questions and methodological approach are then presented before the chapter concludes with a summary and an over-view of the structure of this thesis.

1.4 Educational assessment in England

If students learned everything they were taught, we would have no need for educational assessment; instead, we would just need a record of what we had taught (Wiliam, 2013): as anyone involved in education will probably testify, students do not learn everything they are taught. As a consequence, educational assessment in a wide range of forms has long been accepted as an integral element in the teaching and learning process, be it from the perspective of the learner, the teacher or external users of assessment outcomes such as employers (QCDA, 2011, Wong & Kaur, 2015; CCEA, 2014; Wiliam, 2013; Shaffner, 2008). However, the uses to which educational assessment outcomes are put have increased over time to support; "…many different kinds of decision, process or action" (Newton, 2007, p. 3) ranging from the classroom-based focus of informing teaching and learning to purposes of selection, measuring progress, certification, and system monitoring.[2] In the simplest of terms the range of uses of educational assessment identified by Newton (ibid), can be classified into three main groups:

---

[2] Newton (Ibid) listed twenty-two different purposes in what he describes as a non-exhaustive list "… to provide a sense of the number and range of different kinds of decision, process or action that can be identified" (p. 3).

- as a tool to inform teaching and learning;
- as a form of certification or accreditation: or
- as a tool for accountability.

Of these three broad classifications, the use of assessment outcomes as a key measure of accountability is a relatively recent phenomenon in the history of the English education system and one that has generated discussion, debate and contention within the educational establishment and more widely in the public and political sphere (see for example, Astle, 2017; Pellegrino, 1999).

1.4.1 Teachers' assessments in high-stakes tests and examinations

Since the mid-1950s there has been a rapid expansion in the numbers of students taking high-stakes assessments, during which time the role played by teachers has included the development of locally determined school leavers certificates, regional and national Mode 3 examinations in which syllabuses and the associated assessments were derived wholly at school level, and other school based assessments such as coursework and later controlled assessments used in conjunction with externally set and marked tests. Moreover, up to 2015, teachers were required to submit estimated, or forecast grades to awarding bodies (Gill, 2019) to provide a source of data to assist in the process of awarding final grades and in cases where a candidate missed an examination or requested special consideration, for example due to illness during an examination. In research conducted by Gill and Benton (2015a) for Cambridge Assessment, the accuracy of the estimated grades submitted by schools for OCR GCSE subjects showed that overall, the forecast grade was 'correct' (that is in line with achieved or awarded grade) 43.5% of the time and either correct or within one grade 87.2% of the time (see Figure 1). The research findings also showed fluctuation across subjects and by different school types and concluded that: "… it appears much easier for teachers to accurately forecast grades of C or higher than grades below C" (ibid., p. 4). Since 2016, awarding bodies have not referred to estimated grades, even in cases relating to special consideration: "This is to ensure that all students are assessed according to the same criteria and that the standard of the exam is upheld when awards are made" (AQA, 2020). And the more recent changes to the assessment format used in GCSE, GCE A level examinations and National Curriculum assessments have seen a further reduction in the use of centre-based assessments. Estimated grades for GCE A levels do continue to be used for determining offers for university courses in advance of the publication of actual qualification awards released in late summer and close

to the beginning of university terms. However, Gill and Benton (2015b) also conducted research on estimated grades for OCR GCE A levels and found similar patterns to those shown for the GCSE (see Figure 2).

Figure 1.1: Extract: The accuracy of forecast grades for OCR GCSEs in July 2014
  (Gill and Benton, 2015a; Cambridge Assessment)

Table 1. Overall accuracy of forecast grades

%Accurate	43.56
%Optimistic	42.42
%Pessimistic	14.02

Table 2. Extent of inaccuracy of forecast grades

	%	N
Within 1 grade	87.29	599328
More than 1 grade out	12.71	87268

Table 3: Accuracy of forecast grades by final grade

Final Grade	%Accurate	%Optimistic	%Pessimistic	N
A*	65.52	-	34.48	54844
A	52.08	21.65	26.27	111857
B	49.42	32.25	18.34	150419
C	53.09	39.94	6.97	170059
D	24.28	71.06	4.67	103459
E	18.12	77.26	4.62	49637
F	13.80	81.75	4.46	25719
G	13.79	84.13	2.07	13022
U	11.08	88.92	-	7580
Total	43.56	42.42	14.02	686596

Figure 1.2: The accuracy of forecast grades for OCR GCE A Levels in July 2014
     (Gill and Benton, 2015b; Cambridge Assessment).

Table 1. Overall accuracy of forecast grades

% Accurate	43.14
% Optimistic	43.17
% Pessimistic	13.69

Table 2: Extent of inaccuracy of forecast grades

	%	N
Within 1 grade	88.10	149152
More than 1 grade out	11.90	20143

Table 4: Accuracy of forecast grades by final grade

Final Grade	% Accurate	% Optimistic	% Pessimistic	N	
A*	60.28	-	39.72	16459	
A	58.13	22.81	19.06	32495	
B	47.72	37.73	14.55	42739	
C	40.98	50.95	8.08	37953	
D	24.83	70.76	4.41	24582	
E	17.02	82.31	0.67	11701	
U	5.35	94.65	-	3366	
Total	43.14	43.17	13.69	169295	

It should be noted however that whilst there is a growing body of research on the *accuracy* of teachers' estimated grades, some of which dates back to 1912 (see Starch & Elliot), little is known about *how* teachers go about generating estimated grades (Gill, 2019). Unpublished research by Child & Wilson (2015, cited in Gill, ibid) found that teachers used a range of sources on which to base their judgments with 94% of teachers agreeing that AS level qualification grades were the best predictors of A level grades: "Other commonly used sources of information included observations of the quality of work or of student

commitment and performance in coursework and mock exams" (Gill, op cit., p. 34). However, the Universities and Colleges Admissions Service (UCAS) through which estimated grades are submitted, advise teachers that predicted grades should be: "…aspirational but achievable – stretching predicted grades are motivational for students, unattainable predicted grades are not" (UCAS, 2020). And although UCAS encourages schools and colleges to provide staff training to promote consistency within and across departments, there is no body of work identifying the extent to which schools and colleges adopt such an approach although research by QCA (2006) found that the process of standardisation in schools for GCSE coursework was not common. Further, UCAS (op. cit.) note that teachers may come under pressure from senior leaders in schools to 'inflate' or 'supress' estimated grades. As a part of the arrangements utilised for the awarding of grades in the 2020 pandemic, the Office of Qualifications and Examinations Regulation (Ofqual) has established a project to gain an understanding into how schools determine predicted grades: however, the outcome of this research will not be known for some time. Concerns have also been raised over weaknesses in initial teacher education courses and "…significant gaps in both the capacity of schools and ITT providers in the theoretical and technical aspects of assessment" (Carter, 2015. p. 9. See also McIntosh, 2015). Taken alongside issues of trust in teachers' assessments within the profession in a high-stakes environment (NAHT, 2014), questions over the reliability of teacher-based assessments will persist.

1.4.2 The system of educational accountability in England

The terms 'National Curriculum assessments' or 'tests' and 'general qualifications' are used throughout this study to denote statutory tests used in the primary phase of education and qualifications used in schools and colleges for students up to the age of 18 and recognised and used in performance tables compiled and published by the Department for Education (DfE) and regulated by Ofqual. Since the early 1990s, the outcomes of these assessments have been used as the primary data for the production and publication of school performance tables. Since their inception, performance tables have added to the uses to which assessment outcomes are used, moving the focus beyond the test taker to the institutions in which learners are prepared for such assessments. At this point in time, teachers' assessments played a key role in National Curriculum assessments and in general qualifications. However, the utilisation of these assessments used within measures of accountability presented perverse incentives for schools and teachers; this is discussed further in Section 2.9 of Chapter 2. These unintended consequences led to concern over the reliability of teachers assessments and arguably are the root of the

increased marginalisation of teachers' judgments used as elements of high-stakes assessments:

> …as Daugherty (1995) and Black (1996) have demonstrated, there appears to have been a conscious policy throughout the 1990s to remove teachers from assessment in both the GCSE and the National Curriculum. If one accepts the principle of increasing control through greater centralisation then such a policy makes perfect sense, though it should be pointed out that unions are often keen to reduce the burden placed on teachers by this marking, which is usually unpaid (Lambert and Lines, 2001. p. 77).

The first school performance tables were published in 1992 and were based on secondary phase qualifications with National Curriculum assessment added in 1996 covering the primary and secondary phases of education. As a result, the participation of teachers in generating assessments on which their performance and that of their schools is judged has raised questions of reliability and trust on the grounds that they have conflicted interests and therefore there is need for externally set and marked assessments:

> The logic of this argument is important because it continues to drive policy to this day, despite the change of government; but of course its conclusion rests on a significant value judgement about teachers; one that in our view is deeply flawed (Lambert & Lines, 2001. p. 69).

The use of externally set and marked assessments does not of course in itself prevent acts of malpractice that undermine the involvement of teachers in the test and examination system in England. Malpractice includes the actions of school administrators, teachers and students and although acts by teachers such as over-aiding pupils during tests and examinations are low in comparison to the population of teachers involved each year in high-stakes assessments, "…any malpractice damages the reputation of, and confidence in, the UK's qualification system" (JCQ, 2019. P 10). In recent years, such cases have attracted the attention of the mass media[3] and the number of investigated cases of maladministration for National Curriculum assessments has risen from 524 in 2016 to 793 in 2018 (STA, 2019). However, 48% of cases (n=383) were reported by schools themselves and of the overall cases reported, 32 referred to 'over-aiding' in Key Stage 1:

---

[3] See for example: The Independent on-line (August, 2017); The Telegraph on-line (August, 2017); The Times (February, 2018 & November 2018: Schools Week on-line, (2018).

"The most common allegation type at KS2 in 2018 concerned test administrators over-aiding pupils during the tests. In 2017, 42.3 % of allegations were of this type, compared to 35% in 2018" (Ibid., p. 10). As the Independent Commission on Malpractice (JCQ, 2019) noted, incidents of reported malpractice do however remain 'extremely low' (JCQ, op cit., p. 10) and teacher involvement is in effect a small sub-set of the overall picture.

The literature review presented in Chapter 2 demonstrates that in broad terms, there is a tension between those who argue that the teaching profession is best placed to assess pupils (for example, Bew, 2011a&b: Rimfeld et al., 2019: ATL et al., no date) and others (for example Wolf, 2008) who argue that external assessment is essential if a true measure of the performance of pupils and schools is to be achieved – a view supported by some politicians (for example, Gibb, 2017; Williamson, 2020). A third viewpoint is the development of a coherent and complementary approach using a mixture of internal and external measures to strike a level of balance and improved validity (European Commission, 2018).

1.4.3 Attempts to increase the reliability of teacher based assessments

In an attempt to increase the reliability of teacher assessment in high-stakes assessments in England, successive regulatory bodies such as the SCAA, QCA, QCDA and Ofqual have required examination organisations and in the case of National Curriculum tests, the Standards and Testing Agency and Local Authorities, to develop and implement elaborate systems of moderation for example as currently used for Key Stage 2 assessments. Moderation systems used in England rely on internally moderated marks (that is within a single school) being subjected to the scrutiny of an external assessor who judges the accuracy of the internally generated marks against standards set by awarding organisations. However, the ability of numerous external moderators to maintain the standard set by the awarding organisation has been questioned (Taylor, 1992). One approach to negate claimed deficiencies in teacher assessment is to remove teachers from the assessment process entirely (HMCI, 2013). However, a model based on externally administered written tests places limits on what can be assessed leaving what might be considered as important aspects of learning outside of the assessment process unless teachers are involved as primary assessors in the process (see Stanley et al., 2009). This is why over many years coursework has contributed to the overall assessment framework of most subjects to cover broader elements such as practical work in design

and technology, performance in drama or experimental work in science:

> [...] teachers can sample the range of a pupil's work more fully than can any assessment instruments devised by an agency external to the school. This enhances both reliability (because it provides more evidence than is available through externally devised assessment instruments) and validity (it provides a wider range of evidence) (Mansell et al., 2009, p. 12, cited in Baird et al., p. 44).

But simultaneously, coursework has been the subject of much criticism and concern, in particular around the reliability of teacher assessments be it through conscious or un-conscious bias – for example in favouring particular candidates or groups of pupils, and opportunities for plagiarism or teacher direction and parental support (see for example Colwill, 2007, p. 11). In more recent times, examinations have been re-structured to reduce or replace the more school influenced un-structured coursework with controlled assessments determined by awarding bodies. However as noted by Johnson (2011 & 2012), teachers have advanced notice of controlled assessments and are open to the temptation to assist their pupils and along with the effects of bias, lead to differences in standards of judgments from class to class and school to school. In short, controlled assignments have not prevented concerns over the integrity of teacher-based assessments. And in countries with high-stakes teacher assessments, for example the member states of the Organisation for Economic Co-operation Development (OECD) Programme for International Student Assessment's (PISA) programme, there is little empirical study on the reliability of different tasks, the influence of the teacher or how teachers arrive at their judgments about student attainment (Johnson, 2012; Ofqual, 2013). Further, there is little empirical study to reveal the nature, extent or quality of internal standardisation in schools and colleges - although QCA (2006) concluded that internal standardisation in schools for GCSE assessment was not widespread. Further research conducted by Ofqual (2011) concluded that there was an urgent need for empirical work on the reliability of teacher assessment:

> It is interesting, if somewhat depressing, to note that in several important reviews of the literature on teacher assessment reliability, as in this one, one of the principal findings has been the sparseness of the literature on the reliability of teacher assessment, and in particular of reports offering quantified evidence one way or the other (Wilmut et al., 1996; Harlen 2004; Johnson 2006; Stanley et al., 2009; Harth & van Rijn 2011): (cited in Ofqual, 2011, p. 44).

On the basis of such concern, Michael Gove (2014a), then Secretary of State for Education, announced changes to the role of coursework in GCE and GCSE examinations. The stated goal, as reported by Ofqual (2014) was to reduce the time spent on in-class assessments and reduce malpractice.

However, external paper based examinations have their own limitations ranging from technical issues, for example construct validity where aspects such as practical skills are hard to assess through a written paper, through to negative impacts on teaching such as reducing curriculum content in efforts to 'teach to the test' (Popham, 2001; Vaughan, 2015). Further, there is a lack of confidence in the teaching profession regarding the quality of external marking and even where markers are governed by tight marking schemes, there is much evidence to show variance between markers for the same piece of work (QCA, 2009; Meadows and Billington, 2005). This has been further reinforced by more recent work from Ofqual (2018) on the probability of receiving the 'definitive'[4] grade at qualification level:

> As might be expected there are some clearly identifiable subject patterns. The probability of receiving the 'definitive' qualification grade varies by qualification and subject, from 0.96 (a mathematics qualification) to 0.52 (an English language and literature qualification). The probability of receiving the definitive grade or adjacent grade is above 0.95 for all qualifications, with many at or very close to 1.0 (i.e. suggesting that 100% of candidates receive the definitive or adjacent grade in these qualifications) (p. 4)

The Ofqual report has in turn generated further concern amongst the teaching profession with for example, the Executive Director of the Headmasters Conference (HMC, 2018) stating:

> It is extremely worrying that around a quarter of exam grades are not reliable. This directly affects hundreds of thousands of young people every year…The implications are grave, as a questionable grade can have a significant effect on a

---

[4] The term 'definitive' (eg 'definitive' mark or 'definitive' grade) is based on terminology ordinarily used in exam boards for the mark given by the senior examiners at item level for each seeding response.

pupil's life chances in a high-stakes exam-focused environment (HMC, press release 1st December 2018).

The reported effects on pupil's life chances manifest in the growing number of requests for re-marks in National Curriculum tests, GCSE and GCE examinations and more general concerns with the quality of examinations as cited in *England's 'examinations industry': deterioration and decay* (HMC, September 2012). From this, it would appear that teachers and schools are to some degree 'anti-testing.' However, in 2010, over 6.38 million non-statutory optional tests covering years 3, 4 and 5 alone were ordered by schools. This was the year of a national boycott of statutory tests that was supported by over 26% of key stage 2 schools in England.[5] Even so, the number of tests purchased suggests that in reality, schools have issues beyond the tests themselves.

1.5 The culture of performativity

A now common use of test and examination outcomes as noted above is that of measuring school performance, a manifestation of the growth of what has been described as a culture of performativity (Ball, 2003). Indeed, some would argue examinations are the key measure in what has become described as an 'audit explosion' leading to; "…a decrease in trust and an increasingly defensive (teaching) profession" (Jenkins quoted in Stobart, 2008, p. 135). This has led to calls from the teacher associations for the scrapping of tests and the reinstatement of teacher assessment at key stage 2 and calls for the use of teacher assessments across all high-stakes assessments (Rimfeld et al., 2019). Indeed there are reportedly potential benefits to such a regime including increased validity, a richer evidence base and increased respect of the teaching profession (Johnson, 2013: pp. 91-92; Rimfeld et al., op. cit.). Wiliam, (2001) notes the limitations of relatively short tests with superficial sampling across the subject domain as used in National Curriculum assessments and GCSEs and consequently supports the use of multiple teacher assessments over time as the key to removing bias to improve reliability:

> The experience of GCSE has shown that the danger of bias in teacher assessments can be adequately addressed through standardisation and moderation. By using teacher assessment, we would in effect, be using assessments conducted over tens, if not hundreds of hours for each student,

---

[5] Figures from the Qualifications and Curriculum Development Agency.

producing a degree of reliability that has never been achieved in any system of timed written examinations (Wiliam, 2001, p19).

But others do not reflect Wiliam's confidence in the benefits of moderation, for example in an Ofqual blog (Black, 2019), *11 things we know about marking and 2 things we don't …yet*, one of the two 'unknowns' stated:

> Which has better consistency – marking or moderation? . We are currently conducting a series of studies to help us understand how moderation looks in relation to marking in terms of consistency (accessed on-line, 12th March 2019).

It is also worthy of note that developing standardisation and moderation systems as envisaged by Wiliam (2001) have financial consequences. A report from PwC commissioned by the QCA (2003) estimated that the total annual cost of administering National Curriculum assessments and general qualifications was in the region of £610million in 2003, and by 2008 this figure was thought to be conservative by the House of Commons Children, School and Families Committee report on assessment and testing (House of Commons Children, Schools and Families Committee. 2008); see Figure 1.3 below.

However, Wiliam (ibid., p. 21) does acknowledge the tension between the two concepts of validity and reliability (see also Jellis, 2018; Baird et al., 2011). Validity and reliability are highly dependent to the extent; "….that increasing one tends to decrease the other" (Harlen, 2005b, p. 213), though some, like Harlen prioritise validity; "…since a main reason for using teachers' assessment rather than depending entirely on tests for external summative assessment is to increase the construct validity of the assessment" (Ibid., p. 213). Nonetheless, given the discussions around the importance of the ability of teachers' to undertake reliable assessments in high-stakes examinations, the literature on investigations into the reliability of teacher assessments in high-stakes environments remains scant (Harlen, 2004; Wilmut, 2005; Stanley et al., 2009; Ofqual, 2013) and in need of further research (Ofqual, 2019).

Figure 1.3: Extract from the House of Commons Children, Schools and Families Committee Report on testing and assessment.

Third Report of Session 2007–08. *Volume II. Oral and written evidence. Ordered by The House of Commons to be printed 7 May 2008. (Figures in £millions).*

6. The PricewaterhouseCoopers (PwC) report on examination costs, commissioned by QCA in 2003, published in 2005 a figure of £610 million as the cost of the examination system. ASCL has carried out its own surveys from time to time and our figures suggest that the cost is at least that figure. The costs are broadly consistent between institutions of comparable size.

Table

THE COST OF THE ENGLISH EXAMINATION SYSTEM

	Direct Costs	Time Costs	Total
QCA Core costs	8	—	8
QCA NCT costs	37	—	37
Awarding body costs	264	—	264
Exam Centres—Invigilation	—	97	97
Exam Centres—Support & Sundries	61	9	70
Exam Centres—Exams Officers	—	134	134
Total costs (£m)	**370**	**240**	**610**

Source: PricewaterhouseCoopers

12. None of these figures includes the opportunity cost of the time of staff whose main responsibilities lie elsewhere, though teachers, heads of department, and senior leaders all devote a proportion of their time to setting up, supervising and analysing external examinations, and supporting students through them.

There is also evidence that if teachers' assessments in high-stakes assessments are to be trusted, this must be supported by initial and on-going training and accumulated experience:

> …teachers must have training and the opportunity to accumulate experience and enjoy a sufficient degree of public and professional trust to lend legitimacy to their decisions (Eckstein & Noah,1993, p. 235; in Lambert & Lines, 2001: p. 77).

And with reference to an evaluation of the Assessing Pupils' Progress (APP) pilot (see QCA, 2010), Stanley et al (2009) concluded that the use of a shared language between key stages would increase trust between teachers. However, as Chapters 2 and 3 of this study show, the lack of trust around teachers' assessments, including within the teaching profession, remains an issue. This is particularly evident amongst politicians as illustrated for example in a speech by then Prime Minister John Major (1991) calling for a move away from coursework and a return to more traditional examinations and then Secretary of State

for Education Michael Gove (2014a) calling for more rigorous 'knowledge based' examinations. However, there are further tensions *within* the teaching profession regarding a 'worrying lack of trust' in teacher-based assessment as reported by the NAHT Commission on Assessment (2014) and Ofqual (2012).

This study shows that the extent to which teachers' assessment has counted in high-stakes assessments has fluctuated over time and has more lately eroded as the direction of travel has been away from teachers as assessors through the removal of coursework and controlled assignments for GCE and GCSEs; and from 2023 the removal of statutory teacher assessment at key stage 1 and the statutory requirement for schools to report teacher assessment judgements at the end of KS2: from thereon in, test data alone will be used for accountability purposes (DfE, 2017a).

1.6 Teachers' expertise in assessment

Despite the arguments promoting teacher-based assessments there remains a debate as to whether teachers are capable of conducting reliable high-stakes assessments. There is a broad academic literature covering the technicalities of assessment, but arguably very little is imparted to teachers in their initial training phase or through continuous professional development (see for example Carter, 2015). Other than sporadic training in the use of assessment schemes, for example the now defunct Assessing Pupils' Progress initiative supported by the National Strategies and QCA in the mid to late 2000s, there is little if any focus on the underlying theory of assessment provided to practicing teachers at a national level[6]. This results in a general lack of assessment skills and understanding within the teaching force (Navarro, 2008; CFEY/Pearson, 2020). Further, assessment is still not seen as an integral part of the teaching process alongside content and instruction (Navarro, ibid). The lack of assessment expertise amongst teachers is echoed by the Carter Review of ITT (2015) in concluding that:

> Of all areas of ITT content, we believe the most significant improvements are needed for training in assessment. Findings from the NAHT Commission (2014b) as well as Ofsted have also found weaknesses in assessment training (p. 9).

---

[6] See: Ofsted, 2011.The impact of the 'Assessing Pupils' Progress' initiative.

Stiggins (2008) advocates that teachers' assessments should be more broadly trusted through the provision of support from measurement professionals utilising their greater understanding of assessment practice gained in decades of hands on experience in transforming complex validity, reliability and student-involvement assessment concepts and practices that can be used by teachers. From my experience as an adviser to the DfE and the NAHT Commission on Assessment, I have seen first hand the technically limited and often naïve responses of teachers to the removal of National Curriculum attainment levels. A key argument behind the removal of levels was that schools are best placed to devise their own curriculum and assessment frameworks[7]. However, at an operational level, schools continue to struggle to shake off the ingrained practice of using levels and introduce their own robust assessment systems or to develop a clear language around assessment practice and understanding of expected levels of pupil's performance (Brill and Twist, 2013; DfE, 2018c). It would appear that assessment theory is less understood at 'grass roots' level and how teachers learn to assess and apply standards of performance consistently is, as noted earlier in this chapter, an area of little research. As in the case of the demise of the APP programme and the removal of National Curriculum levels, assessment methodology is subject to continuous change and would require substantial CPD to support the existing teaching workforce and the development of more recent graduates.

1.6.1 The Teachers' Standards

There have however, been attempts to guide the content of ITT courses and support for CPD in England. The Training and Development Agency for Schools (TDA[8]) published a set of professional standards for teachers to define the characteristics of teachers at each career stage (2007). The standards included an element of assessment content. Under the heading of Assessment and Monitoring, the three standards required to gain Qualified Teacher Status (QTS) were:

> Know the assessment requirements and arrangements for the subjects/ curriculum areas they are trained to teach, including those relating to public examinations and qualifications

---

[7] See: House of Commons Education Committee. Primary assessment. Eleventh Report of Session 2106-17. Ordered by the house of Commons to be printed 26 April 2017

[8] The TDA was disbanded in 2012 and some of its duties merged with the National College for Teaching and Leadership (NCTL). The NTL was closed in 2018 with elements of its remit moved to the Teaching Regulation Agency and the DfE.

Know a range of approaches to assessment, including the importance of formative assessment.

Know how to use local and national statistical information to evaluate the effectiveness of their teaching, to monitor the progress of those they teach and to raise levels of attainment. (TDA, 2007, pp. 8-9).

The stated intention was for a progressive set of standards so that teachers would deepen their knowledge, understanding and skills throughout their careers. This was to be supported by; "…a contractual entitlement to effective, sustained and relevant professional development throughout their careers" (ibid., p. 3). In 2011, the standards were simplified to a set of core standards and a narrative description of the attributes of what was described as the 'master teacher standard' (DfE, 2011c). A further review of the standards reported in 2015 with the aim of ensuring; "…that the initial element of training is matched to a rigorous and incrementally supportive professional development programme" (Carter, 2015, p. 3). The Carter Review identified what it described as significant gaps in a range of courses with assessment highlighted as needing the most significant improvement (Ibid*.,* p. 9). However, the Review called for a better and shared understanding of the essential elements of a good preparatory course and recourse to CPD noting that the link between the two is often weak. It further recommended the development of a framework of core content and that:

> …a central repository of resources and guidance on assessment should be developed. This would support the learning of trainees, as well as practicing teachers and teacher educators (Ibid*.,* p. 56).

Following Carter's recommendations, *A framework of core content for ITT* was published (DfE, 2016a) giving further clarification on, but not replacing, the existing Teachers' Standards. In each of the reports noted earlier in this section, the importance of CPD was highlighted and although the government accepted the reports, the central repository of support materials covering educational assessment has yet to be developed. This may be picked up in due course by the Chartered Institute of Educational Assessors (CIEA) or newly formed Chartered College of Teaching (CCT) but at this point there is no indication that such a targeted resource is being developed as an open source. It seems clear from the short history of Teachers' Standards that setting standards alone will not necessarily impact practice. Carter acknowledged the deficiency of expertise in assessment amongst

teacher educators and schools but the proposed solution of experts coming together to form a central repository of resources would require financial support and some level of co-ordination to gain traction. Carter also overlooked the workload demands of already crowded ITT courses and as importantly, the formation of different routes into teaching are as yet unproven vehicles for improving the quality of provision particularly in terms of assessment knowledge, understanding and skills. One of the research questions of this study is designed to gain a deeper understanding of the theory and practice of educational assessment made available to newly qualified teachers through their initial training. The research questions are outlined in section 1.8 of this chapter and in more detail in Chapter 3.

1.7 The rationale and intended contribution of this research project

Based on the previous discussion two broad areas are identified to guide the contribution of this research project. These are discussed in the following sections.

1.7,1 The role of teachers in high-stakes assessment over time

Around the mid-point of the twentieth century, the use of externally set and marked assessments became a central feature of the education system in England, for example the introduction of the eleven plus developed in response to the Education Act (1944), the introduction of the GCE (1952) and the CSE (1966). At those points, teachers were seen as important contributors to the development and administration of assessments as envisioned by Norwood (1943). In more recent times, the role of teachers has diminished through reductions in coursework and even their role as examiners in external assessments has been questioned (Ofqual, 2017). In the narrative of teachers as assessors, some of the changes to their role have been subtle and beyond broader public gaze, for example the arrangements for the reporting of estimated grades, whilst others have been more abrupt due to shifts in government policy for example the change in the grading system for the GCSE. However, an understanding of the history of these changes un-tethered by assumptions of progress and set in view of the broader social and political context is lacking.

1.7.2 Perceptions of the expertise of teachers in educational assessment

There is broad recognition in the educational literature of the centrality of assessment in the process of teaching and learning (for example Wiliam, 2013). However, the earlier discussion has highlighted perceived shortcomings in the level of expertise in educational assessment within the teaching profession and broader concern over levels of trust. And despite moves to increase knowledge and expertise through for example the publication of the Teachers' Standards (TDA, 2007) and the Review of Teachers' Standards (DfE, 2011b&c), deficiencies in the preparation and continuous professional development of teachers, particularly in the domain of educational assessment are evident as highlighted in the Carter Review (2015). Carter also highlighted deficiencies in the expertise found in schools and ITT providers and now that much of the ITT in England is spread through a range of less traditional training schemes, for example School Centred Initial Teacher Training (SCITTS), Schools Direct, Troops to Teachers and Teach First, exactly what assessment theory and practice is made available to trainee teachers sits beyond public access and is difficult to gauge from currently available literature.

It has to be recognised that there is an issue around just how much time can be realistically devoted to assessment theory and practice in initial teacher preparation given the competing demands of other areas of pedagogical study, subject knowledge and school placements. Sitting alongside questions about the assessment expertise of trainers, I would contend that it would be of use to separate out exactly what graduate teachers should know about assessment, what should be developed in the early years of their teaching experience and what should be made available through continuous professional development. (These three groups or stages in the career development of teachers reflect recommendations for teacher development outlined by Alexander et al., 1992). So if teachers are to be involved in a system of high-stakes testing designed to assess a broader range of attributes than can be gained by an externally set test, there is a need to develop a greater understanding of the reliability of teacher assessment and to develop improved training in assessment understanding and practice to build confidence in their judgements (Johnson, 2012). There is also an argument that involving teachers in summative assessment promotes teacher professionalism and expertise more broadly (Harlen, 2004).

However, in order to establish a programme in educational assessment that would prepare and then support teachers, there is a need to understand in detail the current level of

provision in ITT. Studies have been undertaken in the USA showing little emphasis on assessment training and where it was evident, questions were raised about the quality (Wolmut, 1988, in Stiggins and Conklin, 1992; Stiggins, 2008). Yet there is little indication of syllabus content and perceptions of quality of provision relating to educational assessment in the various prospectuses of teacher education courses in England. Further, there is a lack of understanding about teachers' perceptions of their preparedness to enter the classroom experiences as students and teachers and how this sits alongside the views of ITT providers and those who hold positions of key influence in the English education system.

1.8 Research questions

This research study is presented against a backdrop of the 2020 Covid-19 pandemic which despite the reservations over teachers' assessments identified earlier in this chapter have resulted in a volte-face over the role played by teachers in high-stakes assessments. For the first time in the history of the English education system, teachers' assessments will be used to determine the final grades achieved by over 850,000 GCE and GCSE students in high-stakes examinations without any ramifications for the performance measures of schools, teachers or ministers. This is at odds with a status quo built on the marginalisation of teachers and the centrality of examination outcomes in school accountability tables and the internal performance management of teachers. Post summer 2020, Ofqual, the awarding organisations and policy makers will have a rare opportunity to compare data generated by a wholly centre based approach to assessment without any implications for accountability measures, with that of a largely externally set and marked regime.

The intended outcome of this research study is to provide an insight into the role played by teachers in high-stakes assessment over time and an understanding of the extent and perceived quality of current provision in educational assessment theory and practice found in current ITT courses. In the light of the Covid-19 pandemic, it is hoped that this research project will make a further contribution into the post-pandemic review of the role of teachers in high-stakes assessments in England and inform future provision for supporting teachers in their assessment practice.

This can be presented in three key questions:

1. How has the role of teachers in educational assessment changed since the Norwood Review (1943) to the present day?

2. How do teachers and Key Influencers[9] perceive educational assessment and the role played by teachers in the English education system?

3. What does the ITT curriculum in England offer in terms of assessment theory and practice?

The research questions are developed in further detail in Section 3.3 in Chapter 3, Methodology.

1.9 The research approach

This study uses a mixed methods approach based on the view that mixing the analysis of data drawn from different approaches will provide stronger answers to the research questions than might be generated from either a quantitative or qualitative approach alone (Creswell and Creswell, 2018). The research uses a range of approaches including the use of educational literature, policy and other relevant documentation (Scott, 1990); face-to-face semi-structured interviews with teachers and Key Influencers, and surveys of teachers and initial teacher training providers. The chosen methods are driven by the primacy of the research questions in an approach that can be described as being one of 'pragmatism'; "…generally regarded as the philosophical partner for the mixed methods approach" (Denscombe, 2008, p. 273). The pragmatic approach is not tethered by adherence to either a qualitative or quantitative method of research, but rather one driven by an underlying philosophy that endeavours to produce useful, practical outcomes (Cohen et al., 2018). The intention is to offer possibilities for change in educational assessment practice that is relevant and attainable by teachers and policy makers that are focused on educational outcomes rather than on becoming the tools of competing political ideologies. Michel Foucault's pragmatism is central to the philosophical approach that underpins this research project in presenting a critique of the levels of confidence in teachers' involvement in high stakes educational assessments over time in a manner that does not pre-suppose progress but rather examines change. But the intended outcome is not simply one of using Foucault's concepts of archaeology and genealogy as a form of analysis, but rather by applying these concepts as I believe Foucault intended as tools to understand how things change, transform and are displaced; "…in order that it may be of use" (Foucault, 1976, cited in Defert and Ewald, 1988, p. 911-912). The intended usefulness of this critique is to utilise Foucault's pragmatism (See Chapters 3 and 4) to

---

[9] The term Key Influencers is used to describe people in positions of influence in the English education system for example as senior politicians, regulators or leading academics in the field of educational assessment.

develop research evidence that can influence and support strategies to engage and recognise teachers in the process of educational assessment through improved professional competency and building broader confidence amongst the wider public and political communities (see Chapter 9, Discussion). In examining what Reynolds (2004) describes as Foucault's "pragmatic humanism", the author states:

> The aim of critique is thus to function as philosophical resolutions to those problems that can no longer be resolved through traditional means. The critic thus 1) traces through genealogy the contingencies that have made subjects what they are (conditions and consequences); and 2) through doing so, opens up the possibilities to new and experimental forms of subjectivity (p. 959).

A full account of the methodology used in this research project can be found in Chapter 3. This is followed in Chapter 4 with an examination of the role played by teachers in high stakes assessments since the mid-twentieth century using the work of Foucault as a framework of enquiry.

1.10 Chapter 1 summary and structure of this thesis

1.10.1 Summary

Chapter 1 has presented an over-view of the changing role of teachers in what have become high-stakes educational assessments from the mid-twentieth century to the present. It has highlighted a change from one that held a vision for a regime in which teachers would play a central and trusted role to one in which their involvement has become marginalised in a system focused on performativity and high accountability. Opposing views on the place of teachers' assessments have been illustrated with proponents arguing for increased use of teachers' assessment as an aid to improved validity and opponents raising questions over reliability in a system in which assessment outcomes are a central feature of school and teacher accountability. Concern has also been noted over the level of trust in teachers' assessment within the teaching profession and the levels of expertise held by teachers and teacher educators in assessment theory and practice. Yet little is known about the detailed content of ITT provision in this respect or the levels of confidence in the adequacy of such provision. Based on this analysis, this chapter introduced two broad areas on which to focus this research and three research questions. These are addressed in the following chapters.

1.10.2 Structure of this thesis

Following the introduction to this research project, Chapter 2 reviews the academic and other documentary literature relating to teachers' assessment in the English education system. It examines the notions of low and high-stakes educational assessments in what has become a highly centralised system of school accountability and culture of performativity. Questions are raised over the levels of expertise of teacher and initial teacher educators along with the issue of trust in assessment outcomes generated by teachers. Chapter 3 presents the methodology used in this research project with further details on the philosophical underpinnings of this study and research questions. Details on the research design and approach are also presented. Chapter 4 is the first of five data analysis Chapters (4-8): this chapter presents an analysis of a range of documentary and other sources such as parliamentary records and the mass media, and draws on the work of Paul-Michel Foucault as a means of understanding changes to the role of teachers over time in what have become high-stakes assessments. The chapter identifies three epochs, what Foucault describes as epistemes, showing three distinctive periods of change in the role played by teachers. Chapters 5 and 6 present an analysis of the data from the ITT provider and teacher surveys respectively. Chapter 7 provides the analysis of a series of three interviews with six teachers, three from the primary phase of education and three from the secondary phase. Chapter 8 presents the analysis of interviews conducted with fourteen Key Influencers. Chapter 9 presents a discussion of the findings extrapolated from the research data. Chapter 10 presents a set of conclusions drawn from the research data and discussion including the recognition of the limits of this research project, a set of recommendations for further research and some personal reflections on this research project.

Chapter 2: Literature Review

2.1 Chapter overview

This chapter reviews the academic literature and other documentation relating to the role of teachers in educational assessment since the early 1940s. Assessment in the context of education is discussed along with an over-view of the various routes into teaching and their implications for the degree to which educational assessment theory and practice forms part of initial teacher training (ITT). The growth of assessments used as a measure of school accountability is examined in light of their impact on the English education system. In particular, the chapter identifies issues around the reliability and level of trust afforded to teachers' assessments in what can be described as *low* and *high-stakes* assessments; these distinctions are examined in Section 2.6 below. The chapter ends with an examination of calls for increased teacher involvement in high-stakes assessments and the provision of content relating to educational assessment theory and practice in initial teacher preparation.

2.2 Introduction

Much of the literature on educational assessment referenced in this chapter is routed in the United States and the United Kingdom, particularly England, along with policy and other relevant documentation (Scott, 1990). These jurisdictions are bound by areas of common ground, for example on the use of assessment in large scale testing systems and the call for what has been described as assessment literacy amongst the teaching profession and indeed beyond (Stiggins, 1991; Gardner, 2007; Klenowski, 2012). However, on the area of trust and what teachers really need to know about assessment to function in the English education system, the literature is sparse. Reference to trust, or the lack of trust, is common but research into what would re-gain a level of trust is missing. As for what teachers really need to know about assessment, this may be down to the view that there is no real agreement on what newly graduated teachers need to know and be able to do (Ingvarson and Rowley, 2017). Further, concern has been raised about the level of assessment expertise amongst teachers in England and amongst those responsible for educating teachers (Gardner, 2007; Carter, 2015). The chapter begins by looking briefly at the terms educational assessment and high and low stakes assessment and then considers how the role played by teachers in educational assessment has changed over time. Perceptions of trust in teachers' assessment are then considered with reference to the contexts in which such assessments are made and includes a review of attempts to

increase the reliability of teacher based assessments. The chapter concludes that there is a need for better understanding of the range and quality of assessment theory and practice provided for trainee teachers and how this compares with what teachers really need to know in order to be effective practitioners and trusted contributors to the educational assessment system in England.

2.3 Assessment in the context of education

The etymology of 'assessment' traces its origins back to the Latin usage meaning 'to sit beside' and used in the context of sitting beside a judge to assist in determining fines or levels of taxation (etymonline, accessed 12th February, 2020). In modern times, assessment is applied in many contexts, for example in taxation or property valuation, and requires an act of judgment or a decision on an amount, value, quality, risk or importance of something. So for example in terms of personal taxation we see a process of self-assessment: in the field of project management the process of risk assessment and in health care the use of health risk assessments as used by insurance companies. In an educational context, Pellegrino (1999) notes that assessment has become:

> …part of the everyday discussions among citizens, educational professionals and politicians. Not coincidentally, external assessments have become the instruments of the accountability movement (p. 5).

Pellegrino identifies the work of Lee Cronbach in 1957 in "… linking theories and research on learning instruction with the tradition of assessing individual differences in cognitive abilities influences from scientific psychology" (Pellegrino, p. 7) as a 'salient starting point' of key influences on assessment practice. And an analysis by Collins Dictionary (on-line, 2020) would suggest that from the mid-point of the twentieth century, the use of the word 'assess' was becoming more common.

Figure 2.1: Word trend graph on the usage of the word 'assess' [10]

Note: the lines on the x-axis represent 1955, 1979 and 1995. Source, Collins Dictionary. Viewed On-line 13th December 2019.
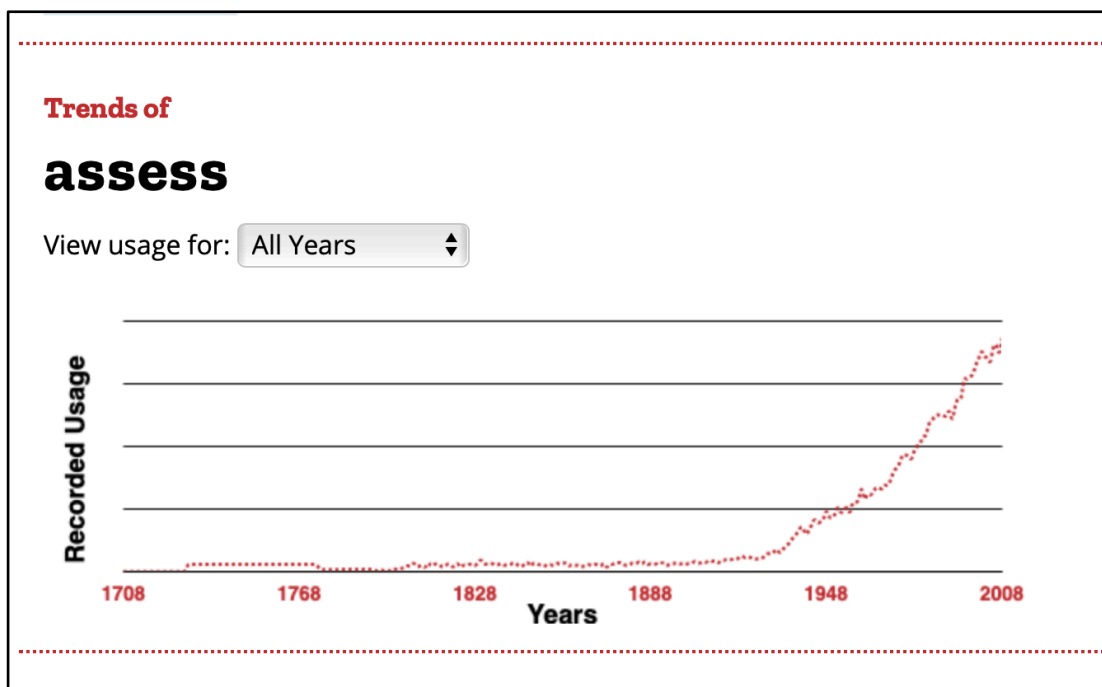
However the term 'assessment' often lacks precision in the field of education (Newton, 2005 and 2007), and is used to describe a range of activities that are put to a variety of uses. Such activities include large-scale assessments, for example:

1. examinations for formal qualifications used to recognise the attainment of individual students, as an accountability measure of institutions in national performance tables and as a measure of educational performance standards over time:

2. end of key stage tests, though they do not provide a qualification as such for the individual pupil, they are frequently used by secondary schools to place individual pupils into ability groups and used as a measure of accountability for primary school performance tables.

---

[10] The word trend graphs are based on information from Google n-grams, and the numbers represent relative frequency of words across books (i.e. written form only) up to 2008. More information on the creations of these graphs can be found at https://books.google.com/ngrams/info.
Source: e-mail from Collins Dictionaries: dictionaries@harpercollins.co.uk

At another level, day-to-day applications of assessments in classrooms are used to inform future teaching and periodic assessments, for example those used at the end of a block of work or end of a school year, are used to measure pupil's progress (QCA, 2009a). In attempting to clarify what makes an assessment educational, James (2010) draws a distinction between educational assessment and educational measurement, the latter described as a term used more widely in the United States to describe formal testing programmes covering the awarding of qualifications, monitoring and accountability. These programmes:

> …are sophisticated procedures and require considerable expertise, time and other resources to develop and implement them properly. For this reason educational measurement is often large scale and carried out, at least in part, by professional agencies external to the schools in which tests are administered (ibid., 2010, p. 162).

However, in the United States, 'assessment' is acknowledged as a broader term commonly describing a process that integrates test information with information from other sources. The 'other sources' cited in The Standards for Educational and Psychological Testing (AERA/APE/NCME, 1999) include examples such as "... information from an individual's social, educational, employment, or psychological history" (p. 3). As such the term assessment includes areas beyond those described as solely educational.

In the England, common usage of terms associated with assessment in the context of education often lack precision as noted by Newton (2007) for example in the way that the words 'exams' and 'tests' are used interchangeably in the mass media. In my view, 'exams' (examinations) should refer to assessments used in certificated qualifications and the use of 'tests' should be clarified in reference to either assessments used in National Curriculum statutory assessments or more locally by schools to assess pupil's progress.[11] My experience of working with teachers in schools also suggests a lack of precision where terms such as formative and summative assessment are used to describe seemingly diverse methods of assessment practice whereas in more technical considerations they are defined by the inferences drawn from what could be the same assessment instrument (Wiliam,1999): and as Paul Black has argued:

---

[11] See QCA, 2009(a) and QCDA, 2010 & 2011 for a useful description of assessment terms for schools to use in relation to their approaches to educational assessment.

…the two functions are two ends of the same spectrum and that there is no sharp difference, and that if the two functions are separated, then teachers' assessment work will be devalued (1998, p. 34).

The inferences drawn from an assessment are of great importance here. In practice, it is as much the inferences drawn of an assessment that are central to its claim to validity as much as the technique of an assessment. Validity is therefore not a property of tests, or even of test outcomes, but a property of the inferences made on the basis of these outcomes. As Cronbach and Meehl (1955) noted: "One does not validate a test, but only a principle for making inferences" (cited in Wiliam,1998. p. 3).

The uses – decisions, processes or actions - to which assessment outcomes are allocated is of central importance in designing an assessment instrument and therefore in deciding its validity and relationship with the educational context. Newton lists 22 purposes for educational assessment (Newton, 2007), and he acknowledges this is not exhaustive. However the list is helpful in establishing a description of educational assessment as each purpose identified can be classified as either applying to the learning of the individual (student) through to purposes concerning larger entities involved in the educational process based on aggregated results, for example at school or cohort level as measures used for school accountability. Nevertheless, Newton (2007) concludes there is a need to use the language of assessment with greater precision (see also Ofsted, 2016, p. 50).

What unites the purposes as identified by Newton (ibid) is the use of assessment as a tool for teaching and learning, as a form of accreditation at the end of a period of teaching and learning or as a tool for accountability. James (2010) provides a useful insight into the term educational assessment as used in the UK as:

> … all those activities that involve eliciting evidence of student learning and drawing inferences as the basis for decisions (p. 162).

Therefore my interpretation of educational assessment is based on James' definition and drawing on the work of Newton (op cit) includes two broad areas:

> 1. in the range of activities that inform or provide evidence of student learning at the classroom level where educational assessment is integral to the process of teaching and learning, for example when teachers use questioning in their lessons

to gauge a pupil's understanding of a concept in order to measure the effectiveness of teaching or deciding the next steps in learning: and

2. in the more formal context where assessments are in the main set by external bodies for example in examinations to award qualifications at the end of a period of study or in National Curriculum assessments introduced as a measure of pupil attainment at the end of key stages of education.

All further references to educational assessment throughout this thesis refer to the above interpretation.

2.4 Initial Teacher Training: routes into teaching and issues

At the time of writing, individuals wishing to train as teachers for employment in state maintained schools in England are required to hold a degree (or equivalent), GCSE grade C/4 in English and maths (or equivalent) – and a science subject for primary trainees, and on completion of their training, achieve Qualified Teacher Status (QTS) which is measured against the Teachers' Standards (DfE, 2011c). However, systematic training of schoolteachers in England has a history covering two centuries. This history reflects an almost continuous cycle of recruitment and retention issues and the establishment of various routes into the profession including apprenticeships in the 1840s (Dent, 1977; Robinson, 2006) and the establishment of training centres running three year courses in the latter end of the nineteenth century (Cross Report, 1888) culminating in numerous teacher training colleges (Maclure, 1979). The Robbins Report (1963) recommended that training centres should become Colleges of Education with closer relationships with universities under whose auspices degrees could be awarded. Following the recommendations of the James Report (1972), the foundations for an all degree profession were formed (Dent, 1977; Maclure, 1979; Robinson, 2006).

Since the early 1990s, there has been a gradual but ever-clearer shift away from university dominated ITT. This is exemplified by the North of England Conference speech by Kenneth Clarke, then Secretary of State for Education who called for an 80% school-based training model (Robinson, 2006). School/university partnerships were developed under the Teacher Training Agency – later re-named as the Teacher Development Agency – with the latter working to a brief to develop more diverse routes into teaching. Much of this was criticised by university providers who arguably had a vested interest in their own position (ibid). More recently, under the 2010 Coalition Government, the debate has become more

heated. The debate has not cooled under the more recent Conservative administration. In short, the emphasis of policy has been in favour of school-centred training apparently led by ideology rather than informed by research and evaluation. As Golding (2015) has noted:

> It is far from obvious that the changes in policy are entirely driven by robust evidence of the needs of either the system as a whole or beginner teachers in particular (p. 118).

What is clear is that there are a growing number of routes into teaching at this time with little in terms of the rationale for the various approaches or clarity in their approach (Roberts and Foster, 2016). And despite criticism in terms of their value for money (Allen et al.*, 2016, p. 5), the current range of training routes continue to offer variations ranging from university based to schools based options each with differing periods of course time and differing course structure (see Appendix 1: Routes into teaching).

Some of the political rhetoric around government policy has also been unfortunate, for example the then Secretary of State Michael Gove referring to any part of the education establishment questioning government reforms as 'the blob', defined as 'academics who have helped run the university departments of education responsible for developing curricula and teacher training courses' (Gove, 2013a). Such a position is unlikely to foster intelligent debate about educational provision. Michael Gove made his intentions very clear: 'We are moving teacher training away from university departments and into our best schools' (ibid). More recently the position on university based teacher training appears to have been softened somewhat, (Hazel and Ward, 2017) perhaps in the light of increasing teacher shortages, the continuing strength of university teacher training and the less certain future of the many small scale school-based providers who are more prone to the movement of key staff, attracting potential candidates in their often specific offerings and ever shifting government policy. However, the future structure and balance of teacher training remains uncertain (Allen, et al., 2016) as can be demonstrated by the debate over teacher apprenticeships and the shifting patterns in entries to the various routes into teaching (see Appendix 1).

2.5 Teacher knowledge of educational assessment theory and practice

The literature covering the role of teachers in educational assessment raises frequent questions in regard to their technical reliability (see for example Opposs and He, eds. 2012; Ofqual, 2013; Rimfeld et al., 2019). However, in the context of a system of school accountability and performance management of teachers centred on test and examination outcomes, the level of trust afforded to teachers' judgments has become a more central topic of concern; this is covered in more detail in Section 2.10. Moreover in more recent times, the focus has shifted to the level to which teachers are educated in the theory and practice of assessment (see for example Carter, 2015). Therefore a particular focus of this study involves developing an understanding of the level, depth and appropriateness of the theory and practice of educational assessment as an element of initial teacher training. Since the requirement for teachers to be suitably qualified to teach was introduced as a part of the national educational policy in England, there has been a number of developments in the recruitment, qualifications and structure of teacher preparation routes over the past years. Changes over the last seven years have created several alternative routes into teaching with a deliberate policy intention of moving the emphasis of training from universities and in to schools. This has particular consequences for this study in that one strand of the project is to gain an understanding of the current content and structure of courses leading to qualified teacher status. However, current pathways into teacher training are varied leading to large numbers of training institutions with differing scales of provision: therefore attempts to investigate the teaching content or syllabuses of courses are more demanding.

2.6 'High-' and 'low-stakes' assessment

There are two broad terms describing educational assessment activity that are relevant to the topic of the role of teachers in the educational assessment system of England – in one teachers are central and broadly left to their own devices: in the other teachers have become increasingly marginalised. The literature on educational assessment frequently refers to *high-stakes* and *low-stakes* assessment (for example Allen*,* 2012; Cole and Osterlind, 2008; Gregory and Clarke, 2003). In essence, these terms describe the uses to which educational assessments are put, what could be described as their function, rather than differences in the design or form of the assessment instrument (Glossary of Educational Reform, 2014). As defined in Chapter 1, high-stakes assessments are frequently characterised as holding substantial or significant consequences for those

involved in the assessment (Stobart, 2008; Gregory and Clarke, 2003; Rosenkvist, 2010). More recently there has been a growing trend across the world for high accountability systems whereby schools and teachers become the focus (see for example OECD, 2010a). In England the introduction of the National Curriculum and associated statutory assessment system in 1988 followed by the publication of results in GCE, GCSE and national curriculum tests in the mid 1990s has raised the stakes for schools and teachers through what are frequently described as school league tables. The publication of outcomes of general qualifications and tests has become a key measure of holding schools and teachers to account: an approach that has become a global feature of high performing countries (PISA, 2011). The Department for Education (DfE) has promoted the approach of robust accountability, aligned with high levels of autonomy, as a key feature of the most effective education systems around the world (DfE, 2013a). However, this interpretation of the 2009 PISA diet (ibid) has been challenged once the data for public and private schools are separated (Benton, 2014), a detail perhaps overlooked by the DfE.

The DfE has produced performance tables for secondary schools in England since 1992 with data from primary schools added in 1996. Performance tables were introduced by then Secretary of State for Education John Patten who stated he was:

> …consigning to the dustbin of educational history a system which denied parents the right to know how schools are performing and prevented them from making informed choices about where they want their children educated (1992).

The actual content of exactly what data are published by the DfE varies from year to year through the publication of an annual statement of intent (see for example, DfE, 2014b & 2015a) but to my knowledge the Department has never used the terminology 'league tables' which is more a creation of the mass media. However, the approach has fuelled broad public interest in the outcomes of high-stakes assessments and as a result impacted on the way schools organise and deliver their programmes of work leading to criticisms such as the narrowing of the curriculum to that which is tested and negatively affecting the wellbeing of teachers and pupils (Education Select Committee, 2017). In the United States, particularly from the passing of the of the No Child Left Behind Act in 2002, performance in high accountability tests has primarily affected teachers and schools through a system of rewards and sanctions for schools (Miller et al., 2009). In England, the publication of results for National Curriculum assessments for key stages 1 & 2 (and up to 2008 for key stage 3), arguably has less impact on individual pupils than the publication of GCE and

GCSE results, but it does heighten the stakes for schools and teachers – and in areas with *'high achieving'* schools, even impacts house prices (Gibbons, 2012).

The unintended consequences of high-stakes testing such as teaching to the test and accusations of cheating have raised broad concern in England, (for example Bew, 2011a&b; Mansell, 2015; Taylor, 2015; Coughlan, 2017; Ward, 2017) – points not missed by authors in other countries where similar high-stakes assessment systems have been considered (Klenowski, 2012). More recently, in an attempt to counter the focus on examination and test outcomes, Ofsted has indicated it will take broader interest in the breadth of the curriculum (Spielman, 2017; Ofsted 2019). Conversely, much less effort is made to define 'low stakes' assessments. In general it would appear to apply to any use of assessment outside of that deemed as being high-stakes and is more confined to internally set assessments used by schools more often as a means of informing learning, commonly referenced as 'formative assessment' where the external validity of a teacher's feedback was no longer important' (Black and Wiliam, cited in Allen, 2012, p. 657). This has in practice created a division between formative assessment, generally perceived as 'low stakes' and 'high-stakes' assessments that are usually in the form of externally set standardised assessments commonly referred to as summative assessments. Of course schools use a range of assessment strategies covering day-to-day, periodic and transitional needs (QCDA, 2011). However, it is end of key stage assessments that focus on achievement that characterise assessments used for purposes of accountability. From my many discussions in schools, teachers have developed an antagonistic view of external high-stakes assessments – largely based on their status as accountability measures rather than anything intrinsic in their form. Unfortunately, their form and function are often inaccurately conflated. This view persists even where teachers' assessments are included in performance data. From 1996, teachers' assessments were collected and included in the published key stage 1 and 2 performance data. However, publication of teacher data ceased in 1999 even though the submission of data continued to be a statutory requirement. In 2010, in my capacity of accounting officer for National Curriculum assessments, I helped to persuade the DfE to re-introduce teachers' assessment data – but failed to persuade them of their equal worth to the externally set tests. My argument was based on the view that teachers' assessment data combined with the external tests would provide a more rounded assessment of pupil's performance; ministers accepted the release of teacher assessment data, but were of the view that external tests were a more reliable measure and therefore of higher weighting. In short, trust in teachers' capacity to produce reliable assessments was lacking, a view that persists (for example NAHT, 2014; Gardner, 2007). But assessment is an essential element in the teaching and learning

process (QCDA, 2011) and the teaching profession should be promoting and developing their expertise in this aspect rather than loading the impact of assessment into what is primarily an argument about excessive workload. This is exacerbated by the view and that teacher assessment does not hold the same significance as external tests. This resulted in the decision by Ministers to the removal of the collection of teacher assessment data for key stage 2 from 2019, subject to a change in legislation, on the grounds that it does not count in accountability measures and reduces teacher workload (DfE, 2017a). Removing the collection of teacher assessment data on these grounds misses the point: it should be a part and be of equal importance (QCA, 2002).

The broad categories of high and low-stakes assessments can be viewed as an over-simplification or distortion of reality. All forms of assessment can be used for formative and summative purposes to some degree and what might appear to be 'low stakes' assessment used by teachers in routine situations such as homework, projects or end of year tests, to a pupil or parent these may be viewed as significant. If an assessment influences what happens to the student or what happens in a school it should be recognised as high-stakes (Popham, 2010). Further, the use of assessment as a central measure of accountability has made the process contentious even leading to the separation of assessment from teaching (Stiggins, R. 2014).

2.7 The growth of high-stakes assessment in England

In the 21st Century, education is established as a basic human right in society (UN, 1948). Educational performance measured by participation rates and formal qualifications have as a result been a long-standing feature of society. In England, there has been enormous growth in the number of qualifications awarded to pupils in the education system from a relatively small proportion of the population in the 1950s. For example in 1953-54, three years after its introduction, 10.7% of the relevant age group passed five or more General Certificate of Education Ordinary levels (GCE O levels) at schools in England and Wales. 5.5% of the relevant age group passed one or more subjects at General Certificate of Education Advanced level (GCE A level). In comparison, by 2010/11 79.6% of pupils in their last year of compulsory education in the UK achieved 5 or more General Certificate of Secondary Education (GCSE) grades A*-C or equivalent and in 2003/04, 39.2% of the relevant age group passed two or more GCE A levels or equivalent (Bolton, 2012).

In examining trends in GCE O level/GCSE[12] achievement from 1954 when the provision of formal qualifications was for the few with 10.7% of pupils passing five or more GCE O levels with the late 1980s, a similar percentage (10%) of school leavers left school with no passes (Bolton, 2012). Conversely, in 2018; "…18% of students (98,779 children) left education without achieving this standard (5 or more GCSEs grade C/4) and above, up from 14% in 2015" (Children's Commissioner, 2019). The growing trend of pupils taking GCE qualifications reflected the view already established by 1959 that success in the GCE examination at Ordinary and Advanced level had become the passport to higher education, the professions or jobs attracting higher remuneration. At this point in time around a third of all GCE entries came from schools other than grammar schools (Beloe, 1960).

The current emphasis on externally set and marked assessments as measures of accountability is some way removed from the situation found in the mid-twentieth century when there was considerable debate around the desirability of formal externally set and marked qualifications. The Norwood Review of the curriculum and examinations in secondary schools (1943) raised questions about the desirability of external qualifications. This view was mainly grounded in fears of a narrowing of the curriculum and teaching to the test raised by the Spens Report (1938). The Norwood review concluded that with regards to the School Certificate Examination (the pre-cursor to the GCE O level), examinations should be entirely the responsibility of the teaching profession; "…making the examination entirely internal, that is to say, conducted by the teachers at the school on syllabuses and papers framed by themselves" (I943, p. 140).

The Beloe Report (1960) re-visited the arguments for and against external examinations. The report acknowledged what it referred to as the 'tonic' effects of external examinations described by teachers and administrators not only for the pupils entered for examinations but also other pupils in the schools in that they provided external validation and improved the tone and self-esteem of the school as a whole. Conversely, the same constituents raised fears of the examination dictating practice, stifling free choice and encouraging wrong values in the class-room: again, viewpoints that still resonate today. At the time, Beloe noted the risk raised in the Ministry of Education Circular 289 that examination

---

[12] The General Certificate in Secondary Education (GCSE) incorporated and replaced the GCE O level and Certificate of Secondary Education (CSE) for 1986 – with the first award in 1988.

results could become an oppressive index of the efficiency of schools (1960, para.77, p. 24.)

By 2003, Murphy (2003) reflected on what he described as an increasingly challenging world in which pupils have to compete for higher education, training and paid employment on the basis of their qualification grades leading to his conclusion that at the end of compulsory education qualifications have become "…of considerable importance to every parent, citizen and employer" (p. 182). The risk of qualifications as a narrow tool for accountability of school performance identified in Circular 289 was a sign of things to come (see Aldridge et al., 1991).

A key driver of general qualifications in England such as the General Certificate of Education introduced in 1951 was one primarily concerned with selection (Bloom, et al.,1971; Wolf, 2002; Beloe, 1960). Over time qualifications have developed to meet the needs of a wider population such as the Certificate of Secondary Education (CSE) in the mid 1960s. At the time, GCE O level was assumed to cater for up to 20% of the 16 year old age group taking four or more subjects. The CSE was then envisaged to target the next 20% below these again in four or more subjects with up to a further 20% of the age group attempting individual subjects (Beloe, ibid).

By 1988, the CSE was absorbed along with the GCE O level into the General Certificate of Secondary Education (GCSE), awarded from 1988 onwards: unlike the GCE O-level and the CSE, there was no predetermined target percentage of the ability range of candidates in mind for the GCSE. (Mobley et al., 1986). Other qualifications to accredit achievement in vocationally orientated education such as the National Vocational Qualification (NVQ), the General National Vocational Qualification (GNVQ) and the GNVQ Part One in the 1990s were also developed. Despite almost continuous changes to the focus and format of these qualifications – and even the demise of some - the result is that the vast majority of the school population now sit examinations for qualifications.

2.8 The changing role of teachers in general qualifications and National Curriculum assessments in England

Throughout the period of development of general qualifications, the role of teachers in the assessment process has changed. Clearly around the time of the Norwood and Beloe reports, the inclusion of teacher assessment was accepted as central to the educational

assessment process. When the CSE was introduced in 1965, teachers had a key role in running the examination through local representation on 14 regional awarding bodies governed by teachers and local education authorities. Committees made up of subject teachers devised syllabuses for the awarding bodies to administer through three modes. Mode 1 was the provision of a syllabus and assessment scheme totally administered by the awarding body including setting and marking an external examination: Mode 2 allowed schools to submit their own syllabus for approval and assessment by the awarding board. For Mode 3, the syllabus and scheme of assessment was determined by the school (or group of schools) and approved by an awarding body. The school(s) set and marked examinations and other aspects such as coursework. These were subject to external moderation by the awarding body.

When the GCSE came in to being in 1986 the use of differing modes of examination continued. There was greater use of coursework than found in the GCE O level to assess aspects such as research skills, practical work, oral skills, and planning and design skills. In some subject, for example Business Studies, there was an upper limit of 40%. In English the whole course could be assessed through 100% coursework. This continued the tradition set by the CSE and was seen to increase teachers' confidence and develop expertise throughout the profession (Mobley, et al., 1986). Since that time there has been a gradual reduction in the role teachers have played in developing examinations such as those covered by Modes 2 and 3 and in the marking of coursework and controlled assignments contributing to the assessment of general qualifications. And despite assertions about a growing reliance on teacher assessment in high-stakes assessment (Johnson, 2013; Stanley et al, 2009 in Johnson 2013), recent changes to GCSE and GCE level qualifications suggests quite the reverse.

National Curriculum assessments have followed a similar pattern. The initial assessment framework set by the National Curriculum Task Group on Assessment and Testing (TGAT, 1988) was based exclusively on teacher assessment with a system of externally set Standard Assessment Tasks (SATs), administered and marked by teachers and externally moderated. However, this approach lasted only a few years due in the main to concerns raised by teachers regarding workload issues. This resulted in the boycotting of early SATs by the teaching associations in the early 1990s with the tactic repeated again in 2009. Following a review of the National Curriculum by Lord Dearing in the mid 1990s, SATs were subsequently replaced with external tests for English, mathematics and science. Whilst understanding the teacher associations' concern about workload, in my view the marginalisation of teachers from key end of phase assessments undermines their position

as professionals. The teacher associations failed to think more strategically about their role in high-stakes assessments. The external tests for key stage 3 were dropped in 2008 in response to pressure from teacher associations and changes to key stage 1 tests were made to give teachers a more central role in the assessment process, but key stage 2 tests continue to this day as do the publication of results in the form of performance tables. This is despite a further teacher boycott in 2009 of about 26% of schools over the impact of tests which at the time were viewed by the NAHT as 'blighting childrens' education' and 'humiliating teachers' (Lightfoot, 2009). However, the term 'SATs' is still used today in teachers' common parlance to describe National Curriculum tests. This reinforces the need identified by Newton (2007) for more precision when using assessment terminology.

2.9 The growth of educational assessment as a measure of school performance in England

Of growing importance during the period under discussion has been the use of assessments from one of recognising or certificating the achievements of the individual student to one of evaluating the effectiveness of the education system as a whole - the performance of schools, individual teachers and even government ministers as a key measure of the school accountability system. The work of Foucault is of relevance here in explaining the use of educational assessment as a technology, an instrument of power that differentiates, ranks and holds the body to account (see Foucault,1976): this presents an interesting perspective that is developed more fully in Chapter 4. However, focus on the examination has intensified since the introduction of school performance tables in the 1990s based on what have become high-stakes assessments covering general qualifications and National Curriculum tests and a culture of performativity (see Ball, 2003, 2013). It is argued by some that the introduction of these accountability measures has resulted in 'unintended consequences' or 'perverse incentives' (for example Klenowski, 2012; Bew, 2011a&b; Mansell, 2015) and has resulted in growing doubt over the efficacy of teacher assessments and a gradual reduction in the part teachers play in high-stakes assessments (Lightman, 2015). Drawing on elements of the work of Foucault, I also argue in Chapter 4 that much of the change in teachers' involvement in high-stakes assessment has been driven more by issues of power and control than a drive to improve the validity and reliability of educational assessment.

Since the first report on Trends in International Mathematics and Science Study (TIMSS) in 1995[13], the Programme for International Student Assessment (PISA) in 2000[14] and the Progress in International Reading Literacy Study (PIRLS) in 2001[15], international assessment programmes have become a key comparator in evaluating the success or otherwise of economies across the world and a key point of reference for Her Majesty's Chief Inspector (HMCI) of schools and politicians (HMCI, 2012). Much of this has been intensified by what Murphy (2003) describes as the now annual "…focus of media hype, in a way that would have been unthinkable thirty of forty years ago" (p. 186). Indeed, some would argue that the now common use of test and examination outcomes to measure school performance has become the key measure (Jenkins, cited in Stobart, 2008). Stobart refers to this as an 'audit explosion' and despite the aim being to support confidence in the system it leads to a decrease in trust and an increasingly defensive profession (ibid). However this view is countered by government ministers, amongst others, who would argue that more robust accountability measures lead to improvements in educational provision (DfE, 2013a&b; OECD, 2010; PISA, 2011).

2.10 Perceptions of trust in teachers' assessments

The level of trust in teachers' assessment has become an area of focus in the educational literature (ATL et al., no date; Brookhart, 2011; Gardner, 2007; Harlen, 2005a; Johnson, 2013; Looney et al., 2017; Stobart, 2008; Wolf, 2008 & 2014). Two main issues appear to underpin much of the discussion and discourse around trust:

> 1. the unintended consequences of a high-stakes assessment system, for example the pressure exerted by the publication of performance tables which have created a perverse incentive to inflate assessments (Lightman, 2015); and

> 2. concern over the level of assessment expertise in the teaching profession (NAHT,2014).

---

[13] Association for the Evaluation of Educational Achievement (IEA). (1995) Report on Trends in International Mathematics and Science Study (TIMSS).

[14] Organisation for Economic Co-operation and Development (OECD) (2000) Report on the Programme for International Student Assessment (PISA).

[15] International Association for the Evaluation of Educational Achievement (IEA) (2001) Report on Progress in International Reading Literacy Study (PIRLS).

The concerns regarding trust are not necessarily raised from outside of the teaching profession with the NAHT Commission on Assessment, (2014) concluding that there are tensions *within* the profession itself:

> The Commission heard from the majority of those submitting evidence that there was a lack of trust in teacher assessment at the present time. There is a worrying lack of trust in individual teacher-based assessment, which emanates from within the profession itself (p. 14).

Concern over levels of trust was also evident in a report produced by the examination regulator Ofqual in 2012 following a 1.5% fall in the percentage of GCSE English grades awarded and for some schools and colleges, unexpected differences in the grades awarded to their pupils. At the time, the GCSE English specification allowed for a 60% controlled assignment element.

> While no school that we interviewed considered that it was doing anything untoward in teaching and administering these GCSEs, many expressed concerns that other nearby schools were overstepping the boundaries of acceptable practice. It is clearly hard for teachers to maintain their own integrity when they believe that there is a widespread loss of integrity elsewhere (Ofqual, 2012, p. 11).

Some of this can be viewed as a broader issue of trust covering the curriculum and associated assessments which was more extensive prior to the introduction of national curriculum (Lightman, 2015*).* When the last version of the National Curriculum was introduced in 2014, it applied statutorily to maintained schools only; academies and free schools being free to develop their own curricula. However, I would argue that in practice most schools are likely to follow the same content and structure if only for the fact that statutory tests for key stage 1 and 2 are based on the National Curriculum and all categories of schools are required to take the assessments for accountability purposes. And although general qualifications such as the GCSE are not a statutory requirement, the syllabuses on which they are based stem from the National Curriculum.

In examining perceptions of trust in teachers' educational assessments within compulsory education, I would suggest the differing levels of concern are dependent on the particular context in which the assessment takes place. For example, the use of teacher assessment on a day-to-day basis tends to draw little attention from the wider community, but where teacher assessment has hitherto contributed to the assessment framework, the

Government has announced that key stage 1 and 2 teacher assessment will cease to be a statutory requirement by 2022 and 2018 respectively (DfE, 2017a). This decision was based on a change in policy to introduce a new baseline assessment from 2020 to be used as the basis of a new progress measure instead of the key stage 1 assessments. As teacher assessments were collected by the DfE but not used in accountability measures, the decision to end the collection of teacher assessment data at the end of key stage 2 was presented as a means of reducing teacher workload. However, Ofsted has raised concern over the quality and use of school assessment in several annual reports raising questions about the reliability, trust and use of teacher assessment (for example, Ofsted, 2007, p. 30; HMCI, 2011, p. 53; Ofsted 2015a, p. 37). The removal of National Curriculum levels in 2014 has prompted further concern around the generation and use of alternative assessment models used by schools and as a consequence, assessment became a key focus of Ofsted inspections in changes to the school inspection framework, September 2014 (Wilshaw, 2014, pp. 1-2).

The focus of Ofsted inspections is important here as it is recognised as being a key influence on the way schools manage and deliver their assessment practice (McIntosh, 2015; Peal, 2014; Ratcliffe, 2014; Harford, 2015). This has been of particular focus in primary schools since 2014 and needs to be viewed in the wider context of not only a revised Ofsted Inspection Framework, but the removal of National Curriculum levels of attainment and the Teacher Workload Survey Working Group Reports - each of which has reflected on the way schools assess their pupils, how this assessment information is recorded, is used to inform teaching and learning and the impact on teacher workload. This has resulted in an Ofsted clarification document referred to as the 'Ofsted Myth busting document' (Ofsted, 2016b) and guidance from the DfE in the form of Teacher Workload Working Group reports on marking and data management (DfE, 2016c & d). Each of these publications has provided advice to schools as to how they might organise assessment practice. However, the removal of levels has further exposed assessment as an area of concern in schools:

> The end of levels has revealed a worrying lack of knowledge and confidence in schools around the principles of assessment and the technical understanding required to enact them (McIntosh, 2015, p. 40).

This was also reported on in the Teacher Workload Group Report on data management (DfE, 2016) which highlighted anecdotal evidence that rather than schools taking the opportunities provided by the removal of levels to develop their own assessment systems,

they were inventing systems that merely mimic them (DfE, 2016). Having led courses for headteachers in approaches to educational assessment, I can only echo these concerns.

Since the introduction of National Curriculum assessments, it has been a statutory requirement for teachers to make an assessment of pupil's attainment. The initial vision of Task Group on Assessment and Testing (TGAT, op cit) was an aggregated assessment made up of teacher assessments and SATs. The mixed economy was viewed as having the potential for increasing teacher's curriculum knowledge and assessment skills, but for some these gains were outweighed by the detrimental effect on manageability (Whetton, 2009). Indeed, manageability issues have become a key feature of the educational landscape (Bew, 2011b, p. 43). In 2008, following failures in the delivery of Key Stage 2 tests, the then Secretary of State Ed Balls scrapped Key Stage 3 tests citing their ever decreasing relevance as parents and others focused more on the GCSE as a measure of performance standards in the secondary phase of education (11 to 16 age group). However, teacher assessments were collected from schools and published alongside test results up to 1999 when publication ceased despite the fact that it was still a statutory requirement to submit the assessment data to the DfE. Publication of teacher assessments was re-instated in 2010. However, teacher assessments at key stage 2 have never been given parity with the test outcomes until teacher assessment of writing replaced testing following recommendations of the Bew Report (2011b).

Teacher assessment has for many years been a feature of general qualifications and was a central feature of assessment schemes produced in the mid 1960s through the Certificate of Secondary Education and the late1980s in the General Certificate of Secondary Education. However, it has in turn attracted criticism for its impact on teacher workload, teaching to the test (or narrowing of the curriculum) and accusations of gaming the examination system drawing calls of cheating in coursework, plagiarism and teaching to the task in controlled assignments (Gove, 2013b). This has resulted in the removal of coursework from the majority of GCE and GCSE syllabuses.

2.11 Perceptions of trust in teachers' assessments in the mass media

Some of the general concern regarding trust in teacher assessments can be illustrated through a range of very public arguments between teachers, administrators, regulators and politicians, for example the rationale for the reduction in coursework in GCE A levels,

the grading of GCSE English in 2012 and concerns over the data showing almost half of pupils in England failed to meet the new tougher performance standards in reading, writing and mathematics at key stage 2 (Education Select Committee, 2017). The contention over assessment has been exacerbated by highly focused media coverage as noted by Murphy (2003) covering issues such as concerns over the appropriateness of GCE A level standards in 2002 and the delivery failure of National Curriculum test results in 2004 and 2008 in a system described by QCA's then Chief Executive as a 'cottage industry' (Sutherland, 2008, p. 11). Each summer on the release of test and examination results, the debate is reinvigorated by the now almost annual national tradition of stories over grade inflation,[16] declines in standards or poor marking. In broad terms, there is tension between the teaching profession and politicians. On the one hand the teaching profession argues that it is best placed to assess pupils, particularly in key stages 1 & 2 and in coursework in general qualifications: this is in spite of an expressed view that teacher assessment results in unreasonable workload as manifest in the disputes between teacher associations and the DfE in the mid 1990s and the outcome of teacher Workload Challenge (DfE, 2015b). On the other hand, politicians argue that performance data from external assessments (with ever decreasing teacher assessment) is essential if a true measure of the performance of pupils and schools is to be achieved (DfE, 2013a & b; DfE, 2014a).

It could be argued that a system using a mix of internal or school based assessment with the results of externally set and marked tests would provide a more reliable assessment system (TGAT, 1988). However, such an arrangement is not on the agenda for either teacher associations or education ministers with the Secretary of State recently stating that; "Exams are the best form of assessment that we can possibly have" (Williamson, 2020a). This may be driven by differences in approach between externally set standardised assessments and classroom based assessments, differences in the

---

[16] See for example: TES, Editorial: *What is grade inflation?* 2nd August 2019. Available on-line: https://www.tes.com/news/gcse-a-levels-grade-inflation Turner, C., and Kirk, A. 15th August, 2019.The Telegraph, *A-level results day 2019: Top grades fall to their lowest level in more than a decade* Available on-line: https://www.telegraph.co.uk/news/2019/08/15/a-level-results-day-2019-top-grades-fall-lowest-level-decade/ Miles, T. Evening Standard, 17th January 2002. Exam 'whistleblower' backed by GCSE Panel. Baker,E., Sutherland., & McGaw, B. 2002. Maintaining GCE A Level Standards. The findings of an independent panel of experts. London: Qualifications and Curriculum Authority (QCA)January 2002. Available on-line: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/605859/0102_Baker_et_al_Maintaining_GCE_A_level_standards_The_findings_of_an_independent_panel_of_experts.pdf

effectiveness of schools to carry out reliable assessments and concerns over teacher bias (Harlen, 2005a; Lamprianou and Christie, 2009; Kirkup and Twist, 2015). Despite the importance of the issue regarding the reliability of teacher assessment, few studies can be found in the literature, particularly regarding high-stakes assessment (Johnson, 2011 & 2013). There would also appear to be little appetite for such studies. In 2011, Ofqual invited tenders to conduct a study into the reliability of teachers' assessments (see Ofqual, 2011a); according to my discussion with officials, the call attracted one submission – which wasn't accepted.

2.12 Attempts to increase the reliability of teachers' assessments

In an attempt to increase the reliability of teacher assessment in high-stakes assessments, elaborate systems of moderation have been developed by examination boards and in the case of National Curriculum tests by Local Authorities. Moderation systems used in England rely on internally moderated marks being subjected to the scrutiny of an external assessor who judges the accuracy of the internally generated marks. This ability has been challenged and where there is disagreement, between the teacher and moderator, who is correct in applying the required standard has been questioned (Taylor, 1992). More recently the NAHT raised concerns that Local Authority moderators were diverging from the Standards and Testing Agency (STA) guidance for teacher assessment moderation at key stages 1 and 2 (NAHT, 2016). One approach to negate claimed deficiencies in teacher assessment is to remove the teachers from the assessment process entirely:

> …inspectors have noted worrying inconsistencies in teacher assessment at the end of Key Stage 1. In infant schools, for example, children are more likely to be assessed as reaching, or exceeding, the standards expected for their age than they are in all-through primary schools. Moreover, uneven moderation by local authorities of the work carried out by schools can lead to poor quality and unreliable assessment. For these reasons, I urge government to consider a return to external assessment at the end of Key Stage 1 (HMCI, 2013, p. 13).

However, a model based on written examinations and tests places limits on what can be assessed, for example speaking and listening in English. This is why over many years coursework has contributed to the overall assessment framework of a range of subjects in order to recognise for example competence in conducting experiments in science,

practical skills in Design & Technology and performance in drama. This is also a concern that needs to be recognised in relation to vocational qualifications where real world contexts or simulations are recognised as key elements of the validity of the qualifications where:

> Clearly, a range of assessment techniques is required to give assurance of what learners have achieved and of their competence, often within a workplace context (Huddleston, 2015, p. 28).

Yet the move to ever-greater regulation driven by fear of teacher influence in the assessment of vocational and educational training (VET) qualifications often undermined their intended purpose and "stifled innovation":

> The emphasis on the regulation of assessment within the English VET system has stifled innovation because too often alternative approaches have been viewed as tantamount to cheating. Assessment of practical activity has been regarded as too subjective and therefore unreliable and, ipso facto, of little educational merit (Huddleston, 2015, p. 33).

Again we see conflict. For example the Part One GNVQ in its early incarnation was designed to reflect the 'real world' needs of the qualification by the use of coursework, simulations and teacher assessment. However, it was gradually brought into line with the assessment regimes found in academic qualifications thus eroding its initial purpose – and leading to its ultimate demise (Huddleston, ibid). But coursework in general qualifications has been the subject of intense criticism and concern in particular the opportunities for plagiarism, teacher direction and support (Gove, 2013b, 2014a). In more recent times, examinations have been re-structured to reduce the more open-ended coursework with controlled assessments determined by awarding bodies. Controlled assessments are based on examination board defined assignments administered in schools under controlled conditions. The levels of control cover task setting, task taking and task marking with the level of control set as high as possible for the skills being assessed to assure validity but ensuring manageability for learners, schools and awarding organisations (Ofqual. 2015). Examples of control include a set period of supervised time or word limit. However attention has been drawn to the temptation for teachers to assist their pupils and issues such as teacher bias, either conscious or otherwise, and variation from teacher to teacher and school to school and; "…the application of different, sometimes personal, assessment criteria" used by teachers (Johnson, 2012, p. 370). Even

where markers are governed by tight marking schemes, there is much evidence to show the variance between markers for the same piece of work (Meadows and Billington, 2005; QCA, 2009; NFER, 2013). And even in countries with high-stakes teacher assessments there is little empirical study on the reliability of different tasks, the influence of the teacher or how teachers arrive at their judgments about student attainment (Johnson, 2012). Further, there is little study to reveal the nature, extent or quality of internal standardisation in schools and colleges. In 2006, QCA concluded that internal standardisation in schools for GCSE assessment was not widespread (QCA, 2006).

On the basis of such concerns Michael Gove, then Secretary of State for Education, announced changes to GCE and GCSE examinations in 2014. The Daily Mail reported:

> Coursework will be axed in almost all subjects – amid concerns it leads to cheating and wastes teaching time – in favour of written final exams (Clark, 2014).

The consequence of the announcement was that assessment in GCSEs and A-levels would be through external tests only, except where the essential skills for a subject could not be tested in an examination, for example in art and design and drama (Ofqual, 2016).

At the time of the Secretary of State's announcement in 2014, Ofqual chief executive Glenys Stacey (2014) stated:

> Non-exam assessments do not always test the skills they are meant to assess, they can disrupt classroom time better spent on teaching and learning and may provide limited evidence of performance across a group of students if they all get limited marks.

> Importantly, non-exam assessments can narrow the focus of what is taught and can be vulnerable to malpractice, meaning the playing field is not level for all students.

Despite successive governments' preference for external assessments and mass media pressure, examinations have their own limitations ranging from technical issues such as validity arguments, for example the alignment of the assessment with the content of the curriculum (Stringer, 2014), through to negative impacts on teaching such as narrowing the curriculum or 'teaching to the test' (Education Select Committee, 2017). Further, there is a lack of confidence in the teaching profession regarding the quality of external marking

undertaken by examination boards (HMC, 2012). From my experience regulating awarding bodies from 2000 to 2003 and accountability for the National Curriculum testing system between 2008 – 2011, I would contest that it is the multiple *uses* to which test and examination results are put that causes discontent amongst the teaching profession. The original purpose of National Curriculum tests was to measure attainment of the National Curriculum and for qualifications, a summative measure of performance for career progression. However over time these original purposes have grown to include other uses such as a measure of the performance of schools and teachers. This has led to an argument that a single test is unlikely to satisfy numerous demands:

> For example, KS2 tests in English, mathematics and science were originally designed to support inferences concerning the level that each pupil had attained by the end of the KS2 programme of study. So, from a level 4 in science, we infer that a pupil has attained the kind of knowledge, skill and understanding in science associated with the relevant level description. Although some pupils will receive incorrect levels, since no assessment is perfectly accurate, it would still be valid to interpret science test results in terms of a certain level of science attainment. This is the kind of inference that we would draw in order to use results for monitoring pupil progress in the subject, or for deciding the most appropriate science set to place a pupil in, upon transfer to year 7. However, for many other uses of results, we need to draw far more 'distant' inferences. (Newton, 2010, pp. 1-2).

2.13 Calls for increasing the status of teachers' assessments

The perceived limitations of tests have resulted in calls for the reinstatement of teacher assessment at key stage 2 and the scrapping of tests (McCann, 2016; Henshaw, 2016) and scrapping the GCSE in favour of school-based assessments covering areas such as teamwork and problem solving (Rethinking Assessment, 2020). Indeed there are arguments of potential benefits to a regime of teacher based educational assessment including assessment of areas not amenable to testing, a broader evidence base drawn over longer periods of time in the classroom, minimising the disruption and psychological effects caused by formal testing, reducing costs and building respect for the professionalism of teachers (Johnson, 2013). Wiliam (2001) adds support to the idea of greater use of teacher assessment that would provide higher reliability as long as the assessments are standardised and moderated to minimize bias. Whilst looking like a reasonable proposition, developing and administering such a system are costly and details of how this might be undertaken in practice are not provided by Wiliam (ibid). In general

terms, teachers' assessments in high-stakes tests do not attract the level of confidence given to other professions such as medicine or the law or in universities where teachers judge the performance of their own candidates (Gardner, 2007; Hargreaves, 1996). Gardner (ibid) suggests that this is due in part to the view that: "Many simply consider teachers to be "partial", that is, subject to bias" (p. 18. See also Lamprianou and Christie, 2009; Kirkup and Twist, 2015). It is of note that at the time of writing, following the cancellation of examinations (Williamson, 2020), the arrangements for awarding general qualifications in 2020 were initially devised to apply statistical moderation through an algorithm based on the previous performances of schools to centred assessed grades (CAGs) and rank ordering of students. Following widely held concern that the algorithm produced favourable bias to centres with small subject entries, many students attended public protests calling for trust in teachers assessments (see Figure 2:2) and following similar protests in Scotland, and Prime Minister Boris Johnson describing Ofqual's approach as the application of a "mutant algorithm" (Johnson, 2021), candidates in England were given the option of taking the highest of either the CAG or the statistically derived result. The impact of such a decision on 'fairness' to students in 2020 and beyond and the implications for the administration of future qualifications will no doubt be the subject of further data analysis and public inquiry in the autumn of 2020 (see for example the Parliamentary Education Select Committee, 2020).

Figure 2.2: Still from BBC Newsnight, 28th August 2020

But in light of the arguments above, are teachers capable of conducting reliable high-stakes assessments? Given some of the criticism regarding the extent of teachers' assessment knowledge, arguably very little is imparted to teachers in their initial training

phase or through continuous professional development. Other than sporadic training in the use of assessment schemes, for example Assessing Pupils' Progress (APP – see QCA 2010), there is little if any focus on developing the underlying theory and practice of assessment provided to teachers by government agencies. This results in observations that teachers have limited skills in assessment (Navarro, 2008).

This has led to calls by some for improvement in what has been described as *assessment literacy* (Gardner, 2007; Stiggins, 2002; Klenowski, 2012). However, assessment is still not seen as an integral part of the teaching process alongside content and instruction (Navarro, 2008). This situation is exacerbated by a view that only assessment (measurement) professionals using external assessments are capable of generating valid and reliable results (though this is often contested by the teaching associations) even though only a fraction of these assessments inform teaching and learning (Stiggins, 2008). Teachers carry out the remaining assessments with insufficient assessment literacy: "Typically, practitioners still are not being trained to assess accurately or to use assessments productively at any level" (ibid., p. 236). Stiggins (op cit) proposes a system in which teachers can develop their expertise through support from what he describes as measurement professionals utilising their greater understanding of assessment practice to develop valid and reliable approaches of direct use to classroom practitioners. From my experience as an adviser to the DfE and the NAHT, I have seen first-hand echoes of what Stiggins describes and the need for improved assessment practice amongst teachers: the reactions of teachers to the removal of National Curriculum attainment levels is a case in point in which many schools found it difficult to envisage new approaches to their assessment practice (see McIntosh, 2015). However, assessment methodology is subject to continuous change and would require substantial CPD to develop the existing teaching workforce and support newly qualified teachers. Wiliam and Thompson (2008) express the view that more recent CPD has improved since Fullan (1991) concluded that teacher professional development in assessment was frustratingly wasteful and made no significant impact in the classroom. This may well be the case, but in more recent times the reduced capacity of local authorities and constraints on the budgets of schools are likely to reduce access to good quality CPD.

A key feature raised in discussions about trust in teacher assessment is that of raising the knowledge and expertise of assessment through the initial preparation of teachers. Stiggins and Conklin (1992) noted how little was taught on assessment theory and practice in initial teacher training and on-going continuous professional development, a situation still in need of remedy (NAHT, 2014; McIntosh, 2015; Carter, 2015). The same level of

concern has been noted in other countries (Isaacs et al., 2013). However, it is not clear on what evidence base this conclusion has been reached. It is equally unclear what constitutes adequate content in assessment theory and practice for initial teacher preparation and CPD in England. Further, my concern here is that now much of initial teacher training in England is devolved to schools through training schemes such as Teach First and more recently Teaching Schools and teaching apprenticeships (see for example Teaching Schools Council, 2020; Ward, 2017), there is even more potential for variance in the content of teacher training with regards to assessment theory and practice; this calls for a better and shared understanding of what essential elements of high quality ITT content look like:

> …we have found considerable variability in ITT content across the system. We have identified what appear to be potentially significant gaps in a range of courses in areas such as subject knowledge development, subject-specific pedagogy, behaviour management, assessment and special educational needs and disabilities (SEND). We believe there may be a case for a better shared understanding of what the essential elements of good ITT content look like (Carter, 2015, p. 6).

The subsequent McIntosh Commission on Assessment without Levels (2015) supported Carter's view concluding that:

> This Commission agrees that the quality of assessment training is currently far too weak. This is not a new phenomenon. In 2007 a major research project (Hobson et al) into the experiences of Newly Qualified Teachers found that, at the end of their ITT courses, just 5% described "Knowledge/understanding of the principles of assessment for learning" as a strength of the teaching. By the end of their NQT year this had fallen to 2%. Moreover the NQTs highlighted "marking and assessment" as one of their top five reported professional development needs (p. 40).

There has however, been research in to teacher-training specifications in the USA to analyse the requirements for assessment training (Wolmut, 1988, cited in Stiggins and Conklin, 1992). The research revealed specific and relevant statements about assessment training in only nineteen out of fifty states in the USA; (Wolmut notes these results being identical to those in a parallel study ten years earlier.) Where training was evident, there were questions about the quality of that training (Wolmut, ibid). By comparison, there is

little indication in the academic literature of the educational assessment content drawn from the various prospectuses of university teacher education course in the UK. Furthermore the current model for initial teacher training makes general access to the content of training more problematic. This would indicate that there is a need to develop a greater understanding of the reliability of teacher assessment and to develop improved training in assessment understanding and practice.

> There is a generally acknowledged need to address the issues more directly, by (re)developing the assessment skills of classroom teachers, and building their confidence for making assessment judgements of their own pupils (Johnson, 2012, p. 40).

The Teachers' Standards provide high-level statements of the attributes a teacher should possess. With regards to assessment, Standard Six 'Make accurate and productive use of assessment' refers[17] (DfE, 2013c). However, the Standards are high-level statements and denote the "…minimum level of practice expected of trainees and teachers from the point of being awarded qualified teacher status" (p. 3). The details of how this manifests in teacher preparation courses is open to interpretation and until teachers can show through research based evidence how their knowledge and understanding of assessment assure validity and reliability, doubts over the efficacy, and therefore a lack of trust in teacher assessment will remain (Hargreaves, 1996).

Further, the calls for assessment literacy have not identified an agreed body of knowledge, understanding and skills (Ingvarson and Rowley, 2017) or that a deep understanding of assessment theory and practice is viewed as an essential area of study and development for teachers. Others have produced lists of what might be required without full explanation as to why (for example Gardner, 2007) and although the Teachers' Standards are a positive step in presenting assessment as a key area of competence for teachers, without exemplification or further guidance they remain open to interpretation. In response to the Carter Review (Ibid), A Framework of Core Content for initial teacher training (ITT) was

---

[17] Standard 6. Make accurate and productive use of assessment
- know and understand how to assess the relevant subject and curriculum areas, including statutory assessment requirements
- make use of formative and summative assessment to secure pupils' progress
- use relevant data to monitor progress, set targets, and plan subsequent lessons
- give pupils regular feedback, both orally and through accurate marking, and encourage pupils to respond to the feedback.

published by the DfE (2016a) and agreeing with Carter's conclusion "…that there is significant variability in ITT courses" (p. 5), the Framework presented more detailed content in support of the Teachers' Standards; for example in relation to Standard 6:

> **Providers should** (original emphasis) ensure that trainees are fully conversant with the fundamental principles of assessment and testing, including the difference between formative and summative assessment; bias, reliability and validity (p. 18).

More recent publications from the DfE have also built on and extended the details contained in the Teachers' Standards. The Early Career Framework (DfE, 2019b) recognises that areas covered in initial teacher training require understanding at a greater depth (p. 5) and offers extended support over the first two years of teaching recognising that:

> Just as with other esteemed professions like medicine and law, teachers in the first years of their career require high quality, structured support in order to begin the journey towards becoming an expert (p. 4).

The roll out of the Framework, supported by part of a £42 million Teacher Development Premium, will begin in September 2020 in four regions across the north of England[18] to provide an evidence base before wider provision from September 2021. The Initial teacher training (ITT): core content framework (DfE, 2019c) has provided further clarity on the Teachers' Standards and the relationship with the Early Career Framework and sets out a clear vision for a three or more years structured package of support including mentoring from "expert colleagues"[19] and a regular review of content in order to "…draw on the best available evidence" (p. 3). The Framework also includes a useful reference section drawing on the research literature in support of each of the standards. However, the Framework is dependent on access to expert colleagues with appropriate knowledge and understanding of assessment theory and practice; attributes that are currently in short supply. Further, in acknowledging the complexity of the process of becoming a teacher, the Framework does not provide detail on what might constitute core knowledge and understanding of educational assessment noting that:

---

[18] The 2020 roll out will cover the North East, Bradford, Doncaster and Greater Manchester.

[19] Expert colleagues are defined as: "Professional colleagues, including experienced and effective teachers, subject specialists, mentors, lecturers and tutors (p. 5).

…it remains for individual providers to design curricula appropriate for the subject, phase and age range that the trainees will be teaching. It will be crucial for providers to ensure trainees have adequately covered any foundational knowledge and skill that is pre-requisite for the content defined in this framework (p. 4).

Gaining insight into the curriculum offered by ITT providers in England through reference to the academic research literature has proved a challenge, an issue identified by researchers at the Sheffield Institute of Education (2019) in a research report produced for Ofsted that concluded: "This lack of research is in itself a finding worthy of note" (p. 6). The research was commissioned to support Ofsted in developing a new inspection framework for initial teacher education (ITE) to complement the revised inspection framework for schools introduced in September 2019. In the HMCI commentary (October 2019), Amanda Spielman noted:

Much like our previous education framework, the current ITE framework places a lot of emphasis on data. For example, it focuses on employment rates, completion rates and individual trainees' effectiveness. Consequently, inspectors have put relatively little weight on what trainees are taught or how well the centre-based and school-based training is combined into a coherent package of learning. The reliance on other outcome measures may, therefore, cover up some kinds of weakness across partnerships, or even mask strengths (p. 3).

Spielman also spoke of the continued diversity of ITE provision since the ITE framework began and large increase in school-led routes opening since 2015, with the growth in small institutions offering ITT to an average of 50 trainees per year (ibid). Importantly, Spielman (op cit) cited the Sheffield Institute of Education research finding that despite the importance of mentors and professional tutors:

…inconsistency in the quality of mentors and placements, often across the same partnership, was a regular concern identified by those trainees who felt that they had received a poor training experience (p. 8).

In a system in which greater reliance is placed on school based ITT, the provision of highly skilled mentors becomes an essential element. In a presentation to the Junior School Collaboration (JUSCO) in February 2019, Professor Sam Twiselton highlighted three types of teacher: task mangers – focused on keeping the class busy, on task with no reference to learning: curriculum deliverers: where the curriculum is a goal in itself. It is

hard to give a reason why the learning is important: and concept/skill builders; the concepts and skills are key. Tasks (are) only a vehicle for learning. The main goal lies beyond the lesson - transferable and transformative learning (see also, Twiselton 2000, 2004 and 2007). A key outcome of Twiselton's research is the conclusion that newly qualified teachers are more likely to become or remain concept/skill builders if they are exposed to contexts that are conducive to articulation, so they are able to compare/contrast/critique approaches:

> Context and culture matters: Learning organisations are most effective when they create a critically reflective community of practice that is constantly relating practice and curriculum to the bigger picture at every level (2019).

This reinforces the need for skilled mentors with deep knowledge of pedagogy and assessment theory and practice and a deep understanding of how to develop the professional knowledge and skills required of teachers (see Ofsted, 2020).

The Carter Review (op cit) compiled an analysis of course information from 145 ITT courses which provides a rare insight into the components of the ITT curriculum on offer, but particular deficiencies in provision were exposed around educational assessment content. For example the analysis looked for evidence that: *"Trainees are given explicit sessions on key concepts in assessment such as validity, reliability, standardisation, norm referencing, criterion referencing"* (DfE 2015c): however, only 17 of the 145 courses provided evidence of meeting this aspect. In terms of; *"Trainees are taught how to use pupil data to plan for individuals and groups, and at a whole school level"*, 69 courses showed evidence of meeting this aspect, but only 28 of the 145 provided trainees with *"…some training in statistics to support use of pupil data"* (ibid).

The Sheffield Hallam research suggested further areas of research including consideration of the advantages and disadvantages of a tightly structured curriculum for ITE in a diverse market place and the potential need for differentiation across subjects and phases of education. However, it is not clear who should take the lead in developing assessment capability. Ministers place emphasis on external standardised assessments *as* "…government simply does not trust teachers' judgements of their students" (Gardner, 2007, p. 20). Conversely, teacher associations contest the reliance on external assessment but have not presented a united view on what an alternative model of assessment would look like or how it would be justified in terms of validity or reliability. I share the view expressed by Johnson (2013) that the emphasis on external assessments

has undermined confidence in the professionalism of teachers and of equal concern, has resulted in the loss of assessment skills (see also Dunford, 2020). This is not simply an issue of making a contribution to high-stakes assessments, but a concern that assessment becomes marginalised or not understood as an essential element of the teaching and learning process. Looney et al (2017) present a further view that assessment literacy is not simply a list of appropriate or agreed knowledge and skills, but a more complex structure of emotions and beliefs influenced by past experiences of being assessed and perception of the social consequences for teachers and pupils through assessment outcomes.

2.14 Chapter 2 summary

The development of general qualifications and National Curriculum assessments in the 5 to 19 age range since the mid-twentieth century have raised debate about the role of teacher assessment in what can now be described as high-stakes assessments. It raises issues about the definition and purpose of assessment and the need for more precision in the use of language in the context of educational assessment. It also raises questions about the degree of trust in the reliability of teachers' assessment and their level of expertise. There are claims and counter claims about these issues, but this points to the need for a better understanding of what constitutes the essential content of initial teacher training and continuous professional development with regards to educational assessment and further understanding of teachers' perceptions about the social consequences of the uses of assessment. The opposing views about the place and trustworthiness of teacher assessment would benefit from clearer evidence to support these claims. In the context of high-stakes assessments, suspicions over the efficacy of teachers' assessments are likely to continue as is the reliance on externally set assessments. This situation constrains the assessment regime to externally set and marked written tests that limit the focus to elements of the taught curriculum that can be assessed through such tests.

In the light of the literature and other sources covered above, issues of trust are of central importance to understand the place of teachers' assessments in the English education system. Moreover, whilst Carter (2015), McIntosh (2015) and the subsequent ITT Core Content and Early Career Framework (DfE,2019c) provide an outline description of what might constitute a reasonable foundation for what teachers need to know about assessment, there is little empirical evidence of the detail of course coverage in current ITE provision. Chapters 5 to 8 therefore look to identify in more detail current ITT course content and the perceptions of teachers and Key Influencers on the adequacy of teacher

knowledge and understanding in this field. If teachers could demonstrate high capability in assessment theory and practice, the possibilities for more valid assessments could be extended.

The literature review has played an essential role in developing the focus of the key research questions used in this study (see Chapter 3, Section 3.3). As the literature has been explored, the research questions have unfolded and then evolved in order to shine a light on areas underrepresented or missing in the body of works available to researchers in the field of educational assessment, a process described by Alvesson and Sandberg (2011) as; "gap spotting" (p. 248). Examples highlighted in this study include the paucity of available detail on the content of initial teacher education courses, or a more nuanced understanding of the perceptions of trust amongst teachers, policy makers, policy influencers, and academics. But this research project is not restricted to filling gaps alone but one that challenges or 'problematises' assumptions as a means of generating novel research questions; "…that may facilitate the development of more interesting and influential theories" (Alvesson and Sandberg, op cit., p. 267). This notion of problematisation is linked to the approach of Michel Foucault as applied in detail in Chapter 4 as an; "…endeavour to know how and to what extent it might be possible to think differently, instead of what is already known" (Foucault, 1985, p. 9, cited in Alvesson and Sandberg, p. 253). Indeed, the title of the thesis has itself evolved since the initial exploratory foray into the academic literature and wider forms of documentation from one concerned with the *development* of teacher-based assessments over time to one seeking to gain an understanding of *change* in a way that holds no assumption of progress or advancement in efficacy as the term 'development' might suggest (see Section 4.1).

Chapter 3: Methodology

3.1 Chapter over-view

This chapter presents the research questions and turns to the methodological approach used in this study, the research design and ethical considerations. The first section of this chapter discusses the philosophical underpinnings of the methodological approach taken in relation to this research project. The second section presents the research questions followed by the research design with five further sections covering the methods of data collection each designed to address aspects of the research questions:

1. a questionnaire survey with teachers;
2. a questionnaire survey with initial teacher educators;
3. interviews with teachers;
4. interviews with Key Influencers (see page 22 footnote for an explanation of this term);
5. documentary sources – primary and secondary texts (Silverman, 2006).

The axiology and ethical considerations conclude this chapter.

3.2 Philosophical underpinnings of this study

The methodological position of this study adopts a mixed methods approach with the assumption that this form of inquiry yields additional insights beyond information provided by either a quantitative or qualitative study alone (Creswell and Creswell, 2018). In considering the methodological position to this research project, I have drawn on the available literature relating to the philosophical underpinnings of educational research. The purpose of presenting the philosophical position is to provide clarity on the chosen conceptual framework or paradigm that has influenced the selection of the research questions and the methods used to study them (Shannon-Baker, 2016). The term 'paradigm' has been adapted from the work of Kuhn (Morgan, 2007; Hall, 2013; Shannon-Baker, ibid) and is used widely in the research literature:

> …but as Morgan (2007) points out it has been given at least four different meanings in the literature. These have been identified as a world view; an epistemological stance; as shared beliefs among a community of researchers and as model examples of research. Although Morgan (2007) argues that the third of these is

closest to what Kuhn defined as a paradigm he does acknowledge that the second meaning, namely a paradigm as an epistemological stance has been the most commonly used meaning in discussions of social science methodology (Hall, 2013, p. 3).

However, according to Hall (ibid), 'world-view', has been adopted by some of the major writers in the field including Teddlie & Tashakkori (2009, p. 84) and Creswell & Plano Clark (2007, p. 21). But not all researchers support the concept of paradigms viewing them as having the potential to marginalise other beliefs or force researchers to 'buy into a set of beliefs' (Maxwell, 2011, cited in Shannon-Baker, ibid). Although Shannon-Baker accepts these points, the conscious use of a paradigm as a means of a framework to guide researchers during the inquiry phase is advocated. Further:

> …when a researcher provides information about their beliefs, it gives their audience a better understanding of the potential influences on the research. Therefore, I approach the paradigm issue asking not whether paradigms are useful but how paradigms can be intentionally used; which paradigm to operationalize then is at the researcher's discretion (Shannon-Baker, p. 321).

Even where the use of paradigms is supported, there has been further tension between what has been described as the traditional positivist and interpretive movements (Hall, 2013). An approach to research from the positivist view "…takes the ontological position that an external objective reality of phenomena exists independently of our interpretations of it" (Wilson, 2017, p. 3) whereas the interpretivist view denies, "…the existence of an external reality that is independent of human cognition and interpretation" (Ibid., p. 3). Over time these positions or paradigms have been presented by purists as extremes of incompatible practice due to the inherent differences in the underlying philosophies leading to 'paradigm wars' (Tashakkori & Teddlie, 1998, p. 11). However, others have argued that there are sufficient similarities in the fundamental values of both paradigms, for example the fallibility of knowledge and that data can be explained by different theories to, "form an enduring partnership" between the two paradigms (Reichardt & Rallis, 1994, cited in Tashakkori and Teddlie ibid., p. 12). Such a viewpoint underpins the growth of a move away from a monomethods approach to one that allows for a mixed methods approach as a 'legitimate methodological choice' (Cameron and Molina-Azorin, 2011, p. 14). This approach has been described as pragmatism, which elevates the importance of the research questions above the method or underlying paradigm (Tashakkori & Teddlie, ibid). The 'enduring partnership' is manifest in the mixed methods research approach

which according to Hanson et al (2005) has become an 'increasingly popular' and legitimate approach in the social sciences:

> In the social sciences at large, mixed methods research has become increasingly popular and may be considered a legitimate, stand-alone research design (Creswell, 2002, 2003; Greene, et al.,1989; Tashakkori & Teddlie, 1998, 2003). When both quantitative and qualitative data are included in a study, researchers may enrich their results in ways that one form of data does not allow (Brewer & Hunter, 1989; Tashakkori & Teddlie, 1998) (p. 224).

Mixed methods can also be viewed as a combination of elements of qualitative and quantitative research approaches in promoting 'broad purposes of breadth and depth of understanding and corroboration' (Johnson et al. 2007. p. 123). My view is that mixing the analysis of data drawn from different approaches will provide stronger answers to the research questions than might be generated from either a quantitative or qualitative approach alone (Creswell and Creswell, 2018). However, others have argued that mixed methods are not possible due to the incompatibility of the paradigms underlying them (Guba and Lincoln, 1994, cited in Hall, 2014), In considering the history of the mixed methods approach, Hanson et al (ibid) have commented on two persistent issues; the paradigm-method fit issue and the "best" paradigm issue:

> The paradigm-method fit issue relates to the question "Do philosophical paradigms (e.g., postpositivism, constructivism) and research methods *have* to fit together?" Some researchers have argued, for example, that a postpositivist philosophical paradigm, or worldview, could be combined only with quantitative methods and that a naturalistic worldview could be combined only with qualitative methods. This issue has been referred to as the "paradigm debate" (Reichardt & Rallis, 1994). From this perspective, mixed methods research was viewed as untenable (i.e., incommensurable or incompatible) because certain paradigms and methods could not "fit" together legitimately (Smith, 1983). Reichardt and Cook (1979) countered this viewpoint, however, by suggesting that different philosophical paradigms and methods were compatible. In their article, they argued that paradigms and methods are not inherently linked, citing a variety of examples to support their position (p. 226).

The second of Hanson et al's issues, the best paradigm issue, raises the question: "What philosophical paradigm is the *best* foundation for mixed methods research?" (Ibid p. 226).

This issue again attracts multiple perspectives including a dialectical approach whereby using competing paradigms gives rise to contradictory ideas and contested arguments that may not be reconciled (Greene & Caracelli, 1997, 2003, cited in Hanson et al., ibid). Hanson et al state:

> Such oppositions reflect different ways of making knowledge claims, and we advocate for honoring (sic) and respecting the different paradigmatic perspectives that researchers bring to bear on a study (p. 226).

And conclude that:

> The best paradigm is determined by the researcher and the research problem - not by the method (p. 226).

In terms of this study, my motivation in selecting a mixed methods methodological approach is driven by the desire to answer the research questions. This approach assists in answering questions that may not be answered by quantitative or qualitative approaches alone and gives access to a broader set of research tools (Creswell and Plano Clark, 2011). Within the mixed methods approach the primacy of the research questions as drivers of the chosen methodology is frequently described as 'pragmatism'. Pragmatism, as described by Hanson et el (ibid):

> …draws on many ideas including using "what works," using diverse approaches, and valuing both objective and subjective knowledge (p. 227).

Denscombe (2014) proffers that:

> Pragmatism is generally regarded as the philosophical partner for the mixed methods approach…social research should not be judged by how neatly it fits with the quantitative paradigm (positivism) or with the premises of the competing camp – the qualitative paradigm (interpretivism). Decisions about which method to use should, instead, be based on how useful the methods are for addressing a particular question, issue or problem that is being investigated (p. 158).

Creswell and Creswell (2018) drawing on Cherrryholmes (1992) and Morgan (2007), note that pragmatism provides a philosophical basis for research (2018, p. 10) which according to Hall (2013) has gained; "considerable support as a stance for mixed methods

researchers" (p. 6). However, pragmatism has been challenged on the grounds of the difficulty of determining what 'works' because of the assumption that the usefulness of mixed methods design can be known before the project is completed. The success or otherwise of a mixed methods approach can only be gauged when the project is completed (Hall, ibid). The reader can therefore make that decision. But pragmatism is not an 'anything goes' type of approach; its rigour is driven by the need to answer the research questions and provide useful, practicable outcomes (Cohen et al, 2018). This research project aims to address criticisms of the pre-occupation amongst educational researchers with disputes over methodology (Hargreaves, 1996) rather than generating research of direct relevance and accessibility to practitioners. This was a theme developed further in 1997 by the then Her Majesty's Chief Inspector of Schools (HMCI) Chris Woodhead who commissioned a report by Tooley and Darby (1998) that echoed Hargreaves' critique. However, the report was met with criticism for its own partisan nature and flawed methodology (Wellington, 2015). This was followed by the Hillage Report (1998) commissioned by the Department of Education and Employment (DfEE) and although more widely respected, held some common ground with Hargreaves and Woodhead. This relatively short period of public critique of educational research, though not applicable to all research, did result in self-reflection and more emphasis on its connection with policy makers and practitioners (Wellington, ibid). However, some would argue that not much has changed in that policy makers use research to; "…legitimise policy rather than to inform it" (Clegg, 2005, p. 418, cited in Wellington, ibid. See also Badley, 2003; Wilson, 2014; EPI, 2020). This impacts this study in that it may result in challenges to policy makers, but conversely support the work of institutions like the Chartered College of Teachers (CCT), the Chartered Institute of Educational Assessors (CIEA) and initial teacher educators with an expressed aim of equipping teachers with access to high quality research (CCT, 2017). The methodological strategies used in this study are undertaken to 'offer evidence to inform judgements, not techniques that provide guaranteed truth or completeness' (Hammersley, 2005, p. 12, cited in Brannen, 2005).

The decision to present an argument for any chosen methodological approach is an important one. First, and in face of claims that mixed methods are not possible due to the incompatibility of the paradigms underlying them (for example Guba and Lincoln, 1994), it becomes necessary to defend the chosen approach. Second, articulating the methodological underpinning is intended to help the reader to understand the potential influences on the research and third, it acts as a guide and a constant check that the research stays true to the project aims and questions (Shannon-Baker, ibid). The following

sections provide details of the research questions, the research design and the approaches to data collection taken in the light of the stated methodological approach.

3.3 The research questions

As discussed in Chapter 1 of this study, the role of teachers in educational assessment has changed over the period 1943 to the present day. These changes have impacted high and low stakes educational assessments and raised issues over the validity and reliability of teachers' assessments and questions of trust in teachers' capacity to deliver effective assessments. These issues are borne out in the literature review presented in Chapter 2.

According to Johnson (2012), if teachers are to be involved in a system of high-stakes testing designed to assess a broader range of attributes than can be gained by an externally set test, there is a need to develop a greater understanding of the reliability of teacher assessment and to develop improved training in assessment understanding and practice to build confidence in their judgements. There is also an argument that involving teachers in summative assessment promotes teacher professionalism and expertise (Harlen, 2004). Given the claims and counter claims about the reliability of teachers' assessments and the level of assessment training in teacher preparation and professional development programmes, this study focuses on the relationship of three areas: first, the role played by teachers in high-stakes assessment over time; second, perceptions of educational assessment and the role of teachers; and third, the extent and perceived quality of current provision in educational assessment theory and practice found in current ITT courses. This study therefore focuses on three key research questions:

3.3.1. Question 1: How has the role of teachers in educational assessment changed from the Norwood Review (1943) to the present day?

This question draws on the changing role of teachers in educational assessment since the mid twentieth century to the present day. This point in time has been selected as the Norwood Review of the curriculum and assessment (1943) raised questions about external examinations that still resonate today and was shortly followed by the Education Act (1944) which; "…replaced almost all previous education legislation and set the framework for the post-war education system in England and Wales" (Gillard, 2011, Chapter 5). It examines the educational, social

and political contexts through which educational assessment has changed and the impact on the role of teachers. Question 1 is the focus of Chapter 4.

3.3.2. Question 2: How do teachers and Key Influencers perceive educational assessment and the role played by teachers in the English education system?

This question examines teachers' perceptions of educational assessment formed through their early experiences of assessment as pupils and as students. The views of Key Influencers are considered as a means of understanding the broader perceptions held by leading academics in the field of educational assessment, education policy makers and regulators of the role played by teachers in the English education system. This question considers educational assessment in high and low stakes contexts. The data used to address Question 2 are presented in Chapters 6 to 8.

3.3.3. Question 3: What does the ITT curriculum in England offer in terms of assessment theory and practice?

The intended outcome of this question is to gain an insight into the content provided in ITT courses and the views of teachers, teacher educators and Key Influencers on the extent to which it presents an adequate grounding for teachers entering the teaching profession in the maintained sector of the English education system. It will consider the perceived quality of instruction in educational assessment in initial teacher preparation at a time of increased diversification of routes into teaching and the way educational assessment is used in the English education system. The data used to address Question 2 are presented in Chapters 5 to 8.

In answering the research questions, this study draws on a mixture of quantitative and qualitative devices and uses a diverse range of sources to generate data as presented in the research questions and data collection overview (Table 3.1 below). The research design is presented in the next section.

Table 3.1: Research questions and the data collection overview

Research Questions	Data collection	Source/Participants
1. How has the role of teachers in high-stakes educational assessment changed from the Norwood Review (1943) to the present day?	Documentary analysis of primary and secondary sources to include: acts of parliament; reports of commissions and select committees; white papers and green/consultation papers; reports from national bodies; ministerial speeches; newspaper reports; academic literature. Interviews – teachers and Key Influencers.	National archives; libraries; Hansard; websites; interviews with experts and policy makers in the field of educational assessment; Interviews with teachers; regulatory bodies; national bodies e.g., teacher associations and awarding bodies.
2. How do teachers and Key Influencers perceive educational assessment and the role played by teachers in the English education system?	Survey questionnaires. Documentary analysis. Interviews. Survey and interview analysis.	ITT providers; practicing teachers; experts in the field of educational assessment; policy makers; regulatory bodies.
3. What does the ITT curriculum in England offer in terms of assessment theory and practice?	Survey questionnaires. Documentary analysis. Interviews with Key Influencers including international assessment experts in the field of assessment, politicians and teachers. Academic literature; policy documents. Analysis and triangulation of all data sources.	Practicing teachers; experts in the field of educational assessment; policy makers; regulatory bodies, awarding bodies.

3.4 The research design

"Research designs are procedures for collecting, analysing, interpreting and reporting data in research studies" (Creswell & Plano Clarke, 2011. p. 53). Creswell and Plano Clarke usefully present what they describe as six major 'prototypical designs' used in mixed

methods approaches. These designs were further reduced to three core or 'primary designs' by Creswell and Creswell (2018); convergent mixed methods design, explanatory sequential and exploratory sequential design. This study adopts a fixed mixed methods approach whereby the procedures are predetermined and executed as planned (Creswell & Plano Clarke, 2011). There is a wide range of typology-based approaches to mixed methods design that can be adopted or modified dependent on the research purpose and questions. Having considered a range of typological designs, I have taken the option of what has been described by Creswell and Plano Clarke (ibid) as a dynamic approach. This approach allows for the adoption of interrelating components from given designs. Creswell and Plano Clarke advocate the use of a typology-based approach particularly for new researchers, but point out that a typology-based approach should not be used like a; "…cookbook recipe but instead use it as a guiding framework to help inform choice" (ibid., p. 60). Using an adaptation of Creswell and Creswell (2018) and Creswell and Plano Clarke's (2011) primary designs, this study adopts a hybrid or dynamic approach (see also Hall and Howard, 2008) bridging elements of the convergent and explanatory sequential approaches (Figure 3.1). The convergent approach is characterised by the merging of qualitative and quantitative data that are collected concurrently then integrated at the point of interpretation (see Chapter 9, Discussion).

In this study, the main emphasis is on a convergent approach but with an early high-level analysis of the data generated by the teacher and ITT provider questionnaire surveys. An early high-level data analysis of the teacher surveys was conducted for two reasons. The first was to identify potential interviewees from the teacher survey for the qualitative face-to-face interviews and to gain an understanding of their survey responses that might benefit from further and more detailed exploration through an interview. Second, it provided a general view of emerging themes from the teacher and ITT provider surveys that again could be explored further with the interviewees from the teacher and key Influencer interviews. This initial high-level data inspection stage is highlighted at arrow A on Figure 3.1.

Figure 3.1: The research design. Adapted from Creswell & Creswell (2018 p. 218)
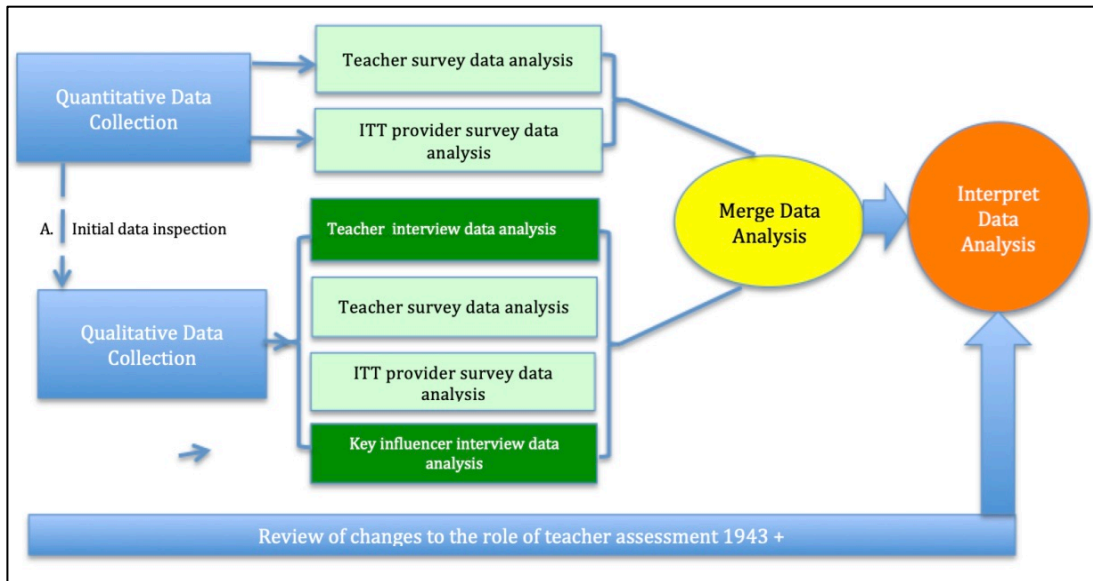
3.4.1 The advantages and challenges of the chosen research approach

The philosophical position underpinning the use of a mixed approach is discussed in Section 3.2 along with the stated goal of combining qualitative and quantitative methods in what I argue is; "…an important way of looking at the social and educational world that is informed by a pragmatic paradigm of practicality in answering research purposes and research question" (Cohen et al., 2018, p. 48-49). However, the mixed methods approach brought to bear a range of advantages and challenges in this research project. I would argue that drawing on quantitative and qualitative techniques provides a method of compensating for inherent weaknesses or gaps in the application of either method alone. For example, although the surveys used in this project provided data on the views of a larger number of participants who had the option to remain anonymous, they did not allow for the deeper exploration of views and perceptions that became accessible through the use of semi-structured interviews with practitioners and Key Influencers. In addition, the use of survey data highlighted particular perceptions that could then be interrogated further drawing on the contexts in which interviewees reside. The use of a series of repeated interviews (Seidman, 2006) with teacher participants proved particularly useful in permitting time for reflection and more meaningful or expanded dialogue (see Section 3.10.1). And interviews with participants who hold positions of influence in policy, regulation, academia and administration provided the facility to gain and interrogate perspectives on the place of educational assessment in England from a range of

perspectives. An equally time-consuming but worthy exercise in considering the role of teachers over an extensive period of time, was the facility to visit archives such as those at the Houses of Parliament and local authority archives held in libraries, providing the opportunity to examine documentary materials not readily accessible through other means. Taken in combination these approaches allow for the triangulation of data, and this adds weight to the overall validity of this study.

However, undertaking such a range of approaches and activities does present challenge, particularly to a lone or single researcher. At one level, it places the need to develop and utilise a broader range of research skills, such as writing surveys that are technically defensible in terms of construct validity for example, and considerations of question or item bias. Further, there is a requirement to develop and apply the skills of quantitative analysis. In addition, the qualitative techniques require the skills of writing interview schedules, an understanding of researcher positionality and the potential for bias (see Chapter 1). Demands are also made in terms of developing robust coding frames, data cleansing and the skills of identifying key aspects and themes generated by the data from a range of sources. The process of arranging interviews brings other demands, not only in gaining access to participants with competing demands on their time and their own competing priorities, but in terms of timing, travelling, and recording, transcribing and analysing the data. Making actual visits to archives adds further time management issues, but also the advantage of viewing materials that do not necessarily appear the academic and historical literature or electronic sources but do provide rich information on prevailing discourses.

However, all of these challenges are surmountable and having had a background in the English education system at a variety of levels including policy development, I have direct experience of managing large scale consultations employing quantitative and qualitative methods – albeit with the support of experts in the field. I also have the privilege of access to a range of players in the English education system which made the arranging of interviews for example relatively straight forward. It must also be acknowledged that access to courses and modules available at the University of Leeds, in particular those covering research methodology, using SPSS, and coding and data analysis tools has played an equally important part in providing the appropriate knowledge and skills required of post-graduate research.

But fundamentally, the choice of mixed methods research is a reflection of my philosophical position of pragmatism in that it is 'practical' and allows the researcher to be

'free' and unfettered; "…rather than the typical association of certain paradigms with quantitative research and others for qualitative research" (Cohen, et al., 2018, p. 13).The adopted approach is driven by the desire to produce valid answers to the research questions aided by the application of educational assessment techniques, which in their simplest form can be expressed as what exactly is it that we want to know, why - then how best can this be achieved. This, I believe, leads to equally practical and useful outcomes of this project.

In practice, the overall approach proved to be iterative as the components provided the opportunity to draw on emerging issues that could be explored further through forays into further literature and in the interviews and cross-checked in the final data analysis. This reflects what Thomas (2013) refers to as a 'recursive' rather than a 'linear plan' (Figure 3.2) in the social sciences whereby the project does not run in a linear fashion but at points stops, is re-planned, changed and even started again (ibid, p. 19). Whilst this project did not 'start again' it did evolve in a non-linear fashion as each stage highlighted un-thought implications for other stages even to the point of refining the research questions, a feature discussed earlier in Chapter 2, Section 2.14.

Figure 3.2: A linear or recursive plan? (Thomas, 2013, p. 19)
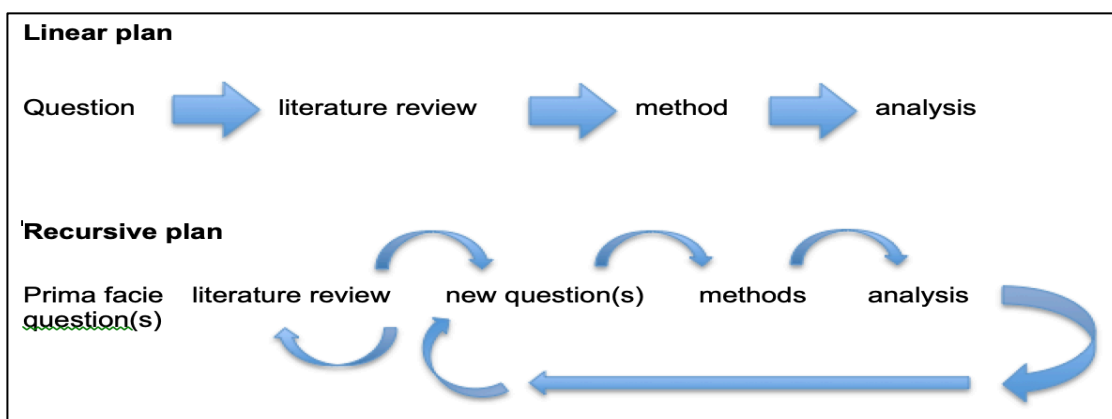
3.5 Data collection and the influence of Michel Foucault

As noted above, a mixed methods approach is used in this project that is designed to draw on the relative strengths of quantitative and qualitative methods in order to address the research questions. The initial phase involved a review of the academic and other literature pertaining to the role of teachers in educational assessment. A key component of this study takes a Foucaultian perspective (see for example Ball, 2013) on the place of

teachers' assessments in the English education system over time as raised by research question 1 of this study: this is covered in detail in Chapter 4. Central to this perspective is that over the given period of time of interest to this study, from 1943 to the present, there is no unquestioning acceptance that the role of teachers in educational assessment has 'developed' in the sense that it has improved, advanced or matured (Cambridge Dictionary, 2019. Oxford Dictionary, 2019); rather the premise is taken that it has changed. Secondly, it examines the role of teachers in the utilisation of educational assessment as a facet of power relations. This element of the study draws on a review of the academic literature together with a review of primary and secondary texts (Silverman, 2006) including policy documents such as white and green papers, parliamentary acts, ministerial speeches and select committee papers along with documentary evidence of discourse such as papers produced by teacher associations, newspapers and journals. The research included reference to archived materials, including those held by teacher associations, for example the NUT; the NFER Archive service, Edinburgh University archive service along with visits to the Houses of Parliament and Kirklees Library Service. The aim is to develop an account of assessment by teachers over time across the compulsory phase of education through what Foucault regards as the prevailing ideologies, archaeology, and transitions in thinking genealogy (Garland, 2014). This goes beyond understanding the phenomena at the centre of this research project through either a positivist or interpretive lens but rather looks to form an appreciation of the political and ideological contexts in which educational assessment takes place. This type of approach can be set within the frame of a critique of educational research that aims to be transformative (Cohen et al, 2007) in that it not only seeks to understand current provision but to promote change, in this case in the interest of improving teaching and learning in schools and increased validity and reliability of teachers' assessments in high and low stakes assessments. Drawing on the work of Foucault as a guide, the overall aim is to develop a cohesive understanding of the changes to teachers as assessors through shifts in what might be described as 'repressive practices' to 'practices of normalisation' (Beasley cited in Lazaroiu, 2013). This genealogical approach departs from conventional histography in favour of what has been described as Foucault's writing of a 'history of the present' (Garland, 2014).

For research questions 2 and 3 of this study, the interpretivist (Cohen, et al., 2018; Duberley et al., 2012.) viewpoint underpins the chosen approach in that the views and perceptions of different stakeholders and practitioners are sought through dialogue around the role of teachers in the process of educational assessment and its future direction in the context of the educational system of England. Question 2 examines teachers'

perceptions of instruction in educational assessment in initial teacher education in terms of the degree to which they feel prepared for the classroom. The views of initial teacher educators in regard to educational assessment theory and practice covered in teacher preparation are also elicited. The question considers further teachers' perceptions of educational assessment in relation to the accountability system, for example through the use of school performance tables published by the DfE and the expectations of school leadership teams, for example through internal appraisal systems.

Question 3 seeks to gain an understanding of the perceptions of a sample of teachers and an identified group of 'Key Influencers' on the future role of teachers in educational assessment. The term 'Key Influencers' is used here to describe a group of selected individuals who hold senior positions in political, academic and administrative spheres and therefore hold positions of broader influence on the English education system. A further discussion of Key Influencers is presented in Section 3.10. In collecting data to address research questions 2 and 3, two approaches were adopted: questionnaire surveys and interviews. These two methods are now presented in the following sections.

3.6 Questionnaire surveys

The questionnaire surveys adopted for this study utilised communication technology to provide wide distribution at little expense and offer anonymity and flexibility of completion to participants. This provides the potential for higher completion rates and therefore more data to inform future phases. Using information from the literature review to formulate the questions, the questionnaires were designed and disseminated using the Bristol on-line Survey tool (BOS, 2017) recommended by the University of Leeds. The survey tool is designed for academic research and complies with UK data protection laws and offers flexibility in question types and analytics and is supported by on-line and telephone guidance. In 2018, BOS was re-named Online surveys.

Two questionnaire surveys were developed to support this study, a survey of initial teacher trainers (ITT) and a survey of practicing teachers. The term ITT was adopted here because of references to the Carter Review of ITT (2015). The central element of both questionnaires is based on aspects regarding trust, quality and range of provision in teacher preparation in educational assessment as identified by the NAHT Assessment Commission (2014) the Carter Review (ibid), elements contained in Teachers' Standard Six '*Make accurate and productive use of assessment'* (DfE, 2011c), and the framework of core content for initial teacher training (DfE, 2016a). Further elements of educational

assessment theory and practice were also included drawing on the Cambridge Approach to Assessment (Oates, 2017) and the broader educational assessment literature. The surveys also examine the types of teacher preparation course offered in the case of ITT providers, or taken in the case of teachers, subject or other specialisms, views on the confidence levels in the quality of content coverage, the identification of areas not covered in initial teacher training courses that would benefit practicing teachers and responses to a critique that there are gaps in both the capacity of schools and ITT providers in the theoretical and technical aspects of assessment (Carter, ibid). A range of question types were developed for both surveys including closed response, multiple choice, free-text, scaled response type and grids (see Appendix 3 & 4: Questionnaire Surveys). Although the questionnaires were developed around the same aspects as noted above, the questions were tailored to meet the ITT provider and teacher audiences. For example, in the case of ITT providers, questions included enquiries into the range of course provided and student numbers. For practicing teachers, questions focused on their individual route into teaching. Responses to these questions provide quantitative and qualitative data as presented in the data analysis provided in Chapters 5 and 6.

3.6.1 Questionnaire pilot

The questionnaire surveys were piloted with six participants and comments were received from my supervisors. Feedback included written and conversational responses and picked up on issues regarding the length of the survey, selection of question types, sequencing and the format of rating scale. A Likert-type (Sauro, 2014) rating scale approach was included in the design of the published survey. Most Likert scales use a five-option format ranging from Strongly Disagree to Strongly Agree with a neutral option in the middle (Sauro, ibid; Bertram, 2017) coded 1 to 5 respectively. Likert scales produce ordinal data, as for example a 2 and 3 on the 5-point scale isn't necessarily the same as the difference between a 4 and 5. However, whilst a Likert item is technically ordinal, the mean of an appropriate set of Likert items can be used as continuous data, standard deviation and confidence intervals (Sauro, ibid). For the purpose of the surveys, I opted for a four-point scale to produce an ipsative or 'forced choice' where no indifferent option is available (Bertram, Ibid). This approach was selected after the pilot to reduce neutral responses. The pilot also used a ten-point scale to produce more granularity but on reflection there appeared to be little advantage in such granularity for what are in effect questions aimed at eliciting broader analysis.

Although the two surveys covered common elements, they required some differences in approach and presentation in response to the targeted participants. The approaches taken for each survey are discussed in the following sections.

3.7 Questionnaire survey for initial teacher trainers

3.7.1 The approach

This survey questionnaire is designed to ascertain ITT providers' course coverage and perceptions of course quality regarding aspects of assessment. For the initial approach to distributing the survey, the top 10 and bottom 10 ITT providers in terms of students numbers were selected from the DfE statistical census statistics on initial teacher training 2016-2017 (DfE, 2018a). I also contacted the Universities' Council for the Education of Teachers (UCET) who kindly advertised the survey in their e-newsletter. The DfE data do not include contact details for the ITT providers and officials would not provide such details due to data protection issues. However, in order to ascertain details of key contacts for each provider, a second database of ITT providers found in the UK Register of Learning Providers (UKRLP, 2017) had to be consulted using the provider unique reference number taken from the DfE statistical census. These designated 'primary contacts' proved to be diverse in terms of their roles for example university vice-chancellors, school principals and academy chief executives and administrators: Figure 3.3 provides an example response.

Figure 3.3: Anonymised extract from an e-mail response from an ITT provider institution administrator

> On 25 Sep 2017, at 14:40, James (NAME) <jimi@INST.org.uk> wrote:
>
> Hi Nick (Sic),
>
> I'm not so experienced in this subject area, but I've passed on your email to my colleagues to see if they can help.
>
> All the best,
>
> James

The initial response to the ITT providers' survey was poor with only 11 responses received by February 2018. A second approach using a stratified sampling approach (Sarantakos, 1998) to target the top 20, middle 20 and bottom 20 ITT providers in England in terms of student numbers was undertaken in July and August 2018 (see Appendix 2). Of these

sixty institutions, the number of trainee teachers in the top provider in 2016/17 was 1,333 with seven in the lowest provider. The median was 46 students. The top 20 institutions totalled 12,985 students, the middle twenty 920 students and the bottom twenty 265 students. This accounts for 14,170 students out of a total of 32,248 or 44% of students enrolled in pre and post-graduate training programmes in 2016/17 (DfE, 2018d&e). Contact was made in July and August to avoid busy teaching and end of course assessment times, although there was a risk that the holiday period would lead to fewer responses. The use of reminders, discussed below, was therefore built into the contact strategy.

For these sixty providers, I used the United Kingdom Register of Learning Providers (UKRLP, 2017) database to establish contact information in order to contact each provider by telephone to identify e-mail contacts for actual course leaders rather than the designated 'primary contact' as identified in the UKRLP listings. Having spoken to staff in each institution, it became clear that in a number of providers, more than one course leader contact was identifiable, for example a primary phase course or secondary phase as shown in Figure 3.4.

Figure 3.4: Anonymised e-mail response providing course leader details within one ITT provider institution

Hi Mick

Programme Leads for our ITT programmes are as follows:

BA Primary Education: (Name provided)
PGCE Primary – (Name provided)
PGCE Secondary – (Name provided)

Kind Regards

(Name)
Programme Administrator
Academic Services, Faculty of Health, Education & Life Sciences
Address Provided
NAME@INST.ac.uk

In such cases, I was provided with contact details for the programme leads (removed here to protect their anonymity) and subsequently made e-mail contact with each of the named course leaders. Some of these contacts led to brief discussions with course leaders and highlighted the point that although they were prepared to submit a survey response, they were keen to point out that they did not have knowledge of course content across their institutions. A further notification was also published in the August 17th 2018 UCET e-newsletter. I also extended the window of access to the survey and sent two reminders to the stratified sample in August and September 2018. According to Survey Monkey (2019):

There's no doubt reminders are useful, but optimizing them effectively—without overdoing it—is a complicated art for even the most experienced survey researcher. (Accessed on-line, April 15th 2019).

In line with this guidance, I limited the reminders to two and included an appreciation for their time if they had already attended to the survey. These efforts did show a distinct rise in completed surveys and resulted in a final total of 49 responses from ITT providers offering a range of routes into teaching (Table 3.5) at the close of the survey in October 2018. This is from a total of 242 ITT providers listed in the DfE ITT statistics for the 2016/17 academic year (DfE, 2018a, d & e). Three further responses were deleted, as they were incomplete.

Figure 3.5: Questionnaire survey responses by ITT route category
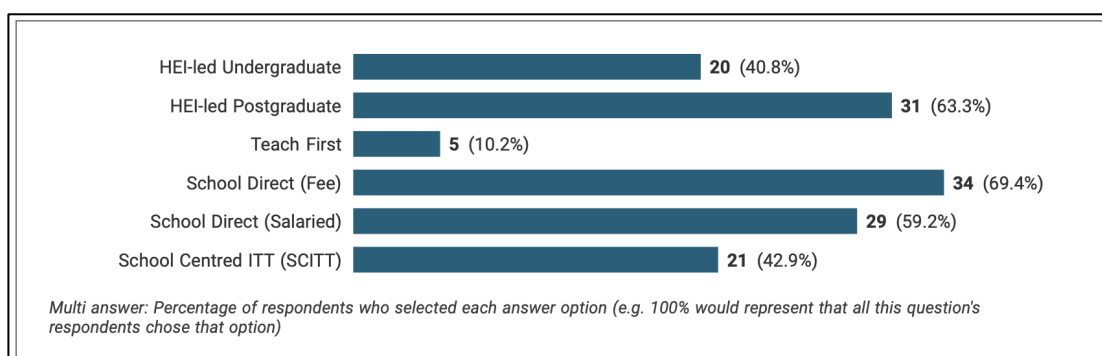
Note: respondents were able to select more than one option.

3.7.2 Questionnaire survey presentation

Having ascertained the contact details for course leaders, initial e-mail contact was made with two follow-up emails as noted above. The survey asked for details on the routes into ITT, phases of education and qualifications on offer and the numbers of students. The major focus then followed the content provided in the survey for teachers on aspects of educational assessment theory and practice covered in their respective courses, confidence in their levels of expertise in educational assessment and any issues that constrain coverage. A copy of the survey of ITT providers can be found at Appendix 3.

3.8 Questionnaire survey for teachers

3.8.1 The approach

The second questionnaire survey was designed to ascertain the perceptions of practicing teachers on the degree to which their initial training in assessment prepared them for practice together with information on their route into teaching and details of their ITT course content. According to DfE school teacher workforce figures:

> There were 498,100 teachers in state-funded schools in England in 2017 (including classroom teachers, headteachers, and deputy and assistant heads) – when part-time work was taken into account, this amounts to the equivalent of 451,900 full-time teachers (DfE, 2018b).

There is no central database or register containing details of practicing teachers. The initial contact strategy therefore was to follow a non-probability sampling approach (Cohen et al., 2018; Dorneyei, 2007). This approach is a 'less than perfect compromise…that reality forces on the researcher' (Ibid., p. 98). In practice, two broad approaches were adopted a) 'snowball sampling' (Dorneyei, ibid; Bryman, 2008) whereby practicing teachers known to me through previous professional engagement were asked to complete the survey and pass on the link to their colleagues and b) 'opportunity sampling' or 'convenience sampling' (Dorneyei, ibid; Bryman, ibid); whereby practicing teachers attending conferences or meetings I attended were invited to take part.

Respondents are all qualified teachers who have trained in England and cover a range of routes into teaching, phases of education and subject specialisms. In this sense as members of a profession, they are homogenous where variation may be less (Bryman, ibid) and may be described as 'appropriate' in that they provide insights into the questions raised by this study and the opportunity to develop rich understandings on which theoretical generalisations may be based (Saunders, 2012).

The initial approach was to send invitations to contacts I had made over my time working in the education system to take part in the survey. This was bolstered by support from professional organisations such as the Chartered Institute of Educational Assessors (CIEA), the @Beyond Levels#Learning First teaching community, FrogEducation and Evidence Based Education all of whom placed notifications on their respective websites or through forwarding Twitter notifications. The response rate at the close of the survey in

November 2018 was 214 in total with 1 deleted questionnaire as it did not contain any responses to the questions. Figure 3.6 shows a breakdown of respondents by the current routes into teaching. Twenty-two other respondents submitted under other categories of routes in to teaching including PGCE, GTP, Cert Ed and BEd.

Figure 3.6: Teacher survey questionnaire respondents by their route into teaching

Multi answer: Percentage of respondents who selected each answer option (e.g. 100% would represent that all this question's respondents chose that option)
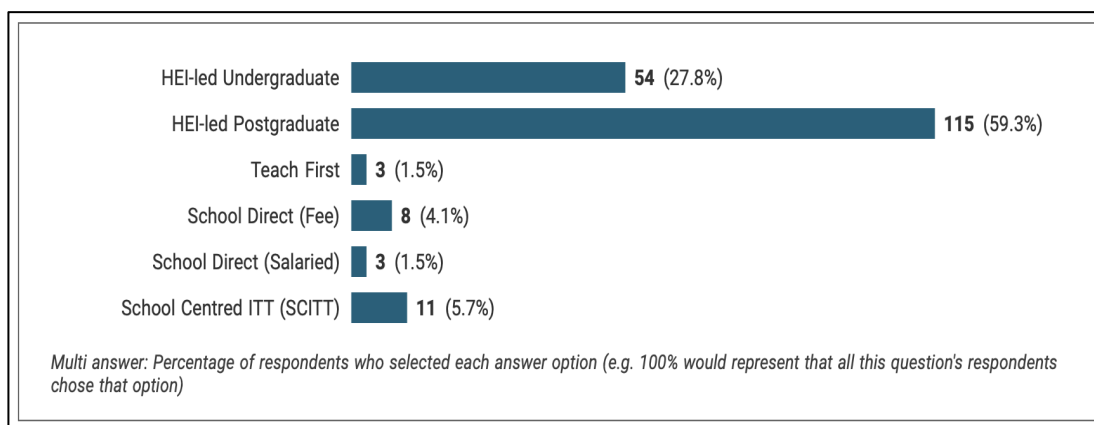
3.8.2 Questionnaire survey presentation

The survey asked for details on the routes into ITT, phases of education and qualifications, date of ITT completion, phase and subjects covered. The major focus then followed the content provided in the survey for initial teacher trainers on aspects of educational assessment covered in their respective courses, perceived confidence in their levels of expertise in educational assessment of their course providers and the extent to which they felt prepared when first entering the classroom as qualified teachers. A copy of the survey of teachers can be found at Appendix 4.

3.9 Interviews

Two sets of interviews are used in this study, interviews with practicing teachers (n=6) and interviews with Key Influencers (n=14): the term 'Key Influencers' is defined in Section 3.11 below. The purpose of the interviews is to gain nuanced insight into 'the lived experience of other people and the meaning they make of that experience' (Seidman, 2006. p. 9). The interviews conducted with practicing teachers in this project aim to complement data derived from the survey questionnaires, allowing the interviewer the flexibility to probe deeper into areas of interest and build insight into the lived experiences and perceptions of subjects (Bailey, 1994, cited in Cohen et al., 2018). The interviews

with Key Influencers aim to provide insight into the views and perceptions of subjects in positions of power or status to influence the role of teachers as educational assessors through their accepted role or expertise. The approaches to interviewing these two populations are discussed in the following sections.

3.10 Interviews with teachers

3.10.1 The interview approach

All participants in this element of the research are fully qualified teachers working in maintained schools in England and educated and trained in the English education system. The selected approach follows a series of repeated interviews, using the three-interview series (Seidman, 2013). This method is underpinned by a phenomenological approach using primarily, though not exclusively, open ended questions with the goal of participants reconstructing their experiences (Seidman ibid), in this case focussing on their early experiences of educational assessment and the knowledge of assessment theory developed during their teacher preparation; their current perceptions of educational assessment and the influence of other forces such as internal sources, for example through performance management systems and external sources such as school performance tables.

The three-interview technique allows; "…the interviewer and participant to explore the participant's experience, place it in context, and reflect on its meaning" (Seidman, ibid., p. 20). Drawing on the work of Seidman, the three-stage approach has been adapted to reflect the goals of this study. Given that all participants in this section of the study are qualified teachers the three stages are:

1. Setting the context: establishing the background, that is their experiences up to the present time; this stage aims to provide insight into their own experience of educational assessment from their time as pupils through to their initial teacher education phase.

2. Establishing the details of their present lived experience, that is to focus on how they view educational assessment and how it impacts and influences their current practice be that through external expectations or the use of assessment in their classroom practice.

3. Establishing perceptions of the function of educational assessment and how they see future possibilities in terms of their role in high-stakes educational assessment and the use of assessment in what might be described as 'low stakes' assessment.

In using the three-stage approach, Seidman recommends interviews of approximately 90 minutes and contained in a two to three week period. The selection of 90 minutes is suggested as a means of reducing a focus on the clock by not using a standard unit of time and the three-week period supports the development of a relationship between the interviewer and interviewee. This time frame is not presented as an absolute and the exploration of alternatives are recognised. However, what is stressed is the need to maintain the overall structure of the process in order for; "…the interviewer and participant to explore the participant's experience, place it in context, and reflect on its meaning" (Ibid., p. 20). This approach arguably aligns with that of the case study approach in drilling down into the perceived reality of participant's experiences. Thomas (2011) in supporting this approach states:

> We must look at our subject from many and varied angles, to develop what the great historian-philosopher Michel Foucault (1981) called 'a polyhedron of intelligibility'. By this he meant that inquires in the humanities and social sciences are too often one-dimensional, as if we are looking at our subject from one direction. In looking from several directions, a more rounded, richer, more balanced picture of our subject is developed – we get a three-dimensional view (p. 4).

Using the three-staged approach therefore provides for more in-depth understanding of the perceptions of the selected participants and aligns with the Foucaultian approach used to address question 1 of this study as presented in Chapter 4. However, it also introduces further manageability issues through effectively multiplying interviews by three occasions. This is further challenged by the recommendation to contain the interviews in or around a three-week period. In interviewing teachers, there is also an issue of determining a schedule that minimises competition with key events in the educational calendar such as end of key stage tests and general examinations. Care was therefore taken to prevent conflict with other major events in the educational calendar of each of the schools and with consideration of the workload of each participant. In order to assist in developing an open and trusting relationship with the participants, a face-to-face approach was selected for the initial interview. Participants were offered the option of holding subsequent interviews

through Skype or telephone. However, in line with the interviewees stated preferences, all interviews with teachers were conducted face-to-face.

3.10.2 The selection of teacher interview participants

To maintain the interviews within a manageable framework the following approach and time frame was adopted. Interviews were conducted with teachers who had indicated in the questionnaire survey that they would be prepared to be interviewed as a follow-up activity. A purposeful sampling approach was adopted whereby participants were selected from the pool of volunteers who have direct experience of the issues at the centre of this study (Creswell and Plano Clarke, 2011). This approach was further refined by homogenous sampling (Dornyei, 2007) through the selection of participants from sub-groups from within the teaching population to include teachers from the primary and secondary phases of education. For balance, three teachers from each phase were selected. Given the practicalities of conducting three interviews with each of the six participants, an element of convenience sampling (ibid) was also at play in that one criterion for selecting the participants was on the basis of geographical proximity. When taken alongside a second set of interviews with Key Influencers, the resources available to the sole researcher, access to resources and time come in to play. This of course needs to be balanced with the generation of sufficient data to support the intent of the study. Seidman's (2013) guidance is helpful in this regard:

> The criteria of sufficiency and saturation are useful, but practical exigencies of time, money, and other resources also play a role, especially in doctoral research. The method of in-depth, phenomenological interviewing applied to a sample of participants who all experience similar structural and social conditions gives enormous power to the stories of a relatively few participants (p. 55).

Arguably, there is also an element of maximum variation sampling (Seidman, 2013; Dornyei, 2007) in the profile of selected participants in that they cover both phases of education of interest to this study, the primary and secondary phase. Of the six teachers selected for interview, four teachers, two primary and two secondary teachers, were chosen on the basis that they hold senior management roles including responsibility for assessment in their respective schools. Two further teachers, one primary and one secondary were selected on the basis that they had been in the profession for five-six years. The rationale behind this is that they are relatively new to the profession so they can relate their training to the realities of their current posts but with sufficient experience

to be able to reflect on how their relative training courses aligned with their developing roles. Given that they have been in the profession for just over five years, they have also cleared what some see as a particular milestone in relation to retaining teachers in the profession (Wilson, 2020; Guardian, 2016; Cuban, 2010). The DfE Analysis of teacher supply, retention and mobility (September, 2018b) reflects the significance of this milestone in that 30% of female and 35% of male NQTs qualifying in 2012 had left the profession by 2017.

3.10.3 Teacher interviewee profiles

The names and locations of the teachers selected for interview have been anonymised to protect their identities in line with the ethical stance taken in this study (see Appendix 5: Ethical Approval). Table 3.2 provides details of the various backgrounds of the teacher interview participants. The reference presented in the table is used to identify each participant in the following chapters.

Table 3.2: Teacher interview participants: background information

Name and Reference	Phase	Current role	Background	Reasons for Selection
Helen HelenP	Primary	Deputy Headteacher	Completed initial teacher training in 1996. Holds a Bachelor of Education degree (Combined Humanities) and Qualified Teacher Status.	Holds responsibility for assessment across the school. Senior Leadership position. Trained through four-year BEd course.
Tim TimP	Primary	Year 6 teacher	Completed initial teacher training in 2013 through the Graduate Teaching Programme (GTP) - School Based after a period as a teaching assistant. Holds a Bachelor's degree in Geography and Qualified Teacher Status.	Qualified for 5 years. Year Six teacher preparing children for National Curriculum Assessments. Qualified through GTP.
Graham GrahamP	Primary	Deputy Headteacher	Completed initial teacher training in 1993. Holds a Bachelor's degree in Mathematics. Completed a Post Graduate Certificate of Education after one year in various jobs.	Holds responsibility for assessment across the school. Senior Leadership position. Has experience as Moderator for National Curriculum Assessments. Qualified through PGCE route.
John JohnS	Secondary	Vice-Principal	Completed initial teacher training in 1999. Holds a Bachelor's degree in Mathematics and a Post Graduate Certificate of Education.	Senior Leadership position. Has experience as a marker and examiner for GCSE Assessments. Qualified through PGCE route.

| Sara SaraS | Secondary | Mathematics teacher | Completed initial teacher training in 2011 through the Graduate Teaching Programme – School based. Holds a Bachelor's degree in Accounting and Finance Six years various jobs before teaching. Qualified Teacher Status. | Teaches GCSE classes. Works with feeder primary schools on mathematics. Qualified seven years. Qualified through GTP. |
| Angela AngelaS | Secondary | Science teacher | Completed initial teacher training in 2001. Holds a Bachelor's degree in Science/Biology. Worked one year before taking a Post Graduate Certificate of Education. | Holds senior management position – head of science. Teaches GCSE classes. Qualified through PGCE route. |

3.10.4 The timing of teacher interviews

Further consideration was given to the timing of the interviews. Finding three periods of around ninety minutes within a three-week timeframe is problematic for teachers with a full-time teaching commitment and other duties relating to planning or management. Given contemporaneous concerns over teacher workload (DfE, 2015b), the approach had to be particularly sensitive. I therefore elected to contact teachers who had made clear that they were willing to be interviewed but mainly from schools with which I was familiar from having previous professional contact. On the one hand, this approach has the advantage that setting up and conducting interviews was easier to organise and develop further as there was an existing relationship of trust. Conversely the familiarity may skew their responses in the light of knowing my professional background. My view is that as these schools and some of the individuals are known to me through former professional engagements – which have always been forthright and professional encounters - the potential benefits outweigh the disadvantages. However, I am aware of the potential bias in the design and conduct of interviews and following the advice of Thomas (2013), I made every effort to take a disinterested stance, that is "…detached, without bias" (ibid., p. 108) but appreciate that:

> Although the interviewer can strive to have the meaning being made in the interview as much a function of the participant's reconstruction and reflection as possible, the interviewer must nevertheless recognize that the meaning is, to some degree, a function of the participant's interaction with the interviewer. (Seidman, 2006. p. 23).

The notions of validity and reliability have been questioned in terms of their relevance to qualitative methods (Bryman, 2008) leading to alternative criteria for evaluating qualitative research. My position is that by making the methodological approach transparent and making my account plausible and credible in relation to the to the kind of evidence presented (Hammersley, cited in Bryman, ibid. See also Aguinis and Solarino, 2019), the project will at least in part provide credible answers to the research questions posed. As Seidman (ibid) states:

> If the interview structure works to allow them to make sense to themselves as well as to the interviewer, then it has gone a long way toward validity (p. 24).

As all of the teachers selected for interview were involved in high-stakes assessments at their respective schools, the chosen timeframe was selected so as to avoid key test and examination clashes (Table 3.3). In practice, the selected timeframe was acceptable to the teachers, although most elected to complete the interviews in less than three weeks so as to avoid conflict with other professional responsibilities. However, with the exception of one interviewee, the format of breaks between the interviews was welcomed by interviewees as offering breaks for reflection between interviews and to accommodate other demands on their time. Although interviews of ninety minutes were offered over a three week period, in practice all interviews were completed within this timeframe, with interviews most commonly running around sixty minutes and within a two week period. In total, the nine interviews resulted in over seventeen hours of interview time (1,061 minutes) with a mean average of just under 177 minutes per interviewee for the three interviews.

The exception was one teacher (GrahamP) who had taken on commitments as a test moderator, which meant his time pressures became more acute. However, given his role in the national test system, I was keen to maintain his involvement in the study. To make this happen, we agreed to complete the interviews in one day but with breaks in between each interview to provide some level of reflection.

Table 3.3: Teacher interview schedule
    In each of three blocks of interviews, 2 participants were identified for interview.

Block 1			Block 2			Block 3		
w/c May 14th	w/c May 21st	w/c May 28th	w/c June 4th	w/c June 11th	w/c June 18th	w/c June 25th	w/c July 2nd	w/c July 9th
Interview 1	Interview 2	Interview 3						
			Interview 1	Interview 2	Interview 3			
						Interview 1	Interview 2	Interview 3

3.10.5 The teacher interview structure

The interviews with teachers followed a semi-structured approach using a set of topics and lead questions as the interview guide (Bryman, 2008; Arksey and Knight, 1999). In this form of approach;

> The interview is loosely structured (thus allowing for some degree of comparability) around an interview guide that contains key questions. Interviewers are free to follow up ideas, probe responses and ask for clarification or further elaboration. For their part, informants can answer the questions in terms of what they see as important (Arksey and Knight, 1999, p. 7).

The set of topics was developed to reflect the key questions of this study and the findings of the literature review. However, the lead questions are open-ended and the sequencing and wording was used flexibly so as to tailor each interview to the individual subject and the responses given (Cohen et al, 2018). The intention here is to allow sufficient flexibility to provide participants with the freedom to raise or develop particular areas related to their own lived experiences to generate findings that can be used to test the findings of the literature review and the outcomes of other data streams: this reflects an inductive approach (Gabriel, 2013; Bryman, 2008). The three interviews developed for teachers followed a progression of stages each designed to fulfil a particular purpose. The main purpose of Interview 1 was to set the context by establishing the background of the subjects by examining their experiences of educational assessment up to the current time

covering their own time as students including their experience of their initial teacher education phase and their entry into the classroom. The focus of Interview 2 was to establish the details of their present lived experience, that is to focus on their perceptions of educational assessment and how it impacts their day-to-day experiences in the classroom be that through external expectations, such as the school accountability system or the use of educational assessment in their pedagogy. Interview 3 provided a focus on establishing the teachers' perceptions of the function of educational assessment and how they see future possibilities in terms of their role in high-stakes assessment and the use of assessment in what might be described as 'low stakes' assessment.

For each topic, a lead question was designed to provide a structure for the interviews with each lead question supported by a number of prompts if needed to support a discussion. For example in Interview 1, under theme 3 Route into teaching, the lead question was: *How and why did you get into teaching?* Copies of the topics and key questions used in each of the three interviews can be found at Appendix 6. However, the interview process allowed for subjects to develop areas that for them best presented their views on the way educational assessment has and still does impact their lived experiences. This approach reduces potential bias emanating from the interviewer as:

> A more standardised and structured approach might overly impose the researcher's own framework of meaning and understanding into the consequent data. It might also overlook events and experiences that are important from the interviewees' point of view, that are relevant to the research but have not been anticipated (Mason, 2011. p. 1021).

All interviews were recorded (voice only) and transcribed. Chapter 7 presents the data from the teacher interviews and is arranged to reflect the topics for each interview.

3.11 Interviews with Key Influencers

The term 'Key Influencers' is used in this study to describe participants who represent a group of people holding particular influence on the policy, regulation and administration of the English education system. Such groups are often described as 'elite' in the research literature (Aguinis and Solarino, 2019; Rubin and Rubin, 1995; Moyser, 2011; Seidman, 2006) and although the term 'elite' may be used; "…in a relative or an ipsative sense in these contexts" (Stephens, 2007 cited in Aguinis et al., ibid., p. 3), my view is that it in common parlance, the term 'elite' confers an often unwarranted aura around such people.

For most part these people are transient, especially in political circles, where it is often their position rather than their knowledge that gives them power. However, some do possess important and sometimes specialist knowledge and experience and have influence on the system at a national level and it is that which makes them key.

3.11.1 The interview approach

The interview approach for the Key Influencers recognises that not all are or have been teachers in the English education system and that they predominantly work in environments where a commitment to three ninety-minute interviews would prove difficult, particularly politicians and regulators. Therefore the topics and lead questions were focused at a level of generality covering for example broad policy issues. Some of those who agreed to be interviewed do not generally grant access to such an open dialogue (Arksey and Knight, 1999) but it was noticeable that the participants responded to the interview questions in a more direct even more efficient way than the teachers. This may well be due to their broader experiences of being interviewed. Although of little direct relevance to the quality of responses, it did mean that interviews were conducted in line with the intended timescale. All interviews were recorded (voice only) and transcribed.

3.11.2 The selection of Key Influencer interview participants

The Key Influencers taking part in this study were selected in order to provide a range of interest groups and perspectives based on their respective roles including politicians, academics, leaders of education related representative bodies, regulatory bodies and those responsible for policy delivery within the English education system. Further selection criteria included their seniority and field of influence, their knowledge of the English education system and the significance of the place of educational assessment. The final 14 participants approached for the purposes of the research represent differing but influential communities and/or enjoy broad reputational recognition as experts in the field. As noted in the literature, such a group; "…can pose considerable problems of access for scholars even in relatively open cultural milieu" (Moyser, 2011. p. 85). However, I already had previous professional contact with all of the Key Influencers during my career in education. This aspect of familiarity with participants is discussed in the axiology section of this chapter.

There is no particular significance in choosing 14 participants. The goal was to provide sufficient coverage of influential bodies and individuals. Given the envisaged length of

each interview, and the use of technology to aid communications, for example Skype, where necessary, there were no concerns regarding the manageability of conducting the interviews. All of those approached agreed to take part in the interviews. As such, the participants represent academics with specialist educational assessment expertise, high profile professional organisations, regulatory bodies, government and local authority officials and politicians with a high level of involvement and experience with educational matters. As with the teachers selected for interview, their identities have been anonymised. With this group of informants, anonymity is not always a concern as they are often widely quoted and used to having their names associated with their words and opinions, for example politicians and academics However, for others within this group, it allowed discussions to move 'off the record' providing greater freedom of expression. This is of particular relevance to those participants who play administrative or advisory roles at a national level. On balance, I think it right to treat the group with the same ethical approach as that for the teacher participants. This approach helps to gain an inner perspective 'which is worthwhile and important' (Seidman, 2006. p. 171). Given the limited number of individuals with such particular backgrounds, their details have been masked to avoid compromising their anonymity. Their anonymised profiles are presented in Table 3.4.

3.11.3 The timing of Key Influencer interviews

Focussing the time for the interviews with teachers towards the end of the summer term provided the opportunity to interview Key Influencers from July to October 2018. This timeframe was in reality chosen independently of considerations of the scheduling of teachers' interviews. In determining the timeframe for interviewing Key Influencers, a number of other practicalities required consideration. The majority of these individuals are in posts that follow the annual patterns of the education system timetable not dissimilar to those of teachers, though several have demands on their time during the main summer break taken by schools. Their roles also mean that for significant periods they are not office based and are somewhat more difficult to 'pin down'. Because of this, a number of the interviews were conducted through Skype or by telephone. However, where ever possible, face-to-face interviews were conducted. With the exception of one interview, all interviews were completed within the designated time frame. The one exception was an interview with a government subject who had to re-schedule because of events in parliament. Given considerations of access and personal resources, the interviews were designed to be completed in one session of approximately one hour. In practice, this worked out well with most interviews completing within the designated time.

Table 3.4: The Key Influencer participants: background information

Name and Reference	Occupation	Relevant roles	Key reasons for Selection & mode of interview
AliA **AliA**	Academic	Policy advice and evaluation. Academic	Experience of regulation. Leading academic in the field of educational assessment. Skype interview.
Ash **AshA**	Academic	Policy advice and evaluation. Academic.	Experience of national policy advice. Leading academic in the field of educational assessment and curriculum development. Skype interview.
Pat **PatA**	Academic	Policy advice and evaluation. Academic.	Experience of national policy advice. Leading academic in the field of educational assessment and curriculum development. Skype interview.
Terry **TerryA**	Academic	Policy advice and evaluation. Academic.	Experience of national policy advice. Leading academic in the field of educational assessment and curriculum development. Telephone interview.
Sam **SamCS**	Civil Servant	Standards and testing.	Experience of policy advice, administration of tests. Face-to-face interview.
Hilary **HilaryLA**	Local Education Authority adviser	Local Authority assessment lead adviser	Experience of local education authority assessment policy development and school support. Face-to-face interview.
Jo **JoPol**	Politician	Government education policy.	Experience at ministerial and cabinet level. Face-to-face interview.
Logan **LoganPol**	Politician	Government education policy.	Experience at ministerial level. Face-to-face interview.
Lou **LouPol**	Politician	Government education policy.	Experience at ministerial and Select Committee level. Face-to-face interview.
Mel **MelPol**	Politician	Opposition party education policy.	Experience as Shadow Minister for education. Face-to-face interview.
Amar **AmarPA**	Professional association executive	Professional association executive. Academic.	Experience of policy advice. Professional association senior executive. Face-to-face interview.
Hayden **HaydenPA**	Professional association executive	Policy advice professional association executive. Academic.	Professional association senior executive. Telephone interview.
Chris **ChrisR**	Regulatory executive	School inspection.	Experience of school inspection and national policy advice. Face-to-face interview.
Jaden **JadenR**	Regulatory executive	Examination and test regulation.	Experience of regulation. Leading academic in the field of educational assessment. Face-to-face interview.

3.11.4 The development of Key Influencer interview topics

In line with the teacher interviews, the topics used in the interviews with Key Influencers were developed to address the key questions of this study and the findings of the literature review. Given the varying roles and backgrounds of the Key Influencers and their time limitations, a more focused and higher paced approach was used in this set of interviews with closer adherence to the pre-determined questions. This did not prevent individuals

from developing areas of particular interest to them further. However, as the topics were developed to address the research questions of this project, they broadly followed the same key issues that were explored in the interviews with teachers. In line with the teacher interviews, lead questions were included in the interview schedule. Copies of the interview topics and lead questions used for the Key Influencers and questions can be found at Appendix 7.

3.12 The data analysis approach

3.12.1 Teacher and Key Influencer interviews

Semi-structured interviews were designed using a deductive approach (Bryman, 2008) to identify topics informed by the literature review and an initial high-level analysis of responses to the teacher survey questionnaires. Lead questions were drafted to introduce each topic supported by a number of sub-questions or prompts. These sub-questions were designed to aid discussion if required. In practice, not all of the questions and prompts were used, or they were adapted during the interviews as the approach adopted a more inductive (Bryman, ibid) manner to allow the flexibility to pursue issues raised by interviewees.

All interviews were recorded using two audio recording devices in order to provide a 'back-up' in case one device failed. In total, the interviews with Key Influencers produced 652 minutes of recording and 1,061 from the teacher interviews. All recordings were downloaded from the recording devices onto an encrypted laptop and deleted from the recording devices. Following completion of the transcription process, audio files were deleted.

The interview recordings were each listened to in order to check sound quality from which notes were taken along with timing references before the recordings were transcribed: this stage of the process provided insight into each interview outcome and the identification of key themes and the views of interviewees. An extract from a set of notes is provided in Appendix 8. Transcripts were then checked for accuracy against the recordings and some corrections were made, most of which related to the use of acronyms or technical words. The transcripts were then analysed in several stages – data reduction, for example the extraction of any extraneous conversational materials or interruptions; and data

organisation and interpretation (Sarnatakos, 1998) to detect early indications of trends or themes.

These stages helped to gain absorption in the content and the identification of emerging themes, phrases and ideas of interest and potential coding categories emanating from the interviewees responses: this process is described by Saldana (2009) as: "…essence capturing…To codify is to arrange things systematically to make something part of a system or classification, to categorise" (p. 80). Consideration was given to the use of either NVivo[20] and/or mechanical methods of coding. Having had previous experience of mechanical methods, such as physically cutting and placing into themes or categories, but not of NVivo, I decided to develop a hybrid using Microsoft Word to develop a data coding and organising grid but taking some of the organisational methods used in NVivo. The interview grids were organised by interview and topics derived from the questions. During the listening to the recordings and checking the transcripts, a list of high-level nodes was identified to categorise the responses. Further analysis provided more nuanced or detailed categories that I described as sub-nodes This stage produced a long list of potential codes that were generated in real time whilst listening to the recording and checking the text. The long list was then re-visited to identify repetition and overlap, an exercise that reduced and focused the list of items that were then grouped by the questions used to guide the interview to form a coding grid. An example extract from a coding grid is shown in Appendix 9.

3.12.2 Coding grids

Topics were denoted in the coding grids by the number of the interview and topic number. So for example, 1T1 refers to interview 1, topic 1 Schooling background and recollections of educational assessment. Sub-topics were generated from the lead questions and are identified by the main topic number and a sub-topic number. So for example, IT1.1 refers to the Early-years and the primary phase of education. The nodes represent themes derived from themes emanating from responses of interviewees, for example, Setting. Sub-nodes were identified to produce more nuanced groupings and were developed in response to their common recurrence most often derived from In Vivo coding, i.e. "…terms used by participants themselves" (Strauss, 1987, p33, cited in Saldana, 2009). Specific coding grids were developed for the teacher and Key Influencer interviews. (See Appendix 9 for a coding grid example).

---

[20] **NVivo** is a qualitative data analysis software package produced by QSR International.

Chunks of text were taken from the transcripts and copied into the grids against relevant nodes and sub-nodes, though during this process further nodes or sub-nodes were identified. The content of the grids was then checked further against the topic questions. During this stage it was possible to highlight elements of the example quotes that were either of clear significance or of secondary interest. This final stage reduced the data to a manageable size in terms of interrogation and the identification of common themes and nodes across both sets of interview data which were used in writing up the analysis for each set of interviews as presented in Chapters 7 and 8.

On reflection, the process provided a good way to engage in depth with the detail within the data, but it did require much more time then envisaged at the start of the process. Some of the data in this phase of coding was removed for lack of relevance to the study, but this sits within the transcripts providing the opportunity for further research.

3.13 Axiology

I had no previous contact with four of the six teacher participants during my professional career. I had however, previous contact with two teacher participants, but wholly in a professional context. With regards to the 14 Key Influencers, I did have professional contact with nine interview participants during my time at the Qualifications and Curriculum Authority and two others since I retired, either in the capacity of providing advice to the DfE or through working groups and commissions. I had no previous contact with the two remaining participants. Prior to agreeing to interview, all interview participants were made aware of my professional career, the intentions of my research and provided with assurance that I had no political or other agenda. Although all interviews were semi-structured, I made every effort to maintain the interview schedule and refrain from interjecting any personal bias.

3.14 Ethical considerations

A number of ethical considerations have been discussed earlier such interviewer bias and insider/outsider considerations. Other key aspects include data protection, anonymisation, storage and sharing of research data. All data used in this research will be stored for no longer than 5 - 7 years as advised by the University of Leeds ethics team.

Before the survey questionnaires were finalised and distributed, ethical considerations were undertaken. This was supported by reference to the University's Research Ethics Policies, discussion with the University's Research Ethics Senior Training and Development Officer and attendance at the Ethics and Ethical Review half-day module presented by the University's Organisational Development and Professional Learning Team (formerly the SDDU). Key considerations included the anonymity and confidentiality of survey respondents in terms of the content of views expressed in their responses and any reference of their data in any outputs of this research study. The survey questionnaire provided participants with the facility to remain anonymous or if they wished to be considered for a follow-up interview, the opportunity to provide contact details. Participants were also informed that by submitting a response to the survey they were giving consent that the information they provided could be anonymised and used for the sole purpose of the study. It was made clear to participants in the questionnaire information sheet that they would not be able to withdraw from the survey once they had submitted their response. Participants volunteering to take part in the follow-up interviews were informed that they could withdraw from the process up to two weeks following the interview.

All interviewees, including Key Influencers, have been anonymised, as have any references to persons made by interviewees during the interviews. A small number of key influences made it known that they were prepared for their identities to be known. However, they accepted that as this was not universal across all interviewees, some of whom preferred to talk 'off the record' and present their personal view rather than those which might be restrained by their professional positions: all Key Influencers agreed to anonymity. An application to the Ethics Committee was submitted on 1st June 2017 and a favourable response from the ESSL, Environment and LUBS (AREA) Faculty Research Ethics Committee was received on 8th November 2017 (Appendix 5).

3.15 Over-view and structure of the data analysis chapters 4-8

The following chapters present an analysis of the data collected through the approaches used in this study, that is: an analysis of documentary and other sources of evidence, interviews with teachers, interviews with Key Influencers, the teacher survey and the ITT provider survey. Each constituency is taken in turn as represented in Figure 3.7, which provides an over-view of the following data analysis chapters.

Figure 3.7: An overview of the data analysis chapters

```
                    ┌─────────────────────────┐
                    │       Chapter 4         │
                    │                         │
                    │  Progress or simply     │
                    │  change? A Foucaultian  │
                    │  perspective            │
                    └─────────────────────────┘
                    ┌─────────────────────────┐
                    │       Chapter 5         │
                    │                         │
                    │  ITT Provider           │
                    │  Questionnaire Survey:  │
                    │  Data analysis          │
                    └─────────────────────────┘
                    ┌─────────────────────────┐
                    │       Chapter 6         │
                    │                         │
                    │  Teacher Questionnaire  │
                    │  Survey:                │
                    │  Data analysis          │
                    └─────────────────────────┘
                    ┌─────────────────────────┐
                    │       Chapter 7         │
                    │                         │
                    │  Interviews with        │
                    │  Teachers               │
                    └─────────────────────────┘
┌──────────────┐    ┌─────────────────────────┐    ┌──────────────────┐
│  Interview 1 │    │      Interview 2        │    │   Interview 3    │
│              │────│                         │────│                  │
│  Context and │    │  Current perceptions of │    │  Educational     │
│  background  │    │  educational assessment │    │  assessment;     │
│              │    │                         │    │  knowledge and   │
│              │    │                         │    │  understanding   │
└──────────────┘    └─────────────────────────┘    └──────────────────┘
                    ┌─────────────────────────┐
                    │       Chapter 8         │
                    │                         │
                    │  Interviews with Key    │
                    │  Influencer             │
                    └─────────────────────────┘
```

Chapter 4: Progress or simply change? A Foucaultian perspective on the role of teachers in high-stakes educational assessment in England since the Norwood Report of 1943

4.1 Introduction

This chapter focuses on Research Question 1, *How has the role of teachers in high-stakes educational assessment changed from the Norwood Report (1943) to the present day?* However, the intention here is not merely to track changes over time in a distinct chronology of events based on a teleology that sees change as a developmental model of progress. In addressing this question, the approach utilises the work of Paul-Michel Foucault as a framework of enquiry rather than consigning his words to the literature review (St. Pierre & Youngblood Jackson, 2014) in order to develop a better understanding of the reasons for change and the perceived decline in trust in teachers' assessment in high-stakes assessments. Following Foucault's approach, this chapter attempts to address the genealogy of a problem expressed in terms current today (Kritzman, 1988, cited in Garland, 2014, p. 367). The aim therefore is to develop an account of educational assessment carried out by teachers in high-stakes assessments over a period of time across the compulsory phase of education that does not assume a model of improved practice or progress towards a recognised vision of perfection or semblance of 'truth'.

4.2 Overview of the approach

This chapter examines the role of teachers in high-stakes educational assessment since the mid twentieth century. This point in time has been selected as the Norwood Report of the curriculum and assessment (1943) brought together questions about external examinations that still resonate today and was shortly followed by the Education Act (1944) which; "…replaced almost all previous education legislation and set the framework for the post-war education system in England and Wales" (Gillard, 2011, Chapter 5). The chapter presents an analysis of the educational, social and political contexts through which educational assessment has changed and the impact on the role played by teachers. The selected approach draws on and extends the review of the academic literature presented in Chapter 2 with a review of primary and secondary texts (Silverman, 2006) including policy documents such as white papers, parliamentary acts, ministerial speeches and select committee papers along with documentary evidence of discourse such as papers produced by teacher associations, newspapers and journals. The aim is to develop an

historical account of assessment by teachers across the compulsory phase of education through what Foucault regards as the prevailing ideologies, archaeology, and transitions in thinking genealogy (Garland, 2014) drawing on a range of available data.

This approach goes beyond understanding the phenomena at the centre of this research project through either positivist or interpretive lenses but looks to form an appreciation of the political and ideological contexts in which educational assessment takes place. This type of approach can be set within the frame of critical educational research that aims to be transformative (Cohen et al, 2007) in that it not only seeks to understand current provision but to promote change, in this case in the interest of improving teaching and learning in schools and increased validity and reliability in high-stakes educational assessments. Using the work of Foucault as a guide, the intention is to develop a cohesive understanding of the changes to the part played by teachers as educational assessors in the English education system over the period of time under consideration through the application of an analysis that does not assume a history of progressive development. Starting from questions conceived in the present, this facet of the study addresses events over time through a contemporaneous lens. This genealogical approach departs from conventional histography to what has been described as Foucault's approach to writing a 'history of the present' (Foucault, 1977a, p. 31. See also Garland, 2014). This concept is addressed in more detail later in the chapter.

4.3 The work of Foucault as a framework for enquiry

It is my intention to gain an understanding of the changing role of teachers' educational assessment over time and a consideration of what factors contributed to these changes. However, to simply chronicle the various changes using a current perspective would not reveal the reality of life for those engaged in the field of education at the point of change. Further, I am yet to be convinced that the various changes to the role of teachers over time - be they initiated by government policy or the teaching profession - were driven by a vision of what ultimate perfection constitutes; in other words, there was change, but did it lead to improvement and therefore produce another step towards a utopian vision of educational assessment? Clearly using a methodical framework to guide this project adds discipline and structure, but I am attracted to an approach that is not tied to a particular school of thought. Rather, an approach that is pragmatic, adaptive and flexible is more appealing. My interest is in producing a functional, fit for purpose account that is not restricted in form by the application of a limited set of tools or by any vision of what constitutes an ideal of what education is or should be (Allen, 2012). It has also become

clear to me that in studying the various works on Foucault, "…interpretations of Foucault's ideas vary almost as widely as the uses to which they are put" (ibid). Therefore my approach is selective and limited in scope drawing only on facets of Foucault's work, principally: 1) that the events of history relating to educational assessment are driven by circumstance and are therefore contingent (see Allen, 2012; Bailey, 2016) rather than determined by a universally agreed vision of perfection or truth: "As a matter of fact, universal truth and value are concepts that he (Foucault) campaigns against" (Wang, 2011, p. 142): and 2) how educational assessment has been used as a technology, an instrument of power that works in a diffuse manner to differentiate, rank and to hold the body to account working through what he described as 'bio-technic-power' (See Foucault,1976):

> Bio-power is 'productive' rather than repressive, acting to control and harness bodies, space and time in more efficient, productive and 'biological/natural' directions. It is a mode of power not directed by anyone, but which increasingly enmeshes everyone. Its purpose is the increase of power and order itself, and it regulates in the name of truth and freedom (Bastalich, 2009, viewed on-line).

My interest here is not in delving into the extended works of Foucault but to apply the way Foucault looked at history to the question at hand, that is, not assuming a model of progress or that actors in other times thought in the way we do now. This approach is illustrated in *The Order of Things* (Foucault, 1970) in which he produced:

> …a history that didn't find continuity, progress origins, and things that were 'the same'. Instead Foucault came up with the notion of periods of history – what he called epistemes – that were organised around their own specific world-views (Danaher et al, 2002. p. 15).

A key notion of Foucault's epistemes is that they are built on discourses consistent with prevailing views of their time with no necessity for linear or progressive links from one to the next, rather they have elements of 'sameness and difference'. Foucault used a pragmatic approach to his work and once remarked that he preferred to utilise the work of other writers rather than enter into elaborate citations; "…he simply *used* (author's italics) the conceptual tools these writers provided, usually transforming them in the process" (Garland, 2014. P. 365). Such an approach does not lend itself to the identification of a grand Foucaultian theory, but rather offers a set of conceptual tools:

…designed as a means of working on specific problems and furthering certain inquiries, rather than as an intellectual end in itself or as a building-block for a grand theoretical edifice (ibid., p. 366).

It is this level of pragmatism that appeals. However, it should be said from the outset that there is an element of risk in using Foucault's works as a methodological framework. Foucault's body of work, though exhibiting a methodological tone, does not add up to a coherent statement of his methodology; "…and they hardly constitute a user-friendly 'how to guide' to Foucaultian scholarship" (Kendal & Wickham, 2003, p. 4). Foucault's approach is essentially utilitarian in drawing influence from other philosophers and applying and modifying their thinking to the issues at hand. Even where Foucault appeared to settle on a particular methodological approach in *The Archaeology of knowledge* (Foucault, 1972), the attempt has been described as a failure:

Why? Because as soon as Foucault resumed his substantive historical work, his new research prompted him to rework his methods and his concepts once again, thereby rendering *The Archaeology* (original italics) redundant (Garland, 2014, p. 366).

However, I view the capacity to modify the approach in response to changing circumstances as an expression of Foucault's pragmatism. His ability to shift approach in order to formulate appropriate ways to gain insight into questions at hand is a strength. This is not a universal view amongst academics some of whom; "… seem exasperated that Foucault does not fit into recognizable categories and does not employ recognizable methodologies" (Marshall, 1996). This reflects Foucault's project as:

…seeking to find a space beyond traditional disciplinary or theoretical positions, from which he could subject those positions to analysis and critique, and trouble the "inscription of progress" in modern politics and scholarship (Ball, 2013, p. 3).

In practice, Foucault developed a set of bespoke tools rather than an overarching methodological framework. The rest of this chapter attempts to utilise Foucault's toolbox as a methodological approach to achieving the research aims of this study with particular reference to research question 1. It is not the intent of this question to find out how the present emerged from the past, rather it's a way of diagnosing the present. Further, it does not see events in a cause effect relationship, but rather the outcome of a set of contingencies leading to 'accidents of history'. This approach views an historical event as

being contingent (see Allen, 2013; Bailey, 2016), that is to say that any outcome is viewed as being but one possible result of a whole series of complex relations between other events (Kendal & Wickham, 2003):

> When we use history, if we are to gain the maximum benefit from the Foucaultian method, …we must ensure that we do not allow this history to stop …to settle on a patch of imagined sensibleness in the field of strangeness (p. 4).

Drawing on what Foucault regards as the prevailing ideologies, archaeology and transitions in thinking genealogy (Garland, 2014) as a methodological approach offers a way of developing an understanding of the way in which teachers' assessment has changed since the early 1940s. Further, and importantly for this study, Foucault's pragmatism offers possibilities that go beyond that of analysis to one of challenge and transformation:

> I do not undertake my analyses to say: look how things are, you are all trapped. I do not say such things except insofar as I consider this to permit some transformation of things (Foucault, 1976a, viewed online).

Despite the question being situated from a point in the history of education in England, the aim is not to compile a history of teachers' assessment as a mere chronology. Foucault does indeed draw on history in his various works, but: "The Foucaultian method's use of history is not a turn to teleology, that is, it does not involve assumptions of progress (or regress)" (Kendall & Wickham, 2003, p. 4). This viewpoint is I believe important in examining how the use of teachers' assessment has varied across the period of time under consideration. Neither does it suggest a development as being the result of an oppressive power driven by a vision of what might constitute perfection. Rather change is the result of a complex web of forces and possibilities:

> '…each discourse undergoes constant change as new utterances are added to it' (Foucault, 1991b, p. 54). New languages continuously plunge us into their inception. No discursive formation is decided in advance, and none appears, fully formed, after examination. Any formation depends upon relations among other types of discourse and within the non-discursive contexts in which it functions— including institutions and social relations, and the conjunction of the economic and the political. The episteme of a period is not a general developmental stage of

reason but 'a complex relationship of successive displacements' (p. 55) (Wang, 2011, p. 148).

The approach therefore is situated in Foucault's 'history of the present' in which there is "…an unequivocal and unabashed contemporary orientation" (Dreyfus & Rainbow, 1982, p. 119 cited in Garland, 2014). Or as Foucault explained in an interview in 1984:

> I set out from a problem expressed in the terms current today and I try to work out its genealogy. Genealogy means that I begin my analysis from a question posed in the present (Kritzman, 1988, p. 262 cited in Garland, 2014, p. 367).

Towards the end of the first chapter of *Discipline and Punish* (Foucault, 1977a), Foucault explains his interest in history as determining an understanding of the past but not in terms of an analysis of the past written in terms of the present. Using Foucault's approach, it is the intention of this chapter to examine the various changes to the role of teachers as educational assessors in a contemporaneous context. The approach is predicated on a view that there is no generally accepted and articulated ultimate state or point of 'truth', that is, that which constitutes the ideal for the role of teachers in high-stakes assessment systems. There are also no predetermined views on what constitutes a particular episteme with regards to educational assessment practice, in other words at what point or why a particular world-view or discourse changed; this is where Foucault's genealogy assists this study in seeking to understand changes in the use of teachers' assessments. Nor does such an approach overlay a model of gradual development or improvement. Rather it seeks to understand why change was undertaken and why particular contingencies were adopted rather than others. Indeed, the 'other' might be of interest to this study, especially ideas or views that failed to predominate at one time only to re-appear down the line. This approach is focused on contemporised thinking manifest in various discourses - the ways in which meaning and values are used to make sense of things; what Danaher et al have described as 'language in action' (2000, p. 31). Foucault developed this approach in *The Order of Things* (1970) in which he describes in the preface how the book came about as a reaction to the humorous and shattering effect of reading a passage from the Argentinian author Jorge Luis Borges on a taxonomy of animals purportedly taken from a Chinese

encyclopaedia that illustrated an alien way of thinking.[21] This caused Foucault; "…to reconsider what is normally understood as the history of ideas and the way in which we understand our connections with the past" (Danaher et al., 2000, p. 14). The result was a line of thinking that people in other times and places may have understood the world in a different way – even in ways we could not imagine (ibid). Foucault utilised a number of approaches to assist in his quest to understand the history of systems of thought – as distinctive from the history of ideas. Of these approaches, archaeology, genealogy and the power/knowledge relationship are of relevance to the use of Foucault as a framework of exploration. Each of these approaches is now considered in turn.

4.3.1 Archaeology

Archaeology was a term much used in Foucault's writing (Foucault, 1970, 1972, 1973). It was a term he developed to describe an approach that excavated the past to reveal layers of discourse in distinct periods of history to reveal distinctive styles of reasoning: "Archaeology wants to show structured order, structural differences and the discontinuities that mark the present from its past" (Garland, 2014). In *The Archaeology of Knowledge* (Foucault, 1972), Foucault describes archaeology as a method by which we can pick at the reality of epistemes through looking at the contemporaneous discourses hidden behind what became the dominant output or product of the period. In identifying key differences between archaeological analysis and the history of ideas, Foucault notes: "They concern the attribution of innovation, the analysis of contradictions, comparative descriptions, and the mapping of transformations" (ibid. p. 155). In practice, the term archaeology as used by Foucault, describes the analysis of particular moments that gave rise to particular thoughts and their ascendency to a position of authority. These systems of thought are underpinned by rules that may not be transparent even to those employing them (Downing, 2008). Foucault's approach was based on a view that for those looking back on these times, these systems of thought may be invisible, or the case of his reaction to reading Borger, viewed as humorous. In discussing his reaction to reading Borges (1942), Foucault notes that the experience was not just laughable but one that; "…shattered, as I read the

---

[21] According to Windschuttle, K. 1997: 'No Chinese encyclopedia has ever described animals under the classification listed by Foucault. In fact, there is no evidence that any Chinese person has ever thought about animals in this way. The taxonomy is fictitious. It is the invention of the Argentinian short-story writer and poet Jorge Luis Borges'. In: The Killing of History: How Literary Critics and Social Theorists Are Murdering Our Past, *from which this article is adapted. Copyright ©1997 by Keith Windschuttle. Reprinted by permission of Simon and Schuster, Inc.*
Cited in: Foucault mis-cited The Chinese Encyclopedia of Animal Classification in The Order of Things http://blog.triciawang.com/post/12151308604/foucault-mis-cited-the-chinese-encyclopedia-of

passage, all the familiar landmarks of my thought – our thought, the thought that bears the stamp of our age and our geography" (1970. p. xvi). Kendall & Wickham (1999) offer two principles of archaeological research:

1. In seeking to provide no more than a description of regularities, differences, transformations, and so on, archaeological research is non-interpretive.
2. In eschewing the search for authors and concentrating instead on statements (and visibilities) archaeological research is non-anthropological (ibid. p. 26).

In providing further clarity relating to principle 1, Kendall & Wickham note the importance of making at least the *attempt* (author's italics) of being non-judgemental and not searching for deeper or hidden meanings. For point 2, the emphasis is on the appearance of statements and the avoidance of seeking to source deep meaning in human beings. In essence, this approach is not therefore a history of ideas, but rather 'a history of the epistemic field that makes certain ideas possible at certain moments, not at others' (Downing, 2008, p. 41). This way of looking at the focus of the question in hand provides a way of bringing to the fore thoughts manifest in discourses related to educational assessment as they existed at certain points in time and why some prevailed, whilst others did not.

4.3.2 Genealogy

In Foucault's later works, he places less emphasis on archaeology with the focus on the dominant discourse at the time, to one more focused on genealogy - the how and why a dominant discourse is replaced. This is of interest to this study in attempting to understand shifts in the role of teachers as assessors and looks to understand contextual influences leading to changes over time and the world we witness today. This approach in Foucault's work is often referred to as 'a history of the present. This has been described as a diagnosis of the current situation (Dreyfus & Rainbow in Garland, 2014). Foucault's genealogy stems from the work of Nietzche which is predicated on the view that any assumption that the epoch in which we live is the high point of civilized achievement is a fiction to be 'debunked' (Downing, 2008) and breaks with traditional historical analysis; the aim of genealogy is to illustrate the singularity of events rather than a search for a point of origin: it thus eschews the notion of relentless progress (Smart, 2002, p. 52). In this sense the events of history are seen:

> …not as the product of destiny, regulated mechanisms or the intentions of a constitutive subject but as the effect of haphazard conflicts, chance, and error, of relations of power and their unintended consequences (ibid., p. 52).

There are several points in time where changes made to the administration of educational assessment have resulted in unintended consequences or perverse incentives; for example the focus on tests used for accountability purposes leading to a narrowing of the curriculum as schools prioritised their teaching on what is measured at the expense of the broader curriculum. Using Foucault's approach, this study is interested in developing an understanding of not only what became the dominant mode of operation, but that of the subjugated knowledge of the time, what Smart has described as knowledge:

> …submerged within functionalist or systematizing modes of thought, and forms of knowledge disqualified by virtue of their location beneath the threshold of scientificity. It is to the realization of this end that genealogy is directed (Smart, 2002, p. 55).

This is not a quest to determine what is right or what is wrong; such an approach would involve judgement, a process Foucault hoped to eliminate: 'I can't help but dream about a type of criticism that would not try to judge' (Foucault, 1988, p. 326, in Kendal & Wickham, 1999, p. 30), and as noted by Sharp (2019), education:

> …is peppered with ...examples of how what we thought was right at the time shaped decisions that were made, and greatly affected people's lives. What these points in history have in common is that we were initially convinced that one approach was right, and then suddenly we weren't (p. 6).

From this, it is the intention of this study to discover '…the small and multiple changes that lead to alterations in trends of thinking and operation in any given epoch' (Downing, 2008, p. 15). Therefore there is no assumption of progress in the understanding and application of educational assessment in the English education system:

> Genealogy ...must record the singularity of events outside any monotonous finality; it must seek them in the most unpromising places, in what we tend to feel is without history – in sentiments, love, conscience, instincts; it must be sensitive to their recurrence, not in order to trace the gradual curve of their evolution but to isolate

the different scenes where they engaged in different roles (Foucault, 1994, pp. 369–370).

4.3.3 Power and knowledge

In reviewing the relationship between various approaches to educational assessment and the role of government and teachers in high-stakes assessments, the work of Foucault provides an 'extra dimension' in the analysis of the intersection of power and knowledge (Kendall & Wickham, 2003). The various modes of teacher assessment within the period under investigation have been central to a range of often confrontational periods between central government and teacher associations more latterly intensified by various models of accountability that have drawn heavily on assessment outcomes. Foucault's 'extra dimension' stems from his genealogical position. In essence, it derives from Nietzsche's genealogy that looked at the historical origins of powerful institutions and discourses making claims of universal and eternal truth (Danaher et al., 2000). Rather than truth and knowledge being the basis of society and culture, Foucault takes the view that they too have a history in a way that is closely related to the way in which power relations have transformed from one of sovereign – subject, based on divine right, to a system of truth and knowledge manifest in in what Foucault describes as the human sciences. Following Nietzsche, Foucault's view is that the dominant disciplines and discourses are the result of power struggles in which they have triumphed over others. Foucault argues that the knowledge and truth produced by the human sciences was, on one level, tied to power because of the way in which it was used to regulate and normalize individuals. It is of importance here to consider how Foucault uses the term power. It is not used in terms that describe violent struggle or oppression, but rather in the way power is in effect a source of energy, the element that makes a machine work; in that sense, Foucault sees power not as repressive but productive; power does encounter, but that resistance is part of how power works (Kendall & Wickham, 1999). In this sense, power permeates all social relationships from the institution to the intersubjective; "…exercised through a web-like structure in which individuals are its vehicle…Power, knowledge and truth underwrite all human relationships" (Lazaroiu, 2013, p. 823-824). In this way, power is deployed in every day transactions and experiences working; "…*within* the social body rather than *from above* it" (author's italics; Foucault, 1980b, p. 39 cited in Jackson, 2013, p. 841). From the point of view of this study, one goal is to understand the variations in power-resistance struggles, not necessarily in terms of a top-down imposition, but in how discourse influenced change and adoption of particular contingent modes of practice:

With regard to resistance to power, because Foucault placed resistance within discourse, the distinction between the oppressors and the oppressed is obfuscated (Hodges et al., 2014, p. 565).

In this, points of transformation of power and knowledge that demonstrate contingency rather than steps of progress towards an ultimate truth can be illustrated. Much of this can be detected through an understanding of everyday life at particular moments of history:

…to recover that which was silenced in a way that enabled the understanding of experiences before they were shaped and redefined by socially sanctioned science and philosophy (Turkel, 1990, p. 171).

In essence, this is an attempt to capture the social context at the point of change. In this sense power is the core from which morality emerges rather than from that of a universal truth (ibid). Foucault presents the view that the aim of the power/knowledge reading is to analyse; "…how things change, transform themselves and migrate" (Foucault, 2000, p. 294 cited in Jackson, 2013, p. 845): he also noted that those involved in power/knowledge relations influence the process because they can; "…in their actions, their resistance, their rebellion, escape them, transform them" (ibid., p. 294). This suggests that in terms of the role of teachers in the assessment process, they have the facility to influence and change practices. However, Turkel suggests that discourses are to some extent controlled, limiting access to speakers who are deemed qualified in terms of formal qualification or professional certification rendering discourse to; "…a form of exclusive communication and interaction" (Turkel, 1990, p. 177) which in practice is the recognition or demarcation of expertise. For Foucault, this can be overcome by methods that criticise the imposition of knowledge practices on things. This again holds interest for this study in seeking to identify 'exclusive communication' and the influence on practice by experts in the field such as examining the views of politicians and those who might be deemed as Key Influencers on the education system.

4.3.4 Foucault's pragmatism

The philosophical underpinnings of this study are driven by the goal of responding to the research questions with useful, practical outcomes that are of direct relevance and accessibility to practitioners (see Section 3.2). Taking a pragmatic approach Garland (2014), supports the intentions of this study by offering a framework described by Foucault as:

"…the most precise and discriminative analyses I can in order to show in what ways things change, are transformed, are displaced…. From this point of view, my entire research rests upon the postulate of an absolute optimism. I do not undertake my analyses to say: look how things are, you are all trapped. I do not say such things except insofar as I consider this to permit some transformation of things. Everything I do, I do in order that it may be of use. (Foucault, 1976a, p. 911).

In an examination of the pragmatism found in Foucault's later work, Fabbrichesi (2016) illustrates the efficacy of philosophy as; "...ethopoietic (habit-forming), not epistemic" (p. 3). This reflects the view proffered by Reynolds (2004) that:

Foucault…sees philosophy's role as one of criticism, not of confirming eternal truths. Moreover, as a form of criticism, philosophy's most vital function is to transform the existing cultural and social order (p. 958).

In addressing what he describes as; "…building bridges across familiar philosophical divides" Koopman (2011, p. 3) questions the notion of genealogy being conceived as looking backwards and pragmatism as looking forward, but rather standing in need of one another:

Any full-scale practice of critical inquiry requires the fulfilment of both intellectual desiderata of reconstruction and problematization - hence critical inquiry itself calls for something like pragmatism that provides a reconstructive service as well as something like genealogy that performs a diagnostic service (p. 6)

In this sense, Foucault's genealogy becomes in itself a pragmatic exercise in that it traces the conditions that make subjects what they are and through doing so, opens up new possibilities (ibid; also see Hodges et al., 2014; Wang, 2011). Fabbrichesi (ibid) notes that Foucault's later reflections on the topic of *parrhesia* (speaking freely – even to power) should not be interpreted as simply speaking the truth; "…but as the "courage" *to act upon what is truthfully held*, (sic) to "work out" the effects of truth so understood. In this way, truth is no longer only speech but becomes pragma" (p. 3). These are important considerations for this research project in identifying possibilities and actions that can be applied in response to problematising the placing of trust in teachers in regard to high stakes educational assessments with the intention of working towards; "…unimagined

forms of social and political life" (Allen, 2012a, accessed online). It is an approach that applies Foucault's purpose:

> …to re-examine evidence and assumptions, to shake up habitual ways of working and thinking, to dissipate conventional familiarities, to re-evaluate rules and institutions starting form this re-problematization (Foucault, 1996b, p. 463; cited in Stickney, 2012).

As noted earlier in this section and in Chapter 3, the underlying philosophy of this research project drives the goal of using 'what works' in order answer the research questions. But the pragmatic approach extends also to finding practical and useful outcomes that can assist in increasing teacher's knowledge of assessment theory and practice, increasing trust in teacher-based assessments and of central importance, supporting the teaching and learning process. Educational research has however been criticised in some quarters for its lack of application to the every-day lives of teachers, their pupils and the wider education community: see for example, Hargreaves and Woodhead cited in Section 3.2. In more recent times, efforts have been made to bridge the space between academic researchers, policy makers and education practitioners, see for example Wilson's (2014) insightful discussion around the development of a sustainable collaboration between these seemingly diverse parties. Yet there is evidence that teachers remain reluctant to use research to inform their teaching be that because of continued issues with the impenetrable language used in research papers, other workload demands on teachers and the lack of practical support from senior school leaders (MacLellan, 2016. See also EEF, 2016; BERA-RSA, 2014; Coldwell et al., 2017).   Badley (2003) also expresses concern for what he describes as the 'crisis in educational research' for which he cites four main causes: 1. *false dualism;* 'the apartheid that divides positivist and constructivist researchers' (See Section 3.2): 2. *false primacy*; 'the domination of the positivist paradigm to the detriment of more open, pluralistic and critically reflective approaches': 3. *false certainty*; that positivist approaches can provide unquestionable solutions to educational problems: and 4. *false expectations*; particularly held by governments, that research should provide urgent solutions to complex educational problems. In response, Badley calls for a pragmatic approach to resolve the inherent issues identified within these four problematised falsehoods. A key part of Badley's suggested pragmatic and 'more modest' approach to both policy and practice is; "…the rejection of the idea that scientific research can be used with certainty to specify educational practice. All it can provide is possible lines of action" (p. 296). This approach aligns with the modest ambitions of this research project and framework of pragmatism provided by Foucault:

> …if I don't ever say what must be done, it isn't because I believe that there's nothing to be done; on the contrary, it is because I think there are a thousand things to do, to invent, to forge, on the part of those who, in recognising the relations of power in which they're implicated, have decided to resist or escape them. From this point of view all my investigations rest on a postulate of absolute optimism (1991b, *Remarks on Marx,* in Ball, 2013, preface).

4.3.5 Foucault's methods

The key aim of this chapter is to apply Foucault's methods to gain an understanding of how changes to educational assessment were influenced by the displacement of dominant discourses through the application of his genealogical approach:

> Let us give the term genealogy to the union of erudite knowledge and local memories which allows us to establish a historical knowledge of struggles and to make use of this knowledge tactically today (Foucault, 1980, p. 83 cited in Ball, 2013. p. 35).

It is hoped that this will provide further insights into change, but in a way more aligned with Foucault's approach to history:

> …what is important to uncover is not so much who speaks but what is spoken, what positions it is spoken from [what Foucault called *authorities of delimitation*], and how this is mediated by the speaking positions of others: an architecture of policy positions (Gale, 2001, p. 389 cited in Ball, 2013, p. 23).

My contention is that over the period of time of interest to this study, there have been key points of change in the involvement of teachers in high-stakes educational assessment. Clearly there is not room here to consider all acts of parliament related to education policy, so the approach I have adopted is to consider key points of change in the way high-stakes educational assessment has been manifest in a set of periods of history in the English education system. However, despite the interest shown in Foucault's methods by educationalists, Foucault himself did not produce any extensive work on education. As Ball notes, what Foucault had to say about education 'was not much' (ibid., p. 26). However, Foucault's approach, particularly that of problematisation, offers a method from

which we can develop an understanding of the drivers of change in educational practice around the area of assessment. Ball notes:

> As Foucault explained, in yet another rendition of his project: "What I tried to do from the beginning was to analyze (sic) the process of 'problematization' – which means: how and why certain things (behavior, (sic) phenomena, process) became a problem" (Foucault, 1983a). That is, he went on to say, he tried to show "…that it was some real existent in the world which was the target of social regulation at a given moment" (2013, p. 28).

Even so, Foucault did engage directly with some aspects of education. In *Discipline and Punish*, Foucault (1977a) refers to the examination (in terms of testing) as a means of an observation hierarchy and those of a normalizing judgment by which it confers a visibility that differentiates and judges: "The superimposition of the power relations and knowledge relations assumes in the examination all its visible brilliance" (ibid., p. 184). The examination therefore; "…makes it possible to qualify, to classify and to punish" (ibid., p. 184). This in practice leads to the objectification of the subject. Foucault notes this as an area that historians of science have not explored:

> But who will write the more general, more fluid, but also more determinant history of the 'examination' – its rituals, its methods, its characters and their roles, its play of questions and answers, its systems of marking and classification? For in this slender technique are to be found a whole domain of knowledge, a whole type of power (Foucault, 1977, p. 185).

4.4 Tests and examinations and the role of the teacher since Foucault's demise

Since Foucault's death in 1984, the examination, or test, has reached new levels as a tool of power and control – a situation that would likely to be of interest to Foucault. The examination is described by Foucault as a disciplinary power that holds subjects in a mechanism of objectification which increasingly made; "…a perpetual comparison of each and all that made it possible both to measure and to judge" (1977a, p. 186). In *Discipline and Punish*, Foucault's focus on the examination centres on the school – teacher – pupil axis. However, since Foucault penned these words, other dimensions of the methods of production, the use and the impact of the examination have emerged most noticeably from

a tool not merely of normalisation (see for example Marshall, 1996, p. 121) or selection but to one of system wide accountability. The system of school accountability adopted in England provides a method of surveillance; "…accompanied by architecture of power ranging from the design of school buildings to the construction and positioning of seating" (Allen, 2012. Viewed on-line). In Discipline and Punish, (1977a) Foucault used the panopticon as a metaphor to explore the relationship between systems of control and discipline and the power-knowledge concept:

> In his view, power and knowledge comes from observing others. It marked the transition to a disciplinary power, with every movement supervised and all events recorded. The result of this surveillance is acceptance of regulations and docility - a normalization of sorts, stemming from the threat of discipline. Suitable behaviour is achieved not through total surveillance, but by panoptic discipline and inducing a population to conform by the internalization of this reality. The actions of the observer are based upon this monitoring and the behaviours he sees exhibited; the more one observes, the more powerful one becomes (Mason, 2019, viewed on-line).

The use of data generated by assessment outcomes taken alongside a monitoring regime predicated on short notice school inspections provides a technology of constant observation through which; "…a consciousness of constant surveillance is internalized" (ibid). The next sections of this chapter therefore develop the analysis of the role of high-stakes educational assessments not just in terms of classification or managing those who are directly subjected to its force, those who may be termed as being the examination candidate, but more widely as a tool to hold a system of education to account.

4.5 1943 to the present: the identification of three epistemes

Using the approach discussed above, I have identified three periods of time or epochs (what Foucault refers to as epistemes), each with what will be shown to hold attributes distinct from other periods.

4.5.1 Episteme 1. The war years and the post war ambition for education (1943 to the mid 1950s)

At the time of the initial period of interest to this study, the centrality of the role of the teacher in conducting the assessment of pupils was a given and largely uncontested facet

of the education system in England. Teachers were in effect trusted to conduct most if not all aspects of educational assessment used in the English education system. Despite their otherwise devastating effect, the war years produced a focus on the role of education in building a more egalitarian post-war system of education culminating in the 1944 Education Act which set in train what some have described as the most significant changes to the education system made by one act of law:

> It is a very great Act, which makes – and in fact, has made – possible as important and substantial an advance in public education as this country has ever known (Dent, 1962, p. 3).

Indeed the White Paper published in July 1943 leading up to the Act was described by The Times as a 'landmark in English education';

> The Times (July 24th 1943) not unjustly called it a landmark in English education and said that it promised "the greatest and grandest educational advance since 1870" (Barnard, 1966, p. 294).

The White Paper was widely supported by Parliament, which in the context of a war-time coalition government, is not inconceivable:

> In two days' debate, on educational reconstruction the House of Commons "showed itself of one mind to a degree rare in Parliamentary annals…Not a single voice was raised in favour of holding up or whittling down any one of the proposals for educational advance" (The Times, July 31st, 1943) (Barnard, 1966, p. 295).

The impact of war on the nation's thinking about education should not be underestimated as thoughts turned to post-war provision:

> It is not without significance that that the Education Acts of 1870, 1902, 1918 and 1944 were passed in a time of war: and it would seem that men's minds, in a revulsion against the folly and waste and false values of war, turn to education as the one hope for the future – though there are not wanting those also who are interested in education primarily as a means of promoting military efficiency (Barnard, 1996, p. 293).

Writing in 1943, Harold C. Dent (editor of the TES, 1940-1950) provides a wider insight into the impact of war on English education and the broader level of support for the proposals in the White Paper of 1943 through recourse to newspaper reports, government papers and memoranda published by a range of organisations including teacher associations. However, Dent also notes some key areas of differences. For example, in reference to the re-design of the system of secondary education and the process of allocating pupils, Dent (1943) notes:

> There is divergence of opinion as to whether these courses should be provided in a single school (the multilateral school) or in separate schools, e.g. grammar, technical, commercial, modern…Opinion differs as to what precisely should make up the basis of allocation, but there is a very general feeling that much reliance should be placed on the record of the pupil's work and progress in primary school. If allocation to secondary education takes place at 11+, the curriculum in all secondary courses should be the same for two years, and there should be a re-allocation of pupils at 13+ (p. 210).

This same level of support expressed in the debates on the White Paper carried the Education Bill through Parliament resulting in the 1944 Education Act that established for the first time a government Minister of Education (and a Ministry of Education). This alone brought significant government control of the state system of education in England leading Sir Richard Denman, the Member for Leeds Central, to remind the House during the debate on the White Paper that:

> Though we have faith in the President of the Board of Education and his Parliamentary Secretary, it is still the fact that we are giving the State immense power, and that is something which should make us regard with open eyes the road we are taking (Hansard, 1943).

Despite such reservation, the 1944 Education Act received broad support including that of the teaching profession. For example, the NUT noted the passing of the Act as its; "…finest hour on a number of accounts" (Cloake, 2017, p. 37) but noted reservations around the introduction of a tripartite system. Nevertheless, at its 1949 Annual Conference, a resolution was passed calling for more rapid implementation of the Act, particularly around the building of new schools and colleges (ibid). However, in a forward in a Parents' Guide to the Education Act (Alexander, 1946), R. A. Butler, president of the Board of Education, noted:

> The overwhelming complications over priorities in building construction must mean
> that the Act will, in the words of the present Minister of Education, 'take a whole
> generation to bring into force' (p. v).

The 1944 Act brought in a range of changes to the education system including free secondary schooling for all up to the age of 15, with a further goal of raising of the school leaving age to 16 at such time as the Minister deemed it as being 'practicable'. It included the re-structuring of Local Education Authorities and the introduction of three progressive stages of education to be known as primary education, secondary education, and further education. There was also a requirement for every Local Authority to submit a development plan for Ministerial approval that demonstrated how the local arrangements for schools met the requirements of the Act that would not be deemed sufficient:

> …unless they are sufficient in number ."character, (original punctuation) and
> equipment to afford for all pupils opportunities for education offering such variety
> of instruction and training as may be desirable in view of their different ages,
> abilities, and aptitudes (Education Act 1944c. 31 Part 11 Section 11).

Local authorities reacted to the requirement to submit their development plans with some speed in order to designate land on which to build new schools as in the extract taken from the Minutes of Huddersfield County Borough Council General Purpose (Education) Sub-Committee in December 1945: Figure 4.1. Even so, it was the mid-1960s before the schools were opened, and by 1973, they had changed designation to become comprehensive schools and sixth form colleges. However, it is important to note that the Act did not introduce the 11 plus test or require the establishment of grammar, secondary modern or technical schools. Local authorities were at liberty to devise local arrangements through their development plans in order to meet the requirements of the Act specifying the arrangements for primary and secondary schools. Despite not being stated in the 1944 Act, the formation of secondary education and the eleven plus became key outcomes of the legislation and central to considerations of the role of educational assessment and the part played therein by teachers. Neither did the Act mention the role or influence of the headteacher but as stated by Dent:

> …in practice, well understood and accepted powers of control and direction are
> vested in bodies and individuals at all levels in the educational system. In some

cases these powers have no sanction whatever in law. An outstanding instance of this is the Head Teacher (1963, p. 66).

This view more closely reflects Foucault's view of power as a dispersed entity rather than located in one powerful and coercive institution (Allen, 2012a).

Figure 4.1: Extract from the Minutes of Huddersfield County Borough Council General Purpose (Education) Sub-Committee, 12th December 1945
(Huddersfield County Borough Council Minutes, p. 259)

GENERAL PURPOSES (EDUCATION) SUB-COMMITTEE.
WEDNESDAY, 12th DECEMBER, 1945.

PRESENT—Alderman Smailes in the chair. The Mayor; Alderman Meadows; Councillors Day, Haywood, Kahn, Smith.

141. Chairman and Deputy Chairman.
RESOLVED—That Alderman Smailes be Chairman and the Mayor Deputy Chairman of this Sub-Committee for the municipal year 1945-46.

142. Development Plan.
(a) County College.
RESOLVED—That the Authority be recommended to reserve the Greenhead School and grounds for county college purposes.

(b) Secondary Schools.
RESOLVED—That the Authority be recommended to provide secondary schools of 3 types, grammar schools for each sex separately, a technical high school (mixed) and mixed modern secondary schools.

(c) That the sites now discussed at Deighton, Fartown, Crosland Moor, Taylor Hill, Hall Bower, Rawthorpe, Waterloo and Salendine Nook be approved for secondary school purposes, that sketch plans be submitted to the Town Planning and Estate Committees for delimination of boundaries and for reservation of these sites for education purposes, and that the necessary steps to adjust and extend the site at Salendine Nook be undertaken.

(d) That further inquiries be made about the site now mentioned for technical high school purposes.

4.5.1.1 The formation of secondary education

The decision to distinguish more clearly between primary and secondary phases of education was established as early as 1926 with the publication of the Hadow Report. This report recommended secondary education for all children, the idea of 'modern secondary schools' to provide an alternative more 'practical' and 'realistic' education as an alternative to the academic course offered by grammar schools, and the raising of the school leaving age to 15. The latter was to be achieved by the provision of maintenance grants to pupils over the age of 14. However, by the time the Report was published changes at ministerial and government level led to the proposed legislation running into trouble over its open-ended financial commitment and disagreement amongst religious denominations (Maclure, 1979, p. 80).

The Report was nevertheless accepted by the Board of Education (without commitment to a date) and paved the way for the tripartite organisation of secondary schools (ibid., p. 80) a system of demarcation, labelling and hierarchy and access to different levels of knowledge and recognition. Changes in government and the role of agencies such as the Board of Education demonstrate how policy reflects the political rationality of a given time (Bailey, 2016). One further key influence of the Hadow Report was to establish the significance of the age of eleven years as a point of change in educational provision.

> There is a tide which begins to rise in the veins of youth at the age of eleven or twelve. It is called by the name of adolescence. If that tide can be taken at the flood, and a new voyage begun in the strength and along the flow of its current, we think that it will 'move on to fortune'. (Hadow, 1926, p. xix).

This conclusion was reached on the basis:

> …of recent psychological research, and with the existing practice of legislators and administrators. The need for a fresh classification of the successive stages of education before and after the age of 11+ (Hadow, 1926, p. ix.)

Hadow was however at pains to stress that:

> Attention should be called to the fact that the expression 'age of 11+' is not intended to be used in a precise chronological sense. The mental as well as the chronological age of the pupil must be taken into account (Hadow, 1926, p. 71).

The Hadow Report also advocated the use of an 'entrance' examination on the lines of what were at the time tests to determine scholarship awards and free places into grammar schools in order to determine which form of secondary school was most appropriate and a final 'leaving examination' to certify achievement of both selective and non-selective 'modern' schools. Unlike the more academic examinations available to grammar school pupils, these newly formed examinations would provide modern schools with a clear purpose and "…a definite standard to guide their work" (Hadow, 1926, p. xx). However, this examination:

> …need not be influenced to the same extent by the requirements of an external examination, and the teachers will accordingly be free to frame courses in the

several subjects of the curriculum (with some bent in many cases towards agriculture, commerce or the local industry or group of industries) (Hadow, 1926, p. 87).

It was clear from this recommendation that such an examination should be voluntary and could be used as a form of encouragement for further study post the age of 15 should circumstance permit. A further report of the Consultative Committee for the Board of Education on the Primary School (1931), also under the chairmanship of Hadow:

> …recapitulated psychological research which suggested that the difference in intellectual capacity between the brightest and dullest children made it necessary to classify a single age-group in several sections (streams) by the age of 10 (Maclure, 1979, p. 188).

In 1938 the Spens Report on Secondary Education with Special Reference to Grammar Schools and Technical Schools, reiterated the psychological research noted by Hadow and 'summarised the consensus of opinion at the time about the validity and usefulness of intelligence tests' (Maclure, 1979, p. 193). It was hailed by Spencer Leeson, Headmaster of Winchester College, who described the report as pointing to changes; "…almost as significant in administration as in educational theory and practice" (Leeson, 1939, p. 63). In particular, the Report promoted the use of intelligence tests noting:

> Intellectual development during childhood appears to progress as if it were governed by a single central factor, usually known as 'general intelligence', which may be broadly described as innate all-round intellectual ability. It appears to enter into everything which the child attempts to think, or say, or do, and seems on the whole to be the most important factor in determining his work in the classroom. Our psychological witnesses assured us that it can be measured approximately by means of intelligence tests…but this is true only of general intelligence and does not hold good in respect of specific aptitudes or interests (Spens, 1938, pp. 123-124).

The Report recommended three types of secondary schools with a target of approximately 15 per cent of pupils having access to a grammar school education, subject to local provision. At the time, such access in England varied by locality from 4.2 to 26.4 per cent (ibid., p. xxxiii). Technical Schools, with equal status to the grammar schools, were also envisaged for which a selection process should be used including reaching the necessary

standard in the examination, parental choice, a report from the headteacher of the primary school and an interview of the child and parents with the headteacher of the Technical School and a representative of the local education authority. The interview was viewed as being of 'considerable importance'. It was further suggested that the curricula of the different types of school should follow a similar pattern for the first two years to allow for possible transfer at the age of 13 for those children who might have been misclassified at the age of 11 or have shown 'later development' that makes clear the need for an alternative form of secondary education (ibid p. xxxiii). The Report however dismissed the provision of what was described as 'multilateral schools' – effectively three streams in one institution, on the grounds that they would be too large, too expensive and that it would prove difficult to find headteachers capable of developing and inspiring the varied curricula across streams. The White Paper on Educational Reconstruction (1943) drew on Hadow and Spens proposing that:

> …in the future, children at the age of about 11 should be classified, not on the results of a competitive test, but on an assessment of their individual aptitudes largely by such means as school records, supplemented, if necessary, by intelligence tests, due regard being had to their parents' wishes and the careers they have in mind (White Paper, 1943, p. 7).

The need for parity of esteem between the different forms of secondary provision was also expressed in the White Paper to counteract the prestige accorded to secondary grammar schools; "…in the eyes of parents and the general public which completely overshadows all other types of school for children over 11" (Para.27, p. 7). However, the ambition of parity of esteem was somewhat at odds with a system generally viewed as being hierarchical and predicated on the classification and control of pupils through the use of intelligence testing and school records.

4.5.1.2 The Norwood Report

The most explicit expressions of a tripartite system of secondary education and linkages between the curriculum and assessment were expressed in the Report of the Committee of the Secondary Schools Examination Council on Curriculum and Examinations in Secondary Schools, 1943 (The Norwood Report). The Report followed some of the themes set out in previous reports in proposing grammar, technical and modern secondary schools from the age of 11 determined by a form of selection dependent on the judgment of primary school teachers "…supplemented, if desired, by intelligence and other tests" (Norwood. p.

17) with due regard to the wishes of parents and pupils. A further opportunity for selection should be made available at age 13 which required the formation of a common approach to the curriculum in what Norwood described as the 'lower school' stage of secondary education. Norwood proffered "…that the curriculum must fit the child, and not the child the curriculum" (p. 9) noting that:

> …transition from primary to secondary education is not a break but a process in which special interests and aptitudes have further opportunity of declaring themselves and of meeting with appropriate treatment. Only on some such reorganisation of secondary education can the needs of the nation and the individual be appropriately met (p. 15).

Norwood did however raise the prospect of multilateral schools rather than three separate institutions noting that although the Technical Secondary School should be separate due to the need for linkages with local industry and commerce, there may be some benefit of scale in placing grammar and modern schools under one roof concluding that:

> We envisage therefore that within the limits of these circumstances experiment will be made with a two-type school (p. 19).

The process of selection or differentiation for the purpose of deciding the most appropriate secondary school, was raised as being problematic in the Report but that it should be based on the judgments of teachers determined by school records produced throughout the primary phase of education compiled by teachers who knew and had taught the child:

> Briefly, by a school record we mean a history of the child's development compiled by teachers who have known and taught him; it would contain an objective record of progress, with notes on special circumstances deserving to be taken into account; it would thus furnish a progressive judgement indispensable for decision as to his most appropriate education in the future, and it would guide those who were later charged with that education. Such a record, compiled by teachers trained to observe and to reflect upon their observations, we regard as the best single means at present available of discovering special interest and aptitude and general level of intelligence (ibid., p. 17).

The use of supplementary information derived through what Norwood described as; "'…intelligence' tests, 'performance' tests and the like" (original punctuation: p. 17) was

presented as an option, but with reservation. The Report noted difficulties in devising tests of this kind but acknowledged more recent evidence of their success, though nevertheless concluding that:

> We suggest therefore that differentiation for types of secondary education should depend upon the judgement of the teachers in the primary school, supplemented, if desired, by intelligence and other tests (ibid., p. 17).

The Norwood Report is notable for its focus on examinations in what was then the proposed formation of a secondary phase of education. The Report laid similar emphasis on the input of teachers to the design and award of qualifications. The Report reviewed existing arrangements for examinations for pupils aged 16 years, the School Certificate and at 18 years, the Higher School Certificate, which in effect were only for those pupils who exceeded the school leaving age of 14 and attended grammar schools. The examinations were taken at the end of the main school course and at the end of the sixth form respectively and were externally set by eight university based examination boards. The Certificate was originally designed to include a written statement signed by the Headteacher of schools "…recognised as efficient by the Board of Education" (Norwood. 1943, p. 27). However, the Report stated that; "…the record of the course pursued at school has now disappeared from the certificate" (ibid., p. 27). The Certificate had two stated purposes: one, to test achievement of the general course of education before a more specialised course in the sixth form, and secondly; "…that success in the examination should under certain conditions qualify candidates for entrance to universities and to professions" (ibid., p. 27). Since the inception of the Certificate, changes had been made to accommodate groups of subjects to reflect changes found in the curricula of schools and to thus continue its matriculation format. The report noted:

> …the degree of importance which the School Certificate has assumed in the view of the schools and the public, for it is implied that whatever curriculum is found suitable for individual pupils shall culminate in a certificate awarded by a University Examining Body, because that certificate has come to be thought of as indispensable to the pupil and the public and to the teacher and his subject (ibid., p28).

The Higher Certificate was designed on the basis that very few candidates would take the examination and that links with the universities in designing the qualification would be

close so as to be used for the award of scholarships awarded by the state and local authorities. Over time, the Report noted, that as the number of candidates had increased, the qualification had moved its function from one of qualification to one of selection or competition. Thus the Report noted that over time the purpose or emphasis of examinations had moved:

> The examinations clearly have moved much from their original purposes, or perhaps more accurately there has been a change of emphasis. Instead of assessing the normal work of a form which has completed a stage of education, the examinations are now a matter of supreme importance to each individual child. What concerns us is to discover how the pupil and the teacher are affected by the atmosphere of examinations which this review suggests as pervading the school at all stages, till it may become the life-breath inspiring all effort (Norwood, 1943, p. 30).

The Report raised concern over the way the examination had become the driver rather than servant of the curriculum by becoming "…the life breath inspiring all effort" (Ibid., p. 30) causing a narrowing of the range of subjects covered and more stress on pupils and teachers. The Report presented a balanced argument for these examinations noting that they provided an externally set and universal standard. This, the Report noted, was of particular importance for those schools "…which have not yet acquired a reputation as high as the longer established schools" (ibid., p. 31). Further, the examination syllabus provided teachers with:

> …a sense of standard; he becomes acquainted with achievement elsewhere as assessed under similar conditions, and in the light of it he can estimate the success of his work (ibid., p. 31).

This externality also provided schools with a measure of comparison against other schools reflecting the wider needs of society – a process of ordering and ranking:

> The test and the verdict must be objective, and conditions must be equal; there can be no prejudice and no favouritism as between school and school or pupil and pupil. Employers, parents and Professional Bodies need the Certificate; employers ask for a disinterested assessment, and would not be satisfied with a Head Master's certificate; parents look for something which will be a hall-mark of their children, valid wherever in the country they may go (ibid., p. 31).

However, Norwood presented a counter argument for the use of externally set examinations. The Report proffered that the external examination restricted the freedom of teachers, restricted the breadth of the curriculum and that:

> Pupils assess education in terms of success in the examination; they minimise the importance of the non-examinable and assign a utilitarian value to what they study. They absorb what it will pay them to absorb, and reproduce it as second-hand knowledge which is of value only for the moment (ibid., p. 31).

The Report also noted what it referred to as 'illusory uniformity" in terms of standards and that pupils were rendered to being no more than an examination number stating that:

> No one can examine better than the teacher, who knows the child; and a method of examination by the teacher, combined with school records, could be devised which would furnish a certificate giving information of real importance to employer or college or profession, and yet would preserve intact the freedom of the school and would rid teacher and pupil of an artificial restraint imposed from without (ibid., p. 32).

In balancing the arguments for and against externally set examinations, Norwood pointed to the success of the School Certificate in establishing more certainty over expected standards of performance. At a time of rapid expansion in the education system, this was viewed as a crucial aid in assisting in particular newly qualified teachers to gain an understanding of the aims and expectations of the curriculum, methods of delivery and a 'sense of direction'. Nevertheless, Norwood concluded that; "…its very success has tended to bring about its progressive disintegration" noting that the certificate had; "…gained more significance than was ever intended at the outset" (p. 32). Given that teachers could no longer be described as being inexperienced, there was little to commend the continuation of an externally administered School Certificate and that the teaching profession should be free to 'enhance its prestige':

> The examination in its present form is having a cramping effect upon the minds of teachers and pupils. On this our evidence leaves no room for doubt (Norwood, p. 33).

In reality, Norwood addressed what became a constant tension over the control of tests and examinations between the state, the universities and teachers by attempting to; "…emphasise the enhanced professional role of school teachers in the new secondary education" (McCulloch, 1993, p. 131). Norwood supported its conclusion by reference to the evidence taken from employers and professional bodies that 'a certificate' gave insufficient information on the pupil as; "…a human being rather than an examinee" (ibid., p. 33). The Higher Certificate Examination drew less pronounced argument either for or against its current formation. The Report noted that it affected fewer pupils but that over time it had developed from an examination of a two-year sixth form course into a more selective tool for the allocation of State Scholarships and local education authority grants. This duality of purpose had resulted in a narrowing of specialism in the sixth form, unnecessary pressure on some students and that:

> …a written examination taken by itself is not the best method of selecting those students who can best profit to the full from a University Course (ibid., p. 34).

However, Norwood recommended a broadly similar format of examination for the Higher Certificate although there was a call for more direct contact between schools and college examiners leading over time to the inclusion of teachers on the university examining bodies. Examination papers should continue to be set by university examination bodies but papers used to award scholarships should be "…free from prescribed books or detailed syllabuses of work" (ibid., p. 39). From this examination, two categories of students would be identified: those capable of obtaining a first class or second class honours degree and those who might be considered if there was further supporting evidence from the school record and an interview. However, more weight would be ascribed to the school record. A further School Leaving Examination was also recommended to be taken by pupils at the age of 18+. Its purpose would include potential as a university entrance examination, exemption from professional entry examinations, entry to training colleges or show that students had 'pursued a course of Sixth Form work with profit' concluding that:

> …this examination should be conducted by external agencies because it is intended to look forward to further study in places other than the schools (ibid., p. 41).

Suggested reforms to the School Certificate were more fundamental with a recommendation that staged moves toward the final end point should be implemented. Central to the proposed approach was a gradual move away from reliance on externally

set examinations. An examination was envisaged, but one set by teachers and subordinate to the more comprehensive school record. The Report noted that not all teachers favoured such an approach and that:

> The public too may not be fully prepared for immediate change, though we think that there is a considerable volume of opinion to which change would be acceptable. But it may be the part of wisdom to take one step at a time, provided always that the ultimate objective is kept steadily in view (ibid., p. 45-46).

A set of stages was provided in the Report. Noting that only a 'small percentage' of students taking the School Certificate go on to university and that the examination is usually taken two years before the usual age of entry into university:

> The direction therefore which change should take is sufficiently clear: it is towards placing the conduct of the examination in the hands of the teachers; they alone can best judge the needs of the mass of their pupils and they ought to be the best judges of the success or failure of the methods they employ (ibid., p. 46).

A further justification was raised.

> Change in this direction is indicated by yet another consideration. If the present School Certificate examination is retained without alteration or prospect of alteration, it will mark off the secondary Grammar School from other forms of secondary education. A system will then be established under which parity in secondary education will become impossible. For the objective of the School Certificate has become so associated in the public mind with secondary education that the establishment of the Technical School and the Modern School as forms of secondary education will be prejudiced from the outset (ibid., p. 46).

Norwood envisaged a period of seven years to make the transition from a university examination board set examination to one that was wholly teacher set at which point the connection would be severed. In the interim university examination boards would be required to establish a Standing Committee:

> …consisting of eight teachers, four members of Local Education Authorities, four University members, and four of HM Inspectors acting as assessors. These

Committees would report to the Secondary School Examinations Council and to the respective Examining Bodies (ibid., p. 46).

Two further recommendations were made; one to encourage schools to offer their own syllabuses and two; the introduction of subject examinations awarded with grades rather than the requirement for matriculation based on achievement across set groups of subjects. These recommendations were to be supported by investigations into the best methods of producing school records and the provision of short courses, conferences and school visits for teachers "to devote attention to their compilation" (ibid., p. 47). Although the Norwood Report focused on secondary school examinations primarily in grammar schools, reference was made to the suitability of children for such a course at the age of 11 noting here again that the predominant determinant should be the opinion of the teacher "…supported or qualified by whatever tests might be though desirable" and that "…parents concurred in these recommendations" (ibid., p. 48). The examination at 11+ is discussed later in this chapter in section 4.5.1.3.

The Norwood Report brings into relief the level of trust assigned to teachers in determining and administering educational assessments. As discussed later in this chapter, whoever controls the means of assessment holds considerable power. Norwood was a clear advocate of teachers as professionals, but noted that:

So long as they accept external control as to what they shall teach and external assessment of the way in which they have taught, they can never rise above the rank of journeymen (Norwood, 1942b. cited in McCulloch, 1993, p. 134).

However, Norwood noted that such status would require teachers to take more responsibility and face up to the task and be supported by appropriate training if the teaching profession was ever to become a self-governing profession akin to the status enjoyed by doctors and lawyers: " In other words, by asserting their own control in this sphere of their work, they would be able to acquire a greater degree of professional status" (McCulloch, 1993, p. 134). Despite the enthusiasm for greater teacher control of the examination system shown by Norwood, this was not echoed by the university examination bodies that clearly had a vested interest. This was illustrated by the then assistant Secretary to the Cambridge Local Examinations Syndicate J. L. Brereton, who was highly critical of Norwood's proposal to replace external assessments with internal examinations: "…a step which I believe will allow arbitrariness, favouritism, and patronage to raise their ugly heads again" (Brereton, 1944, p. 187). Brereton presented The Case for

Examinations (1944) "…because many people now regard them as a necessary evil" (p. vi). See Figure 4.2.

Figure 4.2: The Case for Examinations. J. L. Brereton, published 1944
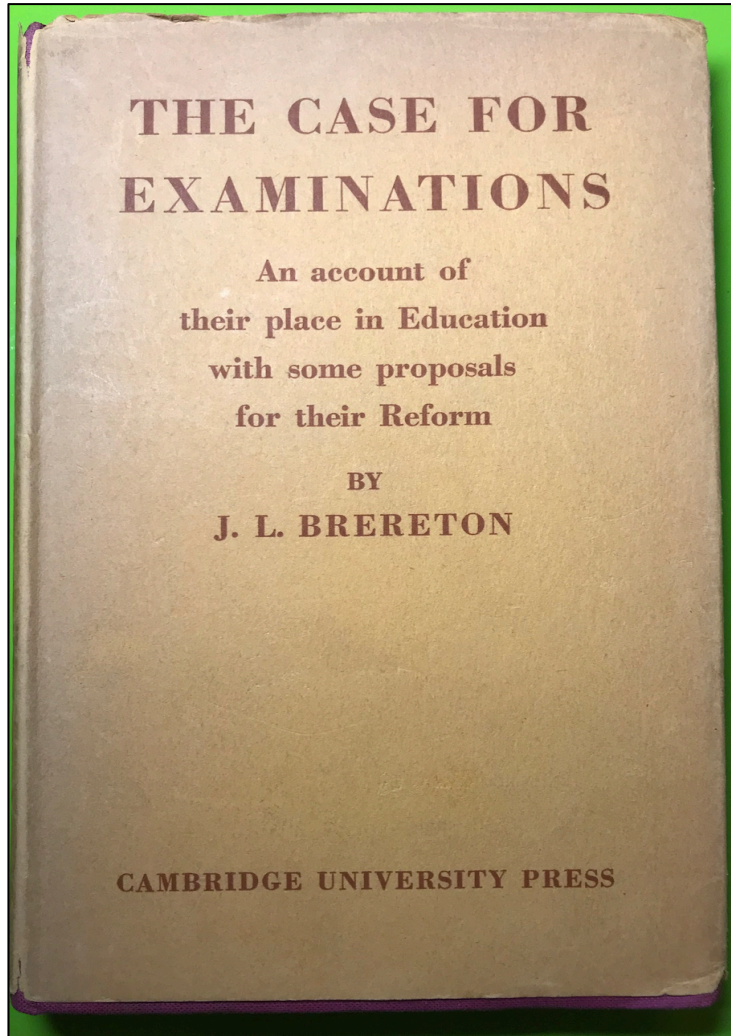
Perhaps of greater pertinence was the reaction of teachers:

> It is noticeable also, however, that teachers' groups were at best lukewarm and often actually hostile in their reactions, and this raises issues about their own perceptions as to the proper role of the teacher (McCulloch, 1993, p. 136).

McCulloch provides the examples of Dr. Terry Thomas, Head of Leeds Grammar School – and member of the Norwood Committee – and Miss W. M. Casswell, Headmistress of Edgbaston High School as defenders of external examinations. Norwood attributed Thomas's dissent to pressure from the Association of Headmasters. Further cautious

opposition came from subject associations such as the Society for Education in Art and the Modern Languages Association who asserted that opposition to external examinations stemmed mainly from those unable to pass them (ibid., p. 137). This opposition reflects what Foucault describes as "Where there is power, there is resistance" (1978, p. 95). Drawing on Foucault Schirato et al (2012) note:

> …that we can see that the flows and manifestations of power are always both disruptive and productive; they produce a series of transformations, realignments, movements and responses that are sometimes radical and dramatic (the French and Russian Revolutions being obvious examples), but generally they are, to use Bourdieu's expression, 'more or less invisible to the cultivated eye' (1990:155) (p. 63).

Foucault posited power "…is not 'held' by any one but, rather, is 'capillary' and therefore flows through all interactions and relations" (Hodges et al, 2014, p. 565). Resistance therefore does not necessitate the formation of an alternative regime but rather constitutes a flow of interactions varying in intensity and prominence.

4.5.1.3 The 11+: selection and competition

The Norwood Report and the Government White Paper (1943) had laid the foundations of the 1944 Education Act that required Local Education Authorities to provide education at secondary level that matched the "abilities, and aptitudes" (Education Act, 1944. Section 8.1.b) of pupils: this became a key determinant of the way educational assessment was used as a means of allocating pupils to schools through what became generally known as the 11+ examination. However, the Act was silent in this regard although Norwood clearly preferred a system based on teacher judgments and school reports supported where necessary by tests. The utilisation of objective and standardised tests was supported at the time of the enactment of the Act by an:

> …orthodoxy that intelligence was measurable by psychometric tests, offering 'a neutral means of assessing the aptitudes of children from deprived backgrounds and of allocating them to appropriate schools' had, by this time, dominated a generation of educational thinking (Whitty & Power, 2015, p. 2).

The education system was also under considerable pressure in terms of providing sufficient and appropriate schools, but although the 1944 Act did not suggest the use of

tests at age 11 as a method of selection, The Nation's Schools, a pamphlet published by the Ministry of Education in 1945 gave a strong steer:

> Methods of selection have not yet been so perfected that it is possible in any way to dogmatise about the best means to employ: those used in the past have often been criticised on the ground that they led to cramming and special coaching; distorted the curriculum in the junior schools, and were unreliable. Further research and experiment are very necessary, but it seems certain that in the meantime no authority can afford to neglect any of the following methods - careful records of the pupil's progress, interests and aptitudes; general objective tests of intelligence and aptitude used, if possible, more than once during the child's primary school years; enquiry into discrepancies between performance in these and school progress and comparison with performance in other tests. These records and tests should make possible a fair assessment of the child's capacity so far as it has developed by the end of the junior school course (p. 27, para. 100).

The pamphlet also raised the need for the training of teachers in producing reports and called for close collaboration and knowledge sharing across junior and secondary schools including details on the content of the curriculum, expected standards and career paths of pupils (ibid., para. 101). Such advice clearly influenced local education authorities as can be seen from the following extract (Figure 4.3) of a draft circular on the Special Place Examination from an un-named Education Department at the time.

According to Gillard (2017), The Nation's Schools received a hostile reception and was withdrawn:

> But as a result of intense pressure from education officials, the policy remained the same and was restated in The New Secondary Education, 1947 (ibid., Education England on-line).

Figure 4.3: Special Place Examination. Un-named Education Department Draft Circular, 1945.
Accessed through the National Archives on-line at:
 https://www.nationalarchives.gov.uk/wp-content/uploads/2014/03/ed147-1341.jpg

3. The following general principles may be accepted:-

(a) Carefully devised objective and standardised tests are needed to assess intelligence and to indicate aptitudes with a sufficient degree of accuracy:

(b) tests must not be such as could dominate or even materially influence curriculum and methods in the primary school, the aims of which must be the full and appropriate development of the personalities and capacities of individual children up to the age of 11:

(c) weight should be given to the judgments of teachers supported by carefully maintained records of individual children.

For the present at any rate some form of qualifying examination will often be necessary and, where the provision of places of a particular type is not adequate, such an examination inevitably becomes competitive. Further research into the process of selection is, however, clearly necessary and is already being undertaken. Local education authorities are asked to experiment and to keep careful records of the subsequent performance of pupils. The Minister proposes to publish from time to time such information as becomes available for the assistance of local authorities and he will also arrange for the problems to be discussed with teachers and others at short courses.

In the New Secondary Education (1947) published by the Ministry of Education to replace The Nation's Schools, testing at age 11 was not specified but the use of intelligence tests as a means of grouping children in secondary modern schools in rural settings was:

> For this purpose teachers have found intelligence tests to be of considerable value, though classification based on such tests should always be checked, and rectified where necessary, in the light of the subsequent performance of the children (p. 44).

However, it was noted in New Secondary Education that:

Some local education authorities and schools have experimented already in methods of recognising and assessing various kinds of aptitude, by including particular tests in their schemes of selection and by systematically recording the "follow up " of school results and after careers. There is room for a great deal of serious investigation along these lines (p. 45).

The publication did note that the essential consideration in any selection process is not the 'intelligence quotient', but more in "…the natural bent of their minds and their outlook, and that of their parents, on their own future" and noted that the intention of the 1944 Act was to provide equitable education for all and that all pupils, including the "brightest and ablest" should attend the schools of their choosing (ibid., p. 45).

As noted earlier, the lack of appropriate school places imposed restrictions on the numbers of pupils being allocated to desired placings leading to wide variations across England's local authorities in the number of available grammar school places that ranged from 10% to 45% (Burt, 1957). Even within an area, there were wide variations; for example in towns in Buckinghamshire, availability ranged from 30-50% whereas in country schools, it could range from 2.5% to 10% (Times Educational Supplement, June 28, 1957). Indeed some commentators have noted that the 1944 Act "…created little that was new: it refined an existing, highly meritocratic system" (Thom, 1986, p. 101) and the principles underlying selective secondary schools were by the late 1950s under challenge as a:

…major cause of 'social waste', as it advantaged the children of middle-class parents and was an impediment to equality (Whitty and Power, 2015, p. 12).

Of further concern, the selection approach was entirely dependent on local authorities with variations in the use of tests, teacher judgments and parental choice. Even so, some form of test was used in the selection process across all local education authorities. Table 4.1 provides a comparison of the tests adopted in English local authorities between 1952 and 1957.

Table 4.1: Percentage of Local Education Authorities Using Standardised or Un-standardised Tests in 1952 and in 1956
(Adapted from Yates & Pidgeon, 1957, p. 28)

Tests Used	County Councils (45)		Borough Councils (71)	
	1952	1956	1952	1956
Standardised intelligence	82.2	91.1	93.0	94.4
Standardised Arithmetic	66.7	73.3	77.5	78.9
Standardised English	64.4	68.9	78.9	78.9
Un-standardised papers in Arithmetic, English or Intelligence	51.1	46.7	31.0	33.8
Standard Battery, Arithmetic, English and intelligence only	48.9	53.3	62.0	60.6
Combination of standardised and unstandardized tests	35.6	37.8	25.4	28.2
Un-standardised tests only	15.6	8.9	5.6	5.6

However, other methods were used in the selection process. For example English essays were used and showed an increase between 1952 and 1957 (Table 4.2). There was also greater use made of teachers' assessments for orders of merit and grading (Table 4.3) and a decline over the same period in the number of authorities stating that they did not make use of assessments provided by primary school headteachers from 26.2% in 1952 to 1.6% (two authorities) in 1956 (Yates & Pidgeon, 1957, p. 29).

Table 4.2: Percentage of Local Education Authorities Using an Essay at some stage of their procedure, in 1952 and in 1956
(Adapted from Yates & Pidgeon, 1957, p. 28)

County Councils (45)		County Boroughs (70)	
1952	1956	1952	1956
62.2	80.0	47.1	71.4

Yates and Pidgeon (ibid) suggested that the 'marked' increase' in the use of essays:

> …can no doubt be attributed in the main to two factors: the desire on the part of the authorities to counteract the undesirable 'backwash' effects that wholly objective tests of attainment in English are said to have on the work of primary schools: and the demonstration provided by Wiseman's research that essays could

be marked with a reasonable degree of reliability (Wiseman, 1949). (1957).

Table 4.3: Percentage of Local Education Authorities Using Order of Merit and Grading in their Teachers' Assessments (1956)
(Adapted from Yates & Pidgeon, 1957. p. 29)

	County Councils (42)	County Boroughs (67)
Order of merit	61.9	58.2
Gradings	26.2	28.4
Both	11.9	13.4

In compiling their assessments, 61.5% of the authorities stated that scores from standardised tests were available to teachers before they made their final judgments (Yates and Pidgeon, 1957, p. 30). The authors also report that in 1956, 33.3% of authorities used school report cards in their allocation procedures; "…and many others stated that they consulted them for candidates in the border-zone" (p. 30). With regards to pupils falling into a 'border-zone' it was further noted that 70% (n= 40) of County Councils and 41.4% (n=70) of County Boroughs, interview children and in 21.3% of authorities interview "…persons other than children, naming mostly primary school heads and class teachers, although a number in certain circumstances see the children's parents" (p. 30). Local authorities applied a mixture of approaches to the selection and allocation process leading Yates and Pidgeon to conclude that:

> The best possible prediction of subsequent success in secondary schools was afforded by a combination of tests and assessments, the measure derived from the primary heads' judgements of the suitability of their pupils for a grammar school course was found to be the best single predictor among those that we examined (p. 77).

In reality, it is a misconception to perceive of the 11+ as a single universal examination. Local authorities were free to determine their own policies and approaches, though most utilised standardised tests developed by two main providers; the National Foundation for Educational Research (NFER) and Moray House tests developed under the auspice of Edinburgh University. An extract from the list of Authorities purchasing Moray House tests in the year ending March 1962 provides an insight into their widespread use (Table 4.4). The centrality of testing across local authorities detracted from the vision of the White Paper (1943) that advocated the use of school records, supplemented if necessary by intelligence tests (p. 7). Neither was this vision carried through to the 1994 Education Bill,

an omission that was not contested:

> The obvious explanation is that the finer points of selection were to be left to the LEAs as compensation for loss of power, because central direction was unnecessary (Thom, 1986. p. 118).

Figure 4.4: An extract from the list of authorities using unpublished Moray House tests during the year ended 31 March 1962
(Moray House Tests, University of Edinburgh Archive)

The NUT (1949) noted the reluctance of some teachers to take responsibility for the selection of pupils due to social pressure and:

> …feel that they should be protected from all charges of favouritism and from all blame if the results do not accord with the wishes of parents (p. 42).

The administration of 'centrally devised objective procedures' was viewed as a way of

mitigating such pressure, but the NUT concluded that the responsibility placed on teachers should prevail (ibid., p. 42). In practice, there was a mixed approach across local authorities. For example, a former director of NFER explained:

> Each LEA could also determine how it used its results: sometimes using two VR (verbal reasoning) tests and taking the higher score of the two; sometimes combining the scores with headteacher's ratings. The tests were marked within the LEAs. Again I think there was a variety of systems, some doing it centrally, but my impression was that it was mostly done in the schools, those being the days of trust (received in a personal e-mail, 1st October 2019).

Here we see reflections of Foucault's position that power is everywhere, wielded by many as it disperses throughout the social body - education system. Further, local authorities were free to change their approaches over time in response to local concerns.   For example the West Riding Education Authority acknowledged concern about the impact of externally set tests as a source of pressure on pupils and teachers leading to what was described as 'excessive coaching' with:

> One of its worst features is that it imposes an intolerable burden on the conscientious Head of the neighbouring Junior School. He knows that to do justice to all his pupils he should not spend undue time on the few, but if he does not his few will lose Grammar School places to less able children in their neighbouring school where excessive coaching takes place (Hyman & Clegg, 1954, p. 16).

As a result of these concerns, the West Riding conducted an experiment with Moray House. Twenty thousand pupils were entered for the County Minor Examination in 1947 and in the same year sat standardised tests set and marked in their own junior schools. Moray house was commissioned to report on which of these two methods was:

> …the more efficient in predicting future ability shown by the order of merit of the same pupils in the Secondary Schools three years later. Their verdict, shown of course, by elaborate statistical calculations, is that there is nothing to choose between the two methods of selection (Hyman & Clegg, 1955, p. 18).

As a result, the West Riding reported that from 1956, pupils admitted to grammar schools would be on the basis of results derived from tests set and marked in the schools and that the first two years of secondary education would be classed as being probationary (ibid.,

p. 18). There were other growing concerns about the influence of the 11+ in what Montague (1958) described as the 'eleven plus battle'. In 1959 Sir Cyril Burt, an advocate of the use of intelligence tests as a means of identifying what he described as the innate nature of intelligence, responded to what he described as "…a heated controversy and vigorous attack (on the 11+)…based on social and political arguments rather than on educational" (p. 99). In presenting his argument, Burt cited his research on identical twins noting:

> From these and many other lines of evidence we may confidently infer that the degree of general ability possessed by each individual is determined mainly by his genetic constitution in accordance with the almost universal laws of Mendelian inheritance (ibid., p. 108).

However, there was growing concern over the veracity of some of the claims of intelligence testing ranging from the lack of clarity in definitions of intelligence and ambiguity of interpretation (see Heim, 1954). Further, criticism of Burt's work began to emerge notably:

> …only a year after his death in 1971, evidence began to emerge that Burt was a fraudster who had simply invented results to fit his theories about the hereditability of intelligence (Socialist Review, 1996).

The use of tests to locate pupils into different institutions became an instrument of hierarchical designation. The majority of pupils were tagged by the term 'failure', one of two outcomes of the 11+, and their attendance at secondary modern schools made their location a highly visible form of surveillance (see for example Hoddle, 2006). Foucault identified the examination as a method of discipline, identification and control:

> Foucault has himself identified educational examinations (i.e. assessment) as one of the more significant disciplinary mechanisms to have emerged….By combining the two principles of hierarchical observation and normalizing judgement, the examination becomes one of the major instruments for locating each individuals place in society (Broadfoot, 1996, p. 99).

Secondary modern schools were established to be 'free' from the pressure of any external examination (Ministry of Education, 1945, para. 77, p. 21) to enable them to produce their own syllabuses and methods (Ministry of Education, 1947). Secondary modern schools though did not in practice confine themselves to an education short of external

examinations and qualifications and the Labour government Circular 10/65 (DES,1965) required local education authorities to provide plans for comprehensive secondary education, a move repeated in 1974 (DES, Circular 4/74) following a Conservative government circular in 1970 (DES, 10/70) that rescinded the 1964 Circular. However, the formation of comprehensive schooling in the mid 1970s does not in itself disassociate from the use of ability groupings, albeit determined by the school. Simon (1954) critiques the multilateral or comprehensive school, as being "schools within a school" and even in junior schools, children are streamed into classes labelled A, B, C etc:

> Penetrate within the school, examine its structure and organisation in detail, and the outward appearance of unity is at once shattered…To label a class "C" makes it fairly clear to the children what their teachers think of them. As a result, these children are not unnaturally liable to get dispirited, and to lose confidence in their own abilities (Simon, 1954, p. 10-11).

4.5.2 Episteme 2. The expansion of high-stakes qualifications (mid 1950s to late 1980s)

4.5.2.1 The General Certificate of Education

In 1951 the General Certificate of Education (GCE) Advanced (A level) and GCE Ordinary level (O level) qualification replaced the Higher School Certificate and the School Certificate respectively. The change stemmed directly from the Norwood Report, but unlike the intentions of the report, the gradual transition from an externally administered system controlled by examination boards to a system of internal examinations for the School Certificate did not materialise. And although the 1944 Education Act ignored Norwood's call for a core curriculum, other than reference to the requirement to teach religious instruction:

> Teachers' freedom in curriculum matters was in reality illusory as the content and standards of the grammar school curriculum were determined by the GCE examinations and by the entrance requirements of the universities. Curriculum control rested, therefore, with the examination boards, which were able to divide their syllabuses and examinations in a relatively unregulated manner. This was the period when the examination boards had considerable influence on the curriculum, both nationally and locally (Tattersall, 2003, p. 13).

However, the qualification regime did move from a system of matriculation based on set groupings of subjects to one where students were free to select individual subjects of choice. This should of course be qualified by the fact that students choices are in reality controlled by what any individual school may offer in terms of the selected syllabus and awarding body. The original target group for the GCE sat within the grammar school and was intended for those pupils aiming to go on to go on to university, the professions and white collar work (Tattersall, 2003). This contrasted with the clear intention that pupils in secondary modern schools would be free from the pressures of any external examination (Ministry of Education, 1945; The Nation's Schools, para.77, p. 21). GCE entry was further restricted by the requirement that pupils had to be aged 16 to enter the examination, an opportunity not available to the majority of pupils who left school at the age of 15. In practice these restrictions resulted in the proliferation of locally produced leaving certificates for pupils in secondary modern and technical schools produced by local authorities as a means of external validation, but; "Such certificates had no national currency" (Tattersall, 2007, p. 57) and further classified secondary modern pupils as 'educational failures':

> The absence of a nationally recognized examination for non-grammar school students reinforced the view that these students were educational failures (Tattersall, 2003, p. 13).

Following representations from teachers, this restrictive requirement for GCE entry was relaxed by the Minister of Education in 1955 and the number of pupils taking the GCE grew as did the prestige of the qualification - now seen by employers and professions in preference to local leaving certificates. The growth in entry was influenced to some degree by the growing school population and the numbers of pupils from secondary moderns taking the examination, though more often not in as many subjects as those pupils from grammar schools:

> As more pupils from Secondary Moderns gained O-level passes, so the underlying logic of academic selection that informed the GCE examination, and the tripartite system more broadly, began to unravel. Within a couple of years, the SSEC recommended further action, laying the ground for the appointment of the Beloe Committee in 1959 (ESRC, 2018).

4.5.2.2 The Beloe Report and the Certificate of Secondary Education

In 1958, the Secondary Schools Examination Council (SSEC) established a committee chaired by Robert Beloe (Chief Education Officer of Surrey) in response to the growing numbers of non-selective school pupils who were by the late 1950s staying on in school up to the age of 16 and taking the GCE examination. The SSEC; "…had become increasingly concerned about the rapid growth of external examinations outside the G.C.E. (original punctuation) framework" (Beloe, para.28, p. 8). In 1955, the then Minister of Education had accepted the view that those pupils attending schools other than grammar schools should be permitted to take the GCE. However,

> …he (Minister of Education) did not favour the establishment of any new general examinations of national standing for secondary schools (Below, para. 25 (b), p. 7).

> But as Beloe went on to state;

> …experiments by groups of schools in organising their own examinations were to be welcomed provided that the schools concerned retained control over their own syllabuses and courses of work (ibid., para. 25 (c), p. 7).

The Beloe Committee's remit was to consider the arrangements and 'desirable' developments for the examination of secondary aged pupils other than by the GCE (Beloe, 1958). The Report noted that the GCE O level examination had in fact raised the standard for a pass grade by making it equal to that of the credit level of the School Certificate, which it replaced, and in effect acted to disbar pupils who would have achieved a pass level under the previous system. The Report also noted that by 1959, about one-third of entrants for the GCE O level were pupils from educational establishments other than grammar schools, many of whom had already left school and were taking the examinations in further education institutions albeit in fewer subjects than pupils in grammar schools. The Report concluded that it had:

> …become apparent in the period since the G.C.E. (original punctuation) examination was introduced, and notably in the last few years, a mounting demand from teachers, parents and pupils alike for examinations of a different kind and less exacting standard. (Here, and in all that follows, we are thinking not of the purely internal examinations set by schools entirely for their own purposes, but of examinations having some degree of "externality".) (Beloe, para.33, pp. 10-11).

As a result, Beloe drew attention to the growth in examinations other than the GCE but noted their lack of uniformity. Such provision was made through examinations offered by groups of schools, bodies with regional coverage offering examinations in individual subjects or groups of subjects, further education examinations and specialist examinations offered by bodies with national coverage with a vocational or professional focus. Beloe illustrated the limitations of local provision as holding only local currency and that they were commonly aimed at pupils in their fourth year of secondary education (para. 40, p. 12). Examinations offered by bodies with regional or national coverage were noted for their wider currency and that their appeal was leading to some decline in examinations designed for further education (para.48, p. 15).

This, as Beloe reported, was in spite of the Education Minister's objections to the widespread use of external examinations as expressed in Circulars 289 and 326 (Ministry of Education, 1955 & 1957 respectively: see Beloe Report, 1960; and Aldridge et al., 1991). It also reflects how schools, teachers and others acted as seats of power in the inter-connected web of power relations. The Beloe Report was published in 1960 and concluded that:

> There appears to be a strong and growing desire, both amongst teachers and amongst parents whose children are thought to be unsuited to attempt the G.C.E. examination, to enter them for some other examination with wider than local currency (ibid., para. 65, p. 19).

The Beloe Report raised concern that as the number of examinations grew, the 'limited group' of examining bodies would hold 'decisive influence' on the development of schools (ibid., para. 66, p. 20). The Report commented on what the Committee saw as very different circumstances than those facing Norwood:

> We have had to think primarily in the context of a newly emerging pattern of secondary education, in which many teachers are still groping their way by experiment, and schools need freedom to grow. Furthermore, they were writing in a period of wartime transition when it was not easy to discern the characteristics and needs of the post-war educational scene. Coming to our problem 15 years after the end of the war and 16 years after the passing of an Education Act which gave a great new impetus to secondary education, we have had the advantage of being able to see more clearly than they could both the emerging pattern and the new needs created by the advance of secondary education on a wider front (Beloe,

para. 72, p. 22).

The Report offered alternative courses of action: to maintain the status quo, or for the Minister to take the initiative in stimulating examinations at the appropriate level, conforming to requirements of (his) choosing (para.90. p. 27): the latter was favoured by the Report as was the recommendation that:

> The teachers in the schools using the examinations must have a major role in operating them and shaping their policy. This means that the Examining Bodies must be neither so large that their administrative offices become, geographically and in other ways, remote from the schools nor so small in number of candidates, nor so located, that they fall under the more or less exclusive influence of a single local education authority (para109, pp. 31-32).

As a result the Certificate of Secondary Education (CSE) was introduced into schools in 1965 with a target group of the 40% ability range below the 20% GCE O level target group. Following the recommendations of Beloe, teachers were central to the preparation of syllabuses, setting papers and marking scripts. Further, regional awarding bodies were created to provide geographical proximity and ease of access and schools could only offer the CSE offered by their regional examination board (see Figure 4.5). The CSE is significant for a number of reasons. First, 14 newly formed regional awarding bodies that provided a publicly recognised qualification for pupils who had previously been disenfranchised carried out the administration. Second, the awarding bodies were governed largely by teachers with subject committees charged with preparing syllabuses, appointing chief examiners and assuring standards across the three modes of examination (see Table 4.4). Importantly, the membership of subject committees; "…was almost entirely made up of practicing teachers, usually elected from within each LEA of the board's region" (Gillan, 2003, p. 101): and thirdly, there was a strong focus on different forms of examination with a strong emphasis on non-academic syllabuses and assessment through coursework. Although the CSE is often recognized for its innovate approach to coursework, a pilot was undertaken by the Joint Matriculation Board in 1965 of 479 candidates in ten schools in Leeds. The experiment was for a GCE O level English qualification based on 100% coursework assessment. By 1993, the qualification extended to over 200,000 entries. The course emerged following criticism that the externally set and marked format was conducive to; "…dull and cramped teaching and to crabbed rote learning and practice" (NEAB, 1994, in Spencer, 2003, p. 121). Despite the apparent success of the coursework based O Level, there were expressions of concern regarding

the workload demands on teachers and schools in running the courses, for example in carrying out the assessments and checking of standards within and across schools, all of which (according to the awarding body) attributed to the "…initial relatively slow growth of the system" (Spencer, 2003, p. 122).

As well as the flexibility of offered by coursework, the CSE made provision for three modes of examination as presented in Table 4.4. For reference, this provision was also carried through to the GCSE from 1986 as shown in the example in Appendix 10, The Salendine Nook School Mode 3 GCSE. However, by 1997, the ever-increasing restrictions imposed by national GCSE criteria resulted in the demise of the modes approach. The reasons behind these restrictions are discussed later in this chapter.

Figure 4.5: The Regional CSE awarding bodies (Gillan, D. p. 99)
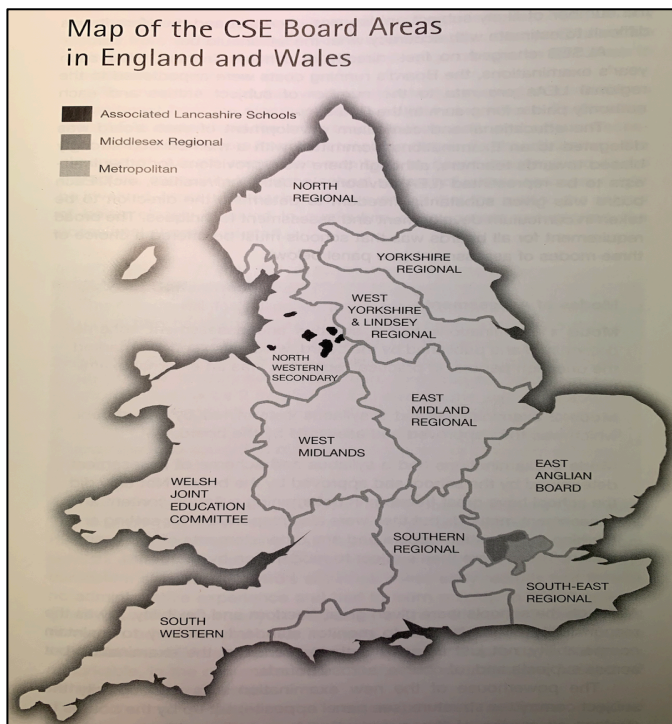
Table 4.4: Certificate of Secondary Education: modes of assessment

CSE Modes of entry	
Mode 1	The syllabus and assessment scheme set and marked by the awarding body.
Mode 2	Schools determined the syllabus. Approved and assessed by the awarding body.
Mode 3	The syllabus and scheme of assessment determined by the school and approved by the awarding body. The school was responsible for setting and marking question papers and other components (e.g. practical work). The awarding body moderated the assessments.

In terms of the power/knowledge dimension, the part played by teachers in the late 1960s and 1970s examination system in England was central with extensive use of coursework and the local production of examinations through the Certificate of Secondary Education. In this manifestation of examinations, the teacher played a central role as the expert professional: "…key constituents in the modern state…who operated on the power/knowledge cusp" (Ball, 2013, p. 15). The involvement of teachers in designing, teaching and administering assessments leading to publicly recognised qualifications from the 1960s through to the late 1980s illustrates the complexity of power relations. The expansion in the availability of qualifications across a wider constituency of the school population brought more pupils into public gaze through their certification and categorization. Here at one we see tension between the purpose of education that strives to raise equality through access to qualifications yet conversely it forms a hierarchy through a mechanism of differentiated examinations and grading systems:

> …education works not only to render its students as subjects of power, it also constitutes them, or some of them, as powerful subjects (Ball, 1990, p. 5).

It also reflects a level of ambiguity in the way teachers view the purpose and administration of high-stakes assessments; on the one hand as an expression of their professionalism,

on the other an administrative burden, a mechanism of comparison between teachers and suspicions of slight in hand in the application of standards for personal or school gains...

In 1964, the Schools Council was created with a remit to co-ordinate the burgeoning examination system. The Council was drawn from a wide constituency including local education authorities, teacher unions, subject associations and awarding bodies in what Tattersall (2007) described as:

> …a far cry from later national overseeing bodies, which, with every new manifestation, accrued ever-greater powers of control over, and regulation of, the examining system (p. 61).

The CSE was linked with the GCE O level standard by designating a grade 1 CSE pass as equivalent to an O level pass, or Grade C from 1975. This was aimed at securing public acceptance of the standard offered by the CSE. A grade 4 CSE was deemed to represent the standard performance of the 'average pupil'. In practice, this cemented the view that the CSE was:

> …inevitably seen as the poor relation of the established examination. A generation later, GNVQ and other vocational examinations would run into the same problem (Tattersall, p. 63).

This was in direct contrast to the common perceptions of the GCE ordinary level:

> Mothers talk about O levels over the garden fence. Fathers talk to fathers about them in their clubs. Offspring are classified, hallmarked and trademarked by the number of O levels they have acquired (Percy, TES, 1967).

And to compound matters, candidates receiving a CSE grade 5 (the lowest grade) were in practice labelled as being; "…below average" (Lawton, 1980, p. 100).

The establishment of coursework as a key feature of the CSE continued into the General Certificate of Secondary Education (GCSE) established in 1986 and first examined in 1988. However, there was no generally accepted rule as to what constituted coursework, for example the day-to-day work produced by students during the course or a specific piece of work designed to meet particular assessment objectives set by the syllabus (Tattersall, p. 65). This raised issues about the comparability of standards across schools.

This was managed by a system of on-site visits, postal moderation and consensus forming through formal meetings of teachers, but was again dependent on the quality and interpretations of the teachers and moderators:

> The efficacy of the system depended on the quality of the assessments of both teachers and moderators, which raised questions about the training they had undertaken in order to make valid and reliable assessments. The consensus approach was later to be favoured as educationally beneficial and effective in the report commissioned by the government, prior to the introduction of the National Curriculum, from the Task Group on Assessment and Testing (DES/WO, 1988a), but its recommendations were judged too costly to implement (Tattersall, 2007, p. 65).

4.5.2.3 The General Certificate of Secondary Education (GCSE)

By the late 1970s the GCE and CSE were established qualifications but concerns about the comparability of standards and the growth of comprehensive education raised questions about the need for a two tier system that with its different grading systems were seen as confusing to those outside of the teaching profession. There were also issues around the costs and administrative burdens generated in running two separate systems of assessment. In the mid 1970s, the Schools Council piloted joint 16+ courses and fuelled by the then Prime Minister James Callaghan's Ruskin College speech in October 1976, there was growing public interest in the school curriculum and the associated assessment regime: the first signs of the end of what has been described as the golden age of teacher control (Le Grand 1977, cited in Whitty, 2001, p. 161) leading to:

> …a shift to regulated autonomy, involving a move away from the notion that the teaching profession should have a professional mandate to act on behalf of the state in the best interests of its citizens to a view that teaching (and other professions) need to be subjected to the rigours of the market and/or greater control and surveillance (p. 161).

This becomes manifest in the third episteme presented in Section 4.5.3.

In 1978 a Government appointed Committee under the chairmanship of James Waddell recommended a single system of examination directed by nationally determined subject

specific criteria. This was followed by the White Paper (1978), Secondary School Examinations: A single system at 16 plus. The White Paper noted:

> The Government also believe it is essential that the universities and further education and representatives of employers and trades unions should be closely associated with the monitoring and co-ordination of the new system of examinations (para. 3, p. 7).

By 1984 agreement was reached on the establishment of the GCSE examination and this was first offered in 1986 with the first qualification awards in 1988. The GCSE reduced the overall number of awarding bodies by the formation of partnerships, which by the mid 1990s had formed into four GCSE unitary awarding organisations in England. A further recommendation of the Waddell Committee was the establishment of a central and strengthened body with responsibility for developing and over-seeing nationally determined subject criteria applicable to school-based and board-based examinations. Teachers were also to be seen as contributors to the system:

> The introduction of a common system is likely to involve more teachers in responsibility for assessment of their pupils' performance, and wider reliance will need to be placed on a number of alternative examining techniques already introduced in O Level and CSE, such as course assessment and practical tests, in which the teacher is often involved, as well as on the more familiar written papers. A common system will continue to provide for school-based syllabuses and many teachers will want to maintain their involvement with syllabus development (Waddell, 1978, para 114. p. 36).

The GCSE carried forward the use of coursework, a key feature of the CSE examination but much less so a feature of the GCE, and: "Unlike the GCE O level and CSE examinations, it (the GCSE) has not been designed with any pre-determined percentage of the ability range in mind" (Mobley et al., p. 12). The format of GCSE examinations was built around subject based syllabuses with differentiated examination papers to provide; "…alternative examination papers and tests at different levels of difficulty" (White Paper, 1978, para.3. p. 3) so as to fit; "…better with comprehensively organised secondary education and assist individual schools to use their resources, not least teachers, to the better advantage of their pupils as a whole" (ibid., para.2 p. 3). Further, all syllabuses had to conform to a nationally determined set of general criteria, with a second set of subject

specific criteria covering twenty subjects which covered around 85% of GCE O level and CSE entries (Mobley et al., p. 31).

It is worthy of note that the development of the GCSE should be seen in the broader educational developments of the time leading up to the formation of the National Curriculum established by the 1988 Education Reform Act. In reality, the seeds of the Education Reform Act had been cast in the late 1970s in what has been described as the Great Debate on Education (see for example Chitty, 1991). Concerns over what had become to be seen as an educationalist managed system of education hidden from public view, or the "secret garden" (ibid) were raised by as early as 1976, by the then Prime Minister James Callaghan who cited the; "contentious issue of examinations" and referred to calls for him to; "…keep off the grass":

> It is almost as though some people would wish that the subject matter and purpose of education should not have public attention focused on it: nor that profane hands should be allowed to touch it (Callaghan, Ruskin College, 1976).

And although much of the focus at that time was on the content and structure of the school curriculum, the associated structure of assessment was of equal significance. Writing in 1980, Lawton noted the influence of the examination system as a major constraint on secondary education and therefore; "…its control is of crucial political significance" (1980, p. 83. See also Ball, 2003, p. 216). This perspective is further noted by Murphy (2003) in viewing the examination as a form of social control and power, particularly vested in those in a position to shape their design. Lawton also expressed concern about the way in which the examination system distorted the curriculum but that:

> …this distortion encourages the classification of pupils into 'successes and failures', 'academic and non-academic' or 'university material or non-starters' (1980, p. 86).

The National Curriculum was the key component of the 1988 Education Reform Act, but as discussed later in this chapter, the associated assessment regime was instrumental in controlling the way schools responded to what became a statutory curriculum and assessment regime, and as Oates (2013) has pointed out; "…schools' behaviour constantly adapts when qualifications change in structure or when pressures from accountability undergo subtle shifts" (p. 3).

4.5.2.4 School Based Assessment (SBA) - Coursework and Controlled Assessments

The national criteria defined coursework as; "…all types of activity carried out by candidates during their course of study and assessed for examination purposes" (SEC, 1986, p. 1). As noted by Crisp (2008), coursework was not a novel approach to assessment but the introduction of the GCSE; "…saw a much increased presence of coursework as part of the assessment culture through its requirement in most subjects" (p. 20). The reason for the inclusion of coursework components were based on arguments of increased validity and for the assessment of objectives for which evidence was ephemeral or difficult to ascertain through written examinations (SEC, ibid). It was also viewed as motivational for students:

> However, if concerns about coursework becoming overly formulaic and predictable in some subjects are well-founded, then coursework may not achieve its intended positive impact. ….Additionally, the heavy workload for teachers and students reported by some constitutes a negative impact of coursework for some of those involved and hence may threaten validity in this respect. (Crisp, 2008, p. 21).

The use of coursework was not universally supported with early concerns not merely focussing on manageability but a:

> …wariness about its (coursework) use, particularly when one English GCSE that used 100% coursework and no exams became so popular that the great majority of 16 year olds were taking it (Opposs, 2016, p. 54).

And by 1991, the then Prime Minister John Major attacked teachers' low expectations and their adoption of 'fashionable theories' and 'hostility to testing' noting that:

> Tests are essential…It is clear that there is now far too much course work, project work and teacher assessment in GCSE. The remedy surely lies in getting GCSE back to being an externally assessed exam, which is predominantly written (Major, 1991).

It is not clear on what basis the government directed 'maximum of 20%' coursework was proposed, however reductions in coursework were duly introduced. Following the

introduction of accountability measures in the 1990s, coursework once again became a focus of attention. Coursework in this context is discussed in Section 4.5.3.8 of this chapter.

4.5.3 Episteme 3. The age of public accountability and performativity (1988 – to the present)

4.5.3.1 The 1988 Education Reform Act

This section of the chapter identifies a third episteme in which the outcomes of what became the first statutory tests and general qualifications formed the base data from which the performance of the school system in England was measured. This period is characterised by the formation of the first statutory national curriculum covering all phases of compulsory education accompanied by statutory assessments of pupils aged 7, 11 and 14 established in law by the 1988 Education Reform Act. The selected ages represented the end of what were designated as key stages; there were no statutory arrangements for assessments at the end of key stage 4 (pupils at the age of 16) as at this point the vast majority of pupils sat general qualifications. However, the content of syllabuses designed for GCSE qualifications were governed by subject and general criteria that reflected the content of the statutory key stage 4 National Curriculum. This ushered in a period of centralised determination and control of the school curriculum through the establishment of core and foundation subjects[22], each containing the designated knowledge, skills and understanding expected of pupils at the end of each key stage that were described as 'attainment targets' and a regime of nationally determined statutory assessments and GCSE qualifications. This was in stark contrast to any preceding period of educational history. The statutory assessments were not designed as qualifications but rather as:

> …the arrangements for assessing pupils at or near the end of each key stage for the purpose of ascertaining what they have achieved in relation to the attainment targets for that stage (Education Reform Act, 1988. Part 1, 2 (c), p. 2).

The National Curriculum was over-seen by the formation of a National Curriculum Council (NCC) charged with keeping the curriculum under review and to advise the Secretary of State on such matters and a School Examinations and Assessment Council (SEAC)

---

22 The core subjects defined in the 1988 Act were: mathematics, English and science; and the other foundation subjects — history, geography, technology, music, art and physical education. In the third and fourth key stages, a modern foreign language was to be specified in an order of the Secretary of State.

charged with keeping examinations and assessments under review and advising the Secretary of State. Both councils were formed of councils appointed by the Secretary of State. Part 1, Section 22 of the Act also made provision for the Secretary of State to make regulations for the provision of information on the educational achievements of pupils; "…including the results of any assessments of those pupils" to be made available either generally or to prescribed persons; "…in such form and manner and at such times as may be prescribed" (p. 18). Further, the Act conferred powers on the Secretary of State to revise the National Curriculum; "…whenever he considers it necessary or expedient to do so" (Section 4, 1 (b), p. 3) and to make changes to attainment targets, programmes of study and assessment arrangements. Powers were also extended to include courses leading to external qualifications:

> No course of study leading to a qualification authenticated by an outside person shall be provided for pupils of compulsory school age by or on behalf of any maintained school unless the qualification is for the time being approved by the Secretary of State or by a designated body (Section 5, (1) ,p. 4).

Such provisions breached the walls of the 'secret garden' as described during the 'great debate' and opened the path for greater visibility – or surveillance- of schools, teachers, pupils and education ministers.

4.5.3.2 The Task Group on Assessment and Testing (TGAT)

In preparing for the 1988 Act, the government formed the Task Group on Assessment and Testing (TGAT) with the remit of advising on assessment and testing in what was then the proposed National Curriculum. Reporting in December 1987, TGAT, together with three supplementary reports published by the DES in 1988 held significance for the functioning of the whole enterprise (Kimberley et al., 1988). According to Kimberley et al (ibid), TGAT had to work within conflicting parameters; "…which they were not magically able to resolve" (pp. 234-5). There were two elements at work here: one, the design of a system that would provide information to students, teachers and parents around the educational strengths, weaknesses and progress of pupils in relation to national standards; and two, to develop systems that would provide statistical data:

> …produced by schools and authenticated by LEAs (DES, 1987, para. 134), for parents, prospective parents, governors, education officers, and a wider public (Kimberley et al., pp. 234-5).

The TGAT report contained a range of recommendations supporting the use of teacher's assessment and externally set standard assessment tasks (SATs). However, the Report noted that:

> A more widely voiced fear is that external tests will impose **arbitrary restrictions on teachers' own work**, (original emphasis) and so limit and devalue their professional role. (para.16).

This resulted in recommendations that assessment should draw on teachers' assessments conducted over time in ' normal learning contexts' and that:

> …the administration, marking and moderation procedures rely on the professional skills and mutual support of teachers, giving them both key responsibilities and communal safeguards against idiosyncrasy (para. 16).

TGAT did however note that experiences derived from the assessment of GCSE coursework and the in-service training for primary teachers highlighted the essential need for support to help teachers to adapt to the procedures that the new system would require (para.17). The report referred to the advantages of assessments conducted by teachers over time that would address all aspects of the curriculum, unlike the more limited approach of specific tasks administered on specific timings. In the latter point, TGAT noted how the performance of pupils in tests can vary from day to day. In effect, TGAT expressed the view that teacher assessments over time would provide greater validity, but supported the use of externally set tasks to be administered with a level of flexibility to sit within 'normal' classroom practice. Notwithstanding, the report did express concerns over local interpretations of standards and potential bias, but noted that such issues could be overcome through training and robust local moderation systems. On the condition that robust moderation procedures were to be in place, TGAT concluded:

> We therefore recommend that the national assessment system is based on a combination of moderated teachers' ratings and standardised assessment tasks (para.62).

There was also a recommendation that standard assessment tasks should be selected by teachers from a national item bank for children aged 7, with tasks supplemented by other options such as tests provided for particular elements of the curriculum available for pupils

aged 11. However, TGAT was not universally greeted with positivity. The formation of assessments designed against attainment targets structured further by levels of attainment presented a model of discrimination:

> As schooling is being restructured to create a 'disparity of esteem' and inequality of resources, competitive self-interest becomes the 'rational' response of parents, governors, pupils and teachers. Thus the whole TGAT report assembles a complex repertoire of assessment proposals, the outcome of which is to discriminate, in both senses of the word. In so doing it promises 'results' which can form the concrete basis on which 'informed choices' can be made. 'Privatisation' now takes on an internalised, ideological form to complement its material expression in the market sphere. Structurally this links directly to social divisions (Kimberley et al., p. 240)

4.5.3.3 The introduction of National Curriculum Assessments

Standard Assessment Tasks designed to assess the National Curriculum were introduced by the government in the early 1990s but almost immediately ran into trouble through opposition to the imposition of the assessments on teachers and pupils and the possibility of less detailed performance tables than envisaged by TGAT: this was manifest in a move by the teacher unions to boycott the assessments:

> The boycott was a symbolic rejection of all that had been imposed on education. The wider opposition to the tests embraced a rejection of the whole Conservative market model of education with its reliance on crude school-by-school comparisons in the form of league tables (Coles, 1994. p. 16).

Coles sites this reaction in the wider context of disillusionment with the then Conservative government – for example the closure of pits, economic crisis, local authority cuts and; "For teacher unions all this came at the end of a number of significant defeats at the hands of the government" (p. 17). This should also be viewed in the light of a contemporaneous speech made by the then Prime Minister John Major noted earlier in that it generated a particularly strong rejection from teachers of English who opposed the dramatic cuts in GCSE from what had been a 100% coursework based assessment. This was exacerbated by the announcement of late changes to key stage 3 assessment arrangements. Mathematics and science underwent pilot runs in 1992, but English was added, without a pilot in 1993. The National Union of Teachers (NUT) balloted members in 1991 regarding

a boycott, but the return was 9 to 1 against such action as; " …many teachers felt that they still had enough room for manoeuvre under the terms of the National Curriculum to shape teaching according to their own priorities…But within a year, the situation had changed dramatically" (Cloake, 2017, p. 75). The NUT put this down to the ever increasing pace of change, increased administrative burden and challenges from the right wing on child centred teaching methods as undermining 'traditional values' (ibid). However, the presence of six teaching unions:

> …each with its own ideas, cultures and approaches, had greatly helped Kenneth overcome opposition to his Education Bill. In later years, Baker sarcastically thanked the then General Secretary Steve Sinnott for presenting him with so many unions to play against each other. "Without that, the '88 Act could never have got through" (Cloake, 2017, p. 75).

In essence, this illustrates Foucault's idea that power is everywhere, yet the spread in this case resulted in forces working in different directions even within one so-called professional body. The boycott was challenged through a government backed Wandsworth Local Authority who took the National Association of Schoolmasters/Union of Women Teachers (NASUWT) to the High Court. However, the Local Authority was unsuccessful, not on the timing or educational merit or otherwise of the tests, but on the unreasonable workload demands made of teachers generated by the marking of assessments. The then Secretary of State, John Patten, took a dogged view on testing and those opposing them. Speaking in a BBC On The Record interview with Jonathan Dimbleby, he stated:

> I said that I thought anyone who thought that testing should not happen to our children was taking a Neanderthal view. That's what I was saying, and I hold to that very, very clearly indeed. I think if we're ever going to have a competitive English, British schooling system we need to have as much testing as we can in order to bring us up to the levels of the French and the Germans (BBC, 1993, interview transcript).

However, the tests did not take place in 1993, but a subsequent report by Sir Ron Dearing (1994) Chair of the newly formed School Curriculum and Assessment Authority (SCAA), resulted in the formation of externally devised and marked National Curriculum tests:

This response, forced by the teachers, paradoxically took the system further from one of the original educational objectives of TGAT, that of the teachers' marking their own students work and deriving useful feedback from the process (Whetton, 2009. p. 145)

The re-casting of National Curriculum assessments from standard assessment tasks to one of tests was significant in moving the tests from a position of high validity to a regime more concerned with reliability and manageability:

…the balance between validity, manageability and reliability gradually (and at times rapidly) swung away from validity toward manageability and reliability. This became inevitable once the teachers' unions, which organised a boycott of the assessments in 1993, had determined that the paramount issue was manageability (Whetton, ibid).

4.5.3.4 National Curriculum Assessments – the middle years

Following the early challenges of introducing National Curriculum assessments and the move to external tests and external marking, the assessment regime remained largely stable up to the mid 2000s. However by 2004, parental concern was developing over the negative impact on pupils of testing, particularly of children aged 7. A new approach was trialled by the Qualifications and Curriculum Authority (QCA) in which teachers could use nationally provided tests, but administered at a time determined by teachers: the outcome of the test was then used to inform overall teacher assessments. The trial was evaluated by the University of Leeds (see QCA, 2004), was judged as being successful and subsequently adopted from 2005 (Whetton, 2009).

During the 2000s, a number of administrative issues occurred with National Curriculum assessments, for example delays to the delivery of materials and late marking of key stage 3 English tests in 2003-4 (see Beasley et al., 2004) and catastrophic failure in 2008 when the United States based Educational Testing Service (ETS), the company contracted by QCA to administer tests, failed to complete the marking process (see Sutherland, 2008). As a consequence, National Curriculum tests at key stage 3 were dropped by then Education Secretary Ed Balls noting the GCSE at the end of key stage 4 was sufficient for providing data used for accountability purposes. Balls also commissioned an Expert Group on Assessment (2009) to provide advice on National Curriculum assessments and wider accountability issues. The Expert Group report stated that; "…insufficient emphasis has

been placed on assessment both in initial teacher training and in later professional development" (para. 146, p. 19) and recommended moves to strengthen the quality of teacher assessment by every school having a lead assessor and by 2020, a qualified Chartered Educational Assessor, which once established would underpin the validity and reliability of teacher assessments to be used as part of the accountability system. Other recommendations included dropping the science test at key stage 2 and replacing it with a national sampling test, and the use of detailed school report cards to replace the achievement and attainment (league) tables used for public accountability (ibid., pp. 8-9)[23]. However, external tests were to be continued at key stage 2 for English reading, writing and mathematics. This was based on the view that:

> Whilst the school system as a whole places a high level of trust in schools and teachers, there is a view that this is not always the case in the area of assessment. The argument has been made that removing all externally marked Key Stage 2 tests and replacing them with teacher assessment only would demonstrate a higher level of trust in teachers. Whilst we have considered this argument, and evidence about the reliability of both tests and teacher assessment, we have concluded that this approach would represent a step backwards, both for pupils' learning and for school accountability (Expert Group, 2009, Para. 9, p. 4).

The decision to recommend the maintenance of external tests at the end of key stage 2 was based on the need for a system of external validation of standards to complement the GCSE at the end of key stage 4. However, the use of tests for primary aged children remained controversial (see for example Curtis, The Guardian, 14th October 2008).

In 2010, further significant changes were made to National Curriculum assessments following a partial boycott of tests – supported by around 26% of primary schools, but fewer than the teacher associations had hoped. Julie Henry (2009), education correspondent for The Telegraph reported wide support from the NUT and NAHT teacher unions noting that Christine Blower (NUT General Secretary) had stated that; "…parents supported scrapping Sats", quoting an NAHT survey of 10,000 parents which found that 85 per cent wanted to abolish the current system (Henry, 2009). However, the tests were conducted as usual, albeit with fewer pupils, and results were issued to participating schools. Schools that did not participate received a reprimand letter from the DfE. In 2011 there was a change of government and a further review of National Curriculum

---

[23] The Report Card was intended for both primary and secondary phase reporting.

assessments was commissioned chaired by Lord Paul Bew. Following the Bew Review report of key stage 2 assessment and accountability (2011), external marking of the English writing test was ended and replaced by teachers marking their own pupils' papers: the teacher generated marks for writing were not however included in the accountability measures. This change was in response to concerns regarding misclassification based on the view that writing poses particular problems regarding reliability:

> We believe that there is fundamental challenge with the marking of writing composition (extended writing of prose, verse, formal letters etc.) because it requires a professional's judgement rather than being empirically 'right' or 'wrong'…the criticism of the marking of writing is not principally caused by any faults in the current process, but is due to inevitable variations of interpreting the stated criteria of the mark scheme when judging a piece of writing composition (Bew, 2011b, pp. 60-61).

However, the approach advocated by Bew that writing is a 'special case' is at odds with the assessment of similar skills at key stage 1 and at GCSE level demonstrating inconsistency of approach across the compulsory phase of education. It further supports Foucault's view that history is not a story of progressive development but rather consists of breaks and contingencies. The creation of performance tables and the reliance on externally set tests of the National Curriculum and general qualifications mark quite a different ethos from that experienced at the time of the Norwood Report in the early 1940s.

4.5.3.5 Performance tables

A key change in the English education system during the period under review of this study is the use of examinations and tests as a measure beyond the individual, the latter an area of Foucault's interest:

> Assessment information has become a proxy measure that is supposed to facilitate judgments on the quality of most elements of our education system: its **teachers**, **head teachers**, **schools**, **support services**, **local authorities** and even the **government itself**. This represents a fundamental change from the situation even 20 years ago, when test and examination results were predominantly meant to serve as indicators of what a **pupil** (all emphasis in original) knew and understood of a subject (Mansell et al., 2009. p. 7).

In the early 1950s, examination results were by and large the property of the individual receiving the qualification. However, over time, and especially over the last 35 years, examination results have produced increasingly large data sets. This has coincided with a growing demand for accountability and user choice (Foley and Goldstein, 2012):

> Growing out of the performance management movement in the private sector, and aided by the increasing availability of large administrative databases, the most visible manifestation of this has been the publication of institutional rankings or 'league tables' based upon particular performance indicators (ibid., p. 7).

According to Foley and Goldstein, performance rankings serve two broad purposes: on the one hand public accountability; "…whereby those who provide resources to run institutions such as schools or police forces can form judgments about where improvement is needed or particular action is required" and on the other hand; "…to provide users of services, such as parents who wish to choose a school, with information to assist them" (p. 7). The overall aim is that institutions and others will react to the publication by engaging in "…competition between institutions in a quasi-market environment " (ibid., p. 7). However, Foley and Goldstein identify:

> A third, and not so obvious, function of league tables and their associated 'institutional targets' is that of control (p. 8).

The assessment arrangements contained within the 1988 Education Act laid the policy foundations for the use of assessment outcomes as measures of accountability. However, the TGAT report cited early concerns with such use noting:

> …there is a fear that results will be published in **league tables of scores**, (original emphasis) leading to ill-informed and unfair comparisons between schools. We believe that most teachers and schools would not object to assessment results being reported to those who know the school and can interpret them in the light of a broader picture of its work and circumstances. They would object, however, to the publication of partial information which is not set in that context and is therefore potentially misleading – particularly where significant decisions are then based upon that information (para.18).

In 1992, the performance tables containing the results of general qualifications were published nationally for the first time as a means of comparing the performance of schools,

effectively applying Foucault's concept of the list at the institutional level (see Figure 4.6). The then Secretary of State John Patten was a keen advocate of publishing performance data stating:

> Your performance and my performance is measured daily and I think it's a very good thing. The professional's performance should be measured, that's why performance tables in the schools are so important (BBC, 1993).

In 1996, results from National Curriculum assessments were added to national data published by the Department for Education (DfE) drawing the performance of the primary phase alongside the secondary phase and into public view (see Figure 4.7).

The content of performance tables has changed year on year since 1992. Each year the DfE, or equivalent, publishes a statement of intent[24] that provides information on forthcoming data requirements: this provides a vehicle for successive governments to drive policy and influence school behaviour. For example in 2014 the then Secretary of State for Education Michael Gove drew on the Wolf Report (2011) to make changes to the secondary performance tables by instigating a focus on what he described as the 'centrality of academic knowledge', largely in response to criticisms over the use of vocational qualifications used by some schools to boost their performance in accountability tables. And despite Gove's rhetoric around the parity of esteem between vocational and academic courses his stated target was:

> …ensuring curricula and exams are more rigorous - with a proper emphasis on the centrality of academic knowledge in the education available to all. Giving all children access to high-quality teaching in maths, English, physics, chemistry, biology, languages and the humanities to the age of 16 provides every child with the opportunity to flourish whichever path they subsequently choose (Gove, 2014, p. 6).

---

[24] The statement of intent fro 2019 and previous years can be found at:
https://www.gov.uk/government/publications/school-and-college-performance-tables-statements-of-intent

Figure 4.6: Extract from the summary of the contents of the first secondary performance tables published by the DfE, 1992
(Source: A History of School and College Performance Tables. DfE, 2012a Internal document)

1992

- First **secondary school performance tables** published by the Department - to inform parental choice and raise standards.

- All maintained mainstream and special schools included. Independent schools included on a voluntary basis.

- Data provided on

   o percentage of 15 year olds achieving:

        5+GCSE grades A-C

        5+GCSE grades A-G
        1+GCSE grades A-C
        1+GCSE grades A-G

   o average point score for 17 year olds entered for at least 1 GCE A/AS exam

---

Figure 4.7: Extract from the summary of the contents of the first primary performance tables published by the DfE, 1996
(Source: A History of School and College Performance Tables. DfE, 2012a, Internal document)

1996

- First **primary school performance tables** published by DfES, showing % of pupils eligible for Key Stage 2 assessment achieving level 4 in tests and teacher assessment in each of the three core subject, plus contextual data on pupil numbers and pupils with SEN.

- Primary tables cover all maintained mainstream schools with KS2 pupils on roll (no special or independent schools)

Secondary Tables

- Data on achievements of 15 year olds in vocational qualifications revamped to focus only on qualifications within the GNVQ/NVQ family, and their pre-cursors.

last revised 16 July 2012   18

A HISTORY OF SCHOOL AND COLLEGE PERFORMANCE TABLES

Secondary and Post-16 Tables

- Achievements in Intermediate GNVQ (and pre-cursors) by 16-18 year olds added.

4.5.3.6 Performativity

The impact of performance tables introduced a new era characterised by the notion of 'performativity' (see Ball 2003 & 2013). Ball describes performativity as:

> …a technology, a culture and a mode of regulation that employs judgements, comparisons and displays as means of incentive, control, attrition and change based on rewards and sanctions (both material and symbolic). The performances (of individual subjects or organizations) serve as measures of productivity or output, or displays of 'quality', or 'moments' of promotion or inspection. As such they stand for, encapsulate or represent the worth, quality or value of an individual or organization within a field of judgement (Ball, 2003, p. 216).

This technology challenges or displaces teachers' values describing a struggle; '…over the control of the field of judgement and its values' (ibid., 2003, p. 216). In England, the publication of examination and test data reduces educational performance to a data set of limited validity. The data are presented by government as a means to encourage teachers:

> …to think about themselves as individuals who calculate about themselves, add value to themselves, improve their productivity, strive for excellence, strive and live an existence of calculation'(ibid., 2003, p. 217).

In some extreme cases, these data are used within schools to rank teachers as this extract from the Times Educational Supplement illustrates:

> To make matters worse, senior leaders have ranked all the teachers in the school. Jessica is near the bottom of the league, while John, across the corridor from Jessica, is well "in the green" with his lot. She knows, she just knows, that she is a better teacher than John and yet she's floundering "in the reds". The public availability of this data makes her feel like everyone is judging her (Rogers, 2017, viewed online).

The impact of performance tables goes beyond those in the teaching profession and holding politicians to account even to the extent of influencing house prices in the proximity of 'high performing schools':

The most recent CEP (Centre for Economic Performance) research for England shows that a primary school one standard deviation above the average in terms of the performance of its pupils in key stage 2 tests (at age 11) attracts a house price premium of around 3%. This means that a school right at the top of the league tables attracts a premium of around 12% relative to one at the bottom (Gibbons, 2012, p. 2).

As an example, a house in Kirklevington was advertised in 2013 as being 'approximately two minutes' away from a primary school 'placed very high on the school league tables' (Rightmove, August, 2013). This phenomenon is driven by the publication of performance data in the mass media: see Figure 4.8. It develops a self-fulfilling prophecy of school performance as parents who have financial power and high performing children fuel rising house prices forcing out those who cannot afford the inflated prices.

Figure 4.8: Media coverage of school performance data

1). The Sunday Times: November 24th 2019

2). The Mail Online: 5th December 2005

The use of performance data has been further intensified by a growing interest amongst policy makers in data drawn from international tests, for example the Programme for International Student Assessment (PISA), a triennial test of the knowledge and skills of 15 year-old pupils administered by the Organisation for Economic Co-operation and Development (OECD) which included 72 countries and economies in the 2015 series and 79 in the 2018 series. Indeed the Bew Review recommended continued participation in specifically designed international comparative studies noting:

> We believe accountability to Government, parents and the public also includes providing a picture of how our education system compares with education systems internationally…as the challenge facing our education system is not only to improve year-on-year, but also to keep pace with the best education systems in the world (p. 42).

Reports from the OECD (2007 & 2010 in Gill and Benton, 2013) suggest that the publication of performance data used to monitor and evaluate teachers or allocate resources, tracked over time by an administrative authority and posted publicly, have a positive impact on student achievement. And despite Gill and Benton's assertion that the effects of the publication of performance data quoted by the OECD; "…were really quite small, and only just reached statistical significance" (p. 8), the government has promoted the view that stronger accountability leads to better results for pupils (Gibb, 2015). Performance tables based on test and examination results have become a key feature of the English education accountability system. So much so that interest and access to international performance data has now become more pervasive with more frequent media coverage not only at national but local level, as illustrated by the article from the Huddersfield Examiner shown in Figure 4.9.

This results in what I believe to be a curious predicament where the institution and the individual are at one time the same and in a state of tension. Using Foucault's analysis, the individual (in one instance the teacher) is objectified through the measurement of their performance against a set of performance indicators through a system of performance management largely set against the achievement of their pupils (at individual, class or school level) – particularly in regard to externally set tests and examinations. At the same time, the institution is evaluated against the performance of other institutions through its place in performance tables measuring limited elements of their overall output – principally

through a concentration on the examination results of externally designated subjects within the broader curriculum offered by schools:

> We are stood before a "permanent economic tribunal" (Foucault, 2010) against which all intellectual activity is judged (Simons & Masschelien, 2006). Furthermore, there is an individualisation of educational institutions as they compete with one another to recruit and perform (Ball, 2013, p. 138).

Figure 4.9: Extract from the Huddersfield Examiner.
   Thursday January 9th 2020
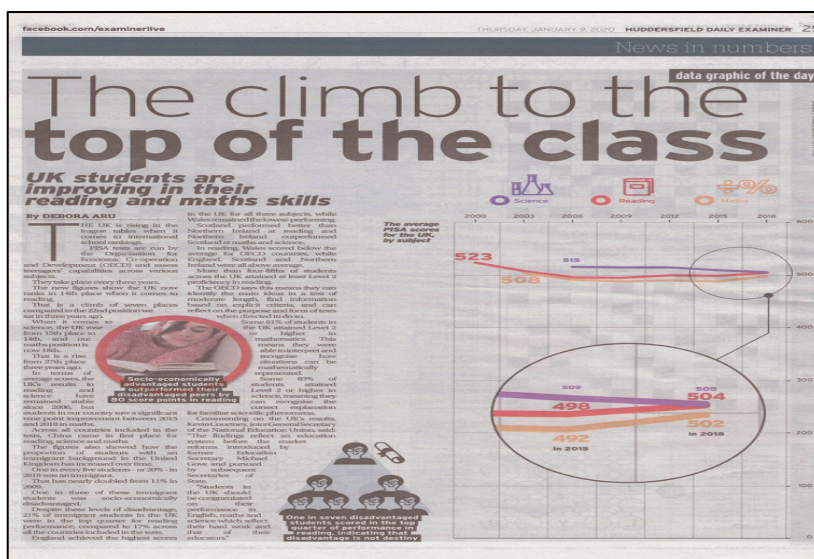
This individualisation of schools drives a climate of competition to perform and recruit, but equally holds ministers and governments to account: "Performativity then, is a "new" moral system that subverts and re-orients us to its truths and ends. It makes us responsible for our performance and for the performance of others" (Ball, 2013, p. 138). The introduction of performance data published by the Department for Education since 1992 has arguably created a new episteme. On one hand, systems of accountability provide public access to information that may be otherwise hidden, whilst on the other hand:

> …it also builds a culture of suspicion, low morale and may ultimately lead to professional cynicism, and then we would have grounds for mistrust…If we want a culture of public service, professionals and public servants must in the end be free to serve the public rather than their paymasters (O'Neill, 2002, in Mansell et al., 2009).

Foucault's notion of the dominant discourse being the product of power struggles (see Section 4.3.3) is relevant here in relation to control over the ways and means used to publicly present school accountability and performativity information through the mass-media. For example, data on school performance can be presented for professional consumption in an informative, dispassionate way such as that produced by the DfE or the Joint Council for Qualifications (JCQ) on National Curriculum tests and GCSE outcomes respectively[25] – or it can be presented in a way that conveys or underpins political or other forms of ideology or as a means of influencing public opinion and wider discourse as often seen in the mass media and more often providing a platform for powerful elites to access audiences en-masse (Khosravinik, 2014). In this latter sense, the examples of mass media coverage of performance data given earlier in this section (see Figures 4.8 and 4.9) demonstrate a shift of discourses:

> "…from 'construals' to 'constructions' (Sayer, 2000), from being just representations and imaginaries to having transformative effects on social reality, being operationaliszed – enacted as new ways of (inter)acting, inculcated in new ways of being (identities), materialized in new instruments and techniques of production or ways of organizing space (Fairclough, 2012, p. 465).

As such, the use of educational assessment data as a means of power becomes 'normalised' (Foucault, 1977a; see Section 4.3.4), absorbed and accepted into every-day discourse as for example in the now annual public debates around examination results each August (see section 2.9), and even to the extent of influencing house prices as noted earlier in this section.

Ball (2003) refers to forms of "immediate surveillance" manifest in appraisal systems, target-setting and the comparison of outputs (p. 219) and quotes a 2000 case study conducted by Troman (2000, p. 349) that found low trust to be in in the ascendant in most schools studied. This marks quite a different ethos from that experienced at the time of the Norwood Report (1943). The age of performativity raises questions over teacher identity. A small scale study by Wilkins (2011) has suggested the 'possible emergence' of a generation of teachers whose experience as pupils has been of an increasingly performative schooling system who:

---

[25] See: DfE; https://www.gov.uk/government/statistics/national-curriculum-assessments-key-stage-2-2019-revised
JCQ; https://www.jcq.org.uk/examination-results/

…cannot be categorised as either 'compliant' or 'resistant' to the demands of performative management systems and government initiatives. They are still largely motivated by affective rewards, but have clear career ambitions; they are aware of the potential conflicts between the demands of accountability and the desire for autonomy, but are generally comfortable with the balance they feel able to strike between these (p. 389).

According to the Organisation for Economic Co-operation and Development (OECD) TALIS Survey 2018, teachers in England are on average 39 years of age, with 18% aged 50 and above (OECD, 2019). Given that the National Curriculum and associated assessments were introduced 31 years ago in 1988, the majority of practicing teachers who were educated in England were either educated, trained or have taught under the direction of the National Curriculum. This may well support Wilkins' findings that presents a view of the erosion of teachers' autonomy and the exploration of the:

 …notion of a technicist model of 'incorporated professionalism' (Troman, 1996., Day et al, 2006). From this perspective, the justifiable political and public pressure for accountability has led to concern that teachers have become increasingly de-professionalised and compliant in their delivery of state-imposed initiatives, be they curriculum 'innovations', new forms of school management or standardised testing programmes (Hatcher 1994) (Wilkins, 2011. p. 393).[26]

In essence, the externally imposed emphasis on 'change and improvement' leads teachers to focus on the outcomes or product of assessment and not the process (see Hanks, 2017, p. 63). In a study of students undertaking initial teacher education courses, Twiselton (2000) identified three categories or stages of development that student teachers often go through as they progress to become effective teachers of English: task managers, curriculum deliverers and concept/skill builders. There is insufficient space here to explore these categories in detail other than to say that a task manger is very product orientated: curriculum deliverers make more connection with the process of learning and concept/skill builders are more focused on the curriculum subject and the deeper understanding of underlying concepts. However, moving from task manager to concept/skill builder is; "…not a straightforward sequence of progression" (ibid., p. 158).

---

[26] See also Wang, 2011, p. 143)

On the one hand the categorisation broadly describes the stages in-line with ITE development, on the other it suggests there is possibly:

> …a danger that even student teachers at the end of their training view the curriculum as an end in itself, without questioning, exploring or fully understanding the rationale underlying it. (and) …that highly prescriptive curricula, such as the National Literacy Strategy and the ITT English National Curriculum, increase this danger, at least in the short term. (Twiselton, 2000, p. 391).

Clearly Twiselton's work is focused on ITT and early career development and sits within the context of the National Literacy Strategy,[27] and whilst the findings point to both positive and negative effects of the provision of a highly detailed and given curricula, it raises questions about how such provision may limit student teachers' ability or decrease their development. As Furlong (2005) notes:

> The state has taken a much more assertive role in defining how to teach as well as what to teach; the result has been the establishment of what I would term 'managed' professionalism. And schools have become the focus of what can be characterized as 'networked markets' (Reid et al, 2004) (p. 123).

As Twiselton's study points out, the role of mentors and the school setting has become crucial in supporting teachers' further development. However, in the context of the limited focus on assessment theory and practice in initial teacher education and the lack of trust and confidence in teachers' assessment more generally as discussed earlier, Twiselton's work suggests that further research should be undertaken to gain an understanding of how teachers develop their knowledge, understanding and practice in a regime of highly prescriptive, and at some points statutory system of educational assessment.

4.5.3.7 Concerns with teachers' assessments

It is worth stressing that during the 1990s as new National Curriculum assessments were introduced, the envisaged role played by teachers changed largely through the interjection of teacher associations raising concerns about the impact on the curriculum, pupils' wellbeing and teacher workload. The initial TGAT model of assessment held teacher

---

[27] See: The National Strategies 1997–2011
A brief summary of the impact and effectiveness of the National Strategies.

assessment as being central to the process. Following the teacher boycotts and a high court action taken by the NASUWT in 1993 that successfully argued that teachers should receive additional pay for marking tests, the Dearing Review's (1994) recommendation that external testing should replace teacher assessments was welcomed by government. In 2011, The Bew Review raised further concerns about evidence it had considered on bias in teacher assessment concluding that: "This evidence suggests that certain pupils could be disadvantaged by a system which relied more heavily on teacher assessment" (p. 49). Bew was not alone in positing such a view. Similar concerns were raised about the reliability of teacher assessments used in GCSE courses even within the same institutions:

> There are issues surrounding teacher assessment that have to do with potential bias, application of different, sometimes personal, assessment criteria, and differences in the available evidence base when implemented curricula and standards of judgement differ from class to class and school to school (Johnson, 2011, p. 5).

Her Majesty's Chief Inspector of schools (HMCI) also noted "worrying inconsistencies" in key stage 1 teacher assessments and called for consideration of a return to external testing:

> In infant schools, for example, children are more likely to be assessed as reaching, or exceeding, the standards expected for their age than they are in all-through primary schools. Moreover, uneven moderation by local authorities of the work carried out by schools can lead to poor quality and unreliable assessment. For these reasons, I urge government to consider a return to external assessment at the end of Key Stage 1 (HMCI, 2013, p. 9).

4.5.3.8 Coursework and Controlled Assessments in General qualifications

By 2005, the coursework components in most subjects had become more restricted, ranging from 20%, for example in GCSE science and mathematics, to 60% in Art and Design and Design & Technology. In GCE A level, coursework ranged from 0% in mathematics, 0 to 30% in English literature and 60% in Art and design. Nevertheless, concerns over the coursework component persisted and arguably increased. Between 2003 and 2004, the Qualifications and Curriculum Authority (QCA) conducted a review of coursework in the GCSE. The key concerns driving the review were threats to the validity and reliability of coursework such as pressures on teachers to obtain the best results, the

influence of the internet on plagiarism, instances of malpractice and inconsistent standards of internal assessment (QCA, 2005). The report however did support the use of coursework albeit in a more tightly regulated format recommending more guidance and support for teachers and a review of subject criteria to consider if coursework was "a necessary and appropriate assessment instrument" (ibid., p. 22). As a consequence, a further independent report was commissioned by the QCA and produced by Ian Colwill (2007) and resulted in the replacement of coursework with what became known as 'controlled assessments' with tighter awarding body controls of what constituted appropriate tasks and closer supervision requirements. The aim of controlled assessments was:

> …to encourage a more integrated approach to teaching, learning and assessment as well as to enable teachers to confirm that students have carried out the work themselves (Opposs, 2016, p. 55).

The approach was accompanied by the application of one of three specified weightings (0%, 25% or 60) and:

> Predetermined levels of control or supervision in controlled assessment (limited, medium and high) applied at three stages in the assessment: task setting, task taking and task marking (ibid., p. 55).

Controlled assessments from 2010 were intended to address the issues around coursework noted above. However, this in turn caused a reaction, such as the tasks would hold limited appeal to candidates, be overly structured and overly focused on particular elements of the syllabus. Crisp (2008) has noted that research conducted by MORI in 2006 reported a general level of recognition of the educational benefits of coursework, although the burden on teacher workload was clearly of concern:

> It is interesting that the Heads of Department interviewed by MORI (2006) were fairly positive about coursework, particularly in subjects with oral or practical coursework tasks, and nearly all acknowledged the benefits to students.

> Concerns about internet plagiarism were not as great as might have been expected (82% of teachers disagreed that students used the internet too much) and whilst more than half felt that students in some schools can gain unfair advantage in the current system the most frequently mentioned drawback was the burden of

marking coursework. The interviews by MORI found that 66% of teachers were opposed to removing coursework and 51% were strongly opposed to its removal. The MORI interview evidence would not seem to support the decisions that have been made (Crisp, 2008, p. 23).

But the introduction of controlled assessments did not resolve the issues found in coursework as noted by research conducted by Ipsos/MORI in 2011 for Ofqual:

> There was a mixed picture, controlled assessment being seen as more of a problem in some subjects than in others. There were also manageability issues but while some schools reported struggling with the practical issues of operating controlled assessments, others reported that they had put in place systems to manage it. Although some suggested replacing controlled assessments by written exams, others did not see that as a viable option as many GCSEs included practical elements that really could not be assessed in a written exam (Opposs, 2016, p. 56).

Ofqual carried out a review of controlled assessments in 2013 and found a number of areas where the format failed to address issues of validity and reliability. These included the memorisation of written pieces that were regurgitated under controlled conditions – becoming a test of memory rather than writing – and teachers interpreting rules and instructions for conducting controlled assessments in different ways, for example candidates using prepared notes that could be referred to in controlled assessments as essay plans - which were not allowed (Opposs, ibid). A further concern was created around GCSE English in 2012 in which the entry patterns of some schools raised suspicions of over-marking of controlled assessment accounting for 60% of the examination total marks. To compensate for the actions of some schools the awarding bodies raised the grade boundaries, an action supported by Ofqual. As a direct consequence:

> A consortium of schools and local authorities pursued a claim for judicial review. There was a High Court hearing in London in December 2012 and the application was subsequently rejected (Royal Courts of Justice, 2013) (Opposs, 2016, p. 57).

Further research commissioned by Ofqual in 2014 highlighted deeper concerns regarding teachers providing over-assistance to their pupils' controlled assessments. A fuller account is provided by Opposs (2016): see also Meadows and Black (2018). The outcome of the work undertaken form 2005 by the QCA and its successor bodies saw controlled

assessments replace coursework and resulted in more recent reform of the GCSE and GCE A level that reduced school based assessment even further: this is picked up on more detail in Section 4.5.3.11.

4.5.3.9 Vocational qualifications in the compulsory phase of education

From the outset of the GCE and the earliest days of the establishment of the secondary modern schools, provision was made for alternative qualifications as noted earlier in this chapter. Local qualifications were perceived as lacking in national standards, but qualifications developed by august or recognisable national bodies did attract more public credibility. Over the period of time of interest to this study, a range of vocational or occupational qualifications have entered, and many left, the national framework of public qualifications: for example the Certificate in Pre-Vocational Education (CPVE), the General National Vocational Qualification (GNVQ) and the GNVQ Part One. Others have survived for some considerable time, for example the Business and Technology Education Council (BTEC) First. Not all of these qualifications were developed for the compulsory stages of education, but from their earliest conceptions, the aim was to provide viable qualifications that would demand broad public recognition and support. An important aspect in the development of these qualification types has been the number of attempts made over time to confer parity of esteem between the so-called academic and vocational/occupational streams of qualifications (see for example, Dearing, 1996; Sharp, 1997; Hodgson et al., 2014). However, as with the earlier CPVE, the assessment regime of the GNVQ was in the main teacher based assessment of coursework and portfolios leading to doubts about the rigour of assessment outcomes:

> During 1991 there were clearly fears that the Government's insistence on the use of conventional examinations in nearly all GCSE and 'A' level courses might be applied to GNVQs, but NCVQ successfully opposed this. There was strong support in the (NCVQ) consultation for assessment by projects and assignments together forming a 'portfolio of evidence', but the question of externally set tests proved much more controversial with a majority against them (Sharp, 1997, viewed online).

Such criticism dogged vocational and occupational qualifications used in the compulsory phase of education (and post-16 phase), for example as in the Review of Vocational Education, (Wolf Report, 2011) which noted that any qualification used within a regime of national performance monitoring must have "…a strong element of external assessment" however:

> *This need not, and indeed should not, mean assessment entirely on the basis of examinations, (author's italics)* which in the case of vocational awards will often be quite inappropriate. But we know that, without regular external referencing, assessment standards in any subject invariably diverge across institutions and assessors (p. 112).

The issue of performance measures is germane. In 1993, the Secondary School performance tables published by the Department for Education included data for the first time on the achievements by 15 year olds in vocational qualifications offered by City and Guilds and the RSA; in 1994, data on achievements by 16-18 year olds in Advance GNVQ; and in 1996, achievements in Intermediate GNVQ (and pre-cursors) by 16–18 year olds with data on achievements by 15 year olds in full GNVQs and Part One GNVQs; "…merged with GCSE on the basis of equivalence" (DfE, 2011). The issue of equivalence raises questions as to the veracity of such claims. There is sparse reference to the origins of such equivalence. For example, it was widely published that a GNVQ Intermediate was the equivalent to four GCSEs at grades A to C, but:

> …at no point have these notions been researched and systematically tested. From the outset they were asserted rather than investigated and established and this has continued (Sharp, 1997).

But schools were quick to see the opportunities to exploit the so-called equivalence in response to the growing pressure exerted by school and college performance tables. This fed derisory accusations such as schools entering candidates for GNVQ's in order to 'game the system' and schools and government(s) using outcomes in vocational courses as a means of boosting performance table ratings and falsifying claims of improved standards:

> A pass in GNVQ information technology, for example, is deemed to be equivalent to no fewer than four passes at GCSE (and is included in the data used to make up the school exam tables as if it were four GCSE passes). I asked the DFES how it has come to decide that one GNVQ is equivalent to four GCSEs, but the frosty press officer failed to provide a reason (Clark, viewed online, 2004).

The way in which schools navigated their way through the accountability measures illustrates how; "….power is always present" (Foucault, 1994, p. 291), pedagogical

institutions abound in opportunities for power reversal: whether in teaching students or initiating teachers into reforms" (Stickney, 2012, p. 656). However, over time, regulatory control brought the format of the GNVQ more and more akin to that of the GCSE examinations with increased external assessment that in reality negated its initial purpose and resulted in its ultimate demise in 2007. The move away from internal or school based assessment (SBA) in vocational qualifications has been a constant theme, in particular where qualifications were the basis of measuring school or college performance. The Wolf Report (2011) was instrumental in maintaining this move in providing the argument for changes in future government policy made under the direction of Michael Gove in 2014 calling for qualifications with:

> …a rigorous marking structure, external assessment, robust content and real stretch, or must be redeveloped to meet that standard. As a result there is - at last - the prospect of a genuine equality of worth and parity of esteem between all qualifications (Gove, 2014b).

The approach further reduced the use of teacher or school based assessment and moved the focus of accountability towards 'academic subjects' and resulted in the formation of Progress 8, a measure of progress across eight subjects introduced in 2016 and the English Baccalaureate, which became an accountability measure under the Conservative/Liberal Democrat Coalition Government in 2010 (See Long and Danechi, 2019).

4.5.3.10 The removal of National Curriculum levels of attainment

From September 2014, a revised National Curriculum was implemented, a key component of which was the removal of levels. Attainment targets and levels of attainment had been a central feature of the National Curriculum since its inception in 1988. Levels of attainment were used as tools to describe learning progression in the statutory orders (describing curriculum content), presented as a series of 'grades' that were used in determining cut scores in National Curriculum tests and within schools to measure the progress of pupils. Prior to 2014, each attainment target had eight levels of attainment. However, as one of the stated intentions of the National Curriculum was that pupils were expected to make two levels of progress in each key stage, early concerns were raised by teachers that pupils who made progress but by remaining at the mid-point, their progress would not be recognised by the system of levels. In response, further sub-levels were awarded through the division of levels into three points: these divisions were made by the designation of two points evenly spaced between the cut scores of two levels giving for example level 4a,

4b and 4c. However, unlike levels, there was no description of what might constitute a level 3b for example as there was no central policy decision to include sub-levels nor were they recognized formally by Ofsted's inspection framework (Boylan, 2016, p. 2). Even so, pupils were designated sub-levels, and this was often used for the performance management of teachers in schools. Such practice had consequences. In announcing the removal of levels, the DfE stated:

> We believe this system is complicated and difficult to understand, especially for parents. It also encourages teachers to focus on a pupil's current level, rather than consider more broadly what the pupil can actually do (DfE, 2013).

Levels had become a shorthand for describing pupil's attainment, but it is not clear how well the levels were understood or applied within and between schools. Sub-levels compounded this further. Levels became a sole focus for some teachers and children – sometimes demotivating. Further, levels were criticized for encouraging undue pace in learning over and above a deeper mastery of curriculum content – a result of the accountability system that encouraged schools to move pupils up through levels quickly, rather than securing knowledge in each subject area. Pupils were merely labelled or categorised rather than being seen as individuals with particular strengths and learning needs. As the announcement was made that levels would be removed, the NAHT responded by setting up a Commission on Assessment in 2014. The Commission's report raised further concerns around teachers' assessments:

> The Commission heard from the majority of those submitting evidence that there was a lack of trust in teacher assessment at the present time…There is a worrying lack of trust in individual teacher-based assessment, which emanates from within the profession itself (NAHT, 2014, pp. 15-16).

Ofsted also reported concern over the level of trust between schools, particularly around the transition from the primary to the secondary phase of education "…that contributes in a failure to share information about assessment" (Ofsted, 2016, p. 21. See also Ofsted, 2015, p. 37).

The removal of levels of attainment exposed gaps in teachers' understanding of assessment theory and practice. In practice, many schools have 'designed' their own assessment systems or purchased commercial packages, most of which reflect the previous system of levels, an approach described by Boylan (2016) as; "...the ghosts of

levels – levels, although 'chased away', come close again" (p. 16). Again teachers here invert and appropriate power. The various approaches adopted by teachers following the removal of levels reflects the findings of a YouGov poll published in a report produced by think tank LKMCo and Pearson Education (Millard et al., 2017) which highlighted teachers general lack of confidence in assessing pupils with a fifth of teachers stating that they didn't know where to look for information on assessment. Further the report noted that the majority of teachers completing the poll did not receive instruction on assessment during their initial teacher training, with only a third of teachers reporting that they feel "very confident" in their ability to assess their pupils' work (ibid., 2017). These aspects are covered further in Chapters 5 and 6, which presents the results of surveys on the extent of assessment theory and practice covered in initial teacher training courses conducted for this study.

4.5.3.11 More recent changes to testing and assessment arrangements

Despite the many changes to the format and uses of assessment over the period of time covered by this study, the level and intensity of discourse over tests and examinations continues. Some of this is in direct response to the introduction of new tests, for example the times tables tests introduced for children in year 4 (age nine) proposed for spring 2020 and newly designed baseline assessments for children at the start of schooling also from 2020. Teachers have thus continued to voice concerns over testing. For example:

> Teachers object passionately to the accountability agenda imposed on them because of the consequences that flow from it. These are undermining creative teaching and generating labels which limit students' learning. Crucially, they also threaten children's self-esteem, confidence and mental health (Blower, 2015).

The threats to children's, and teacher's, wellbeing have become more to the fore in recent times, for example the House of Commons Education Committee report on Primary Assessment concluded:

> …the high-stakes system can negatively impact teaching and learning, leading to narrowing of the curriculum and 'teaching to the test', as well as affecting teacher and pupil wellbeing (Education Select Committee, 2017, p. 3)

Yet teachers continue to focus their teaching on preparation for tests often to the detriment of the wider curriculum and further impact on children's health, for example as reported in the TES: "Nine in 10 teachers believe Sats (sic) preparation harms children's mental

health, survey finds" (Busby, 2016. See also Bradbury, 2019): and in a speech to the NUT Conference in 2018, delegate Katharine Lindenberg stated in reference to the proposed baseline assessments: "They are unnecessary, they are pointless, they are expensive and above all they are damaging, and they are immoral" (Turner, 2018). Yet teachers continue to administer the tests, prepare their pupils for such and administer commercially available tests to pupils. Reservations about tests may be due to the pressures exerted by the accountability uses to which statutory assessment and general qualification outcomes are put holding sway rather than being driven by objections to testing per se. Even during periods of strong opposition to statutory testing for example, schools purchased optional tests and copies of key stage tests in large numbers from QCA for their own internal usage (see Table 4.5). In similar fashion, secondary schools purchase past papers and practice papers from examination boards, not solely used as 'mock papers'. These actions would suggest that testing itself is not the driver of teachers' concerns. Further, in terms of preparing for National Curriculum assessments, schools and teachers do have a choice. A group of schools in Hampshire for example have developed the Charter: Assessment for Children (see for example Pilgrim's Cross C of E Primary School, Hampshire[28]). These schools, supported by their governing bodies have publicly stated that:

> … assessment is only used to support further learning for our children (and) always prioritise children's personal growth, long term development and engagement as learners and their well-being over SATs (Sic) tests scores (Pilgrim's Cross Primary School, 2019).

Over the last two to three years, further changes have been made to the part played by teachers in high-stakes assessments at key stage 2 and in GCE and GCSE qualifications. Over the lifespan of National Curriculum assessments, teachers have been required to provide assessments in parallel to the tests. For most of this period, the data have been published by the DfE. However in 2017, the DfE announced that these data would no longer be required and collected. This decision was announced in response to the Teacher Workload Survey conducted in 2016 in which 56% of the 43,832 respondents cited compiling assessment data as a major drivers of workload, along with "…excessive/depth of marking – detail and frequency required" (DfE, 2015b, p. 8) cited by 53% of respondents. (See also DfE, 2016d). In response, the Standards and Testing Agency Assessment and Reporting Arrangements (STA, 2018) stated:

---

[28] Other schools which have signed up to the charter: Appleshaw; Portway Junior; and Bransgore Primary.

Schools are no longer required to make statutory teacher assessment (TA) judgements in English reading and mathematics. This change has been made in order to reduce assessment burdens on schools, as set out in the government response to the 2017 public consultation on Primary assessment in England. Test results in English reading and mathematics will continue to be used in school performance measures (Section 2.1, p. 6).

Table 4.5: National Curriculum Optional Tests. Bulk Orders 2008 to 2012 (Source STA. 2013. Internal document).

The numbers relate to the number of packets of tests sent to schools. Each pack contains 10 papers. Papers for KS1 & KS2 are those ordered by schools for their own use and not those distributed by the agencies for actual statutory tests (unpublished internal QCDA/STA data).

	2008 ORDER	2009 ORDER	2010 ORDER	2011 ORDER	2012 ORDER
KS1	7869	12449	10077	4950	3668
KS2	33591	30736	20853	13036	11185
Yr3	28153	249087	200174	180143	156701
Yr4	286277	254210	203755	178901	156866
Yr5	313644	287507	234089	204403	171604
Yr7	155974	95055	73636	49269	27046
Yr8	132975	80522	62944	42999	24225
Yr9	0	0	144026	66591	30414

The move was intended to reduce teacher workload, but in practice further dilutes the part teachers' play in high-stakes assessment. Further, major reforms of GCE and GCSE qualifications have also been made by Ofqual. Following concerns with coursework and controlled assessments discussed earlier in this chapter: "Ofqual decided that assessment should mainly use exams, with other types of assessment used only where they are really needed" (Opposs, 2016, p. 58). This move was driven by what Ofqual describe as 'threats to validity'. Such threats include the external pressures caused by accountability or selection systems that may; "…make it harder for teachers to mark accurately; challenges to authenticity or from the manageability or time taken to undertake the assessment" (ibid., p. 58). The changes have seen significant reduction or removal of school-based

assessment used in general qualifications, in particular amongst subjects used in the English Baccalaureate - a key components of school accountability measures. However, one of Ofqual's most contentious decisions around school based assessment (SBA):

> …was in the assessment of practical work in A level biology, chemistry and physics. Ofqual carefully looked at the current arrangements for SBA of practical skills and found that highly predictable assessments were leading to narrow teaching of these skills. Most students were getting similar results bunched around the top mark scale and the assessments were open to malpractice (Opposs, 2016, p. 59).

In GCSE English, the marks for spoken language would not contribute to the overall grade and would be reported separately from the externally marked examination.

The DfE has in recent times made efforts to reduce the impact on workload caused by teacher generated data for centralised data collection, much of which has been driven by concerns over teacher recruitment and retention (see for example the reports of the DfE workload review groups, 2016). Further, Ofsted has reduced its inspection focus on data favouring instead a re-focus on the breadth and balance of the curriculum offered by schools and has issued guidance on what it describes as 'myths' around expectations of data requirements including:

> There are no predetermined expectations on how schools present performance information or data.

> Inspectors do not expect school leaders to set teacher performance targets based on commercially produced predictions of pupil achievement, or any other data set, from which they would then hold teachers to account (Ofsted, 2018., viewed on-line).

Despite the reduction in school based data previously required for National Curriculum assessments, and the reduction in teacher based assessments for use in general qualifications, schools continue to generate excessive data for internal purpose (see for example the DfE, 2017a & 2018f). However, much of this appears to be driven within schools rather than imposed from without. A recently published survey on data use in schools reported that 66% of primary schools and 63% of secondary schools use commercially produced pupil progress trackers with 12% and 30% respectively using

Excel as a tracking tool (n=2,135). Only 5% and 10% respectively reported that they do not use any tracking tools (FFT/education datalab, 2019, p. 5). 52% (n=2,969) of respondents reported that their school collected data on a half-termly or more frequent basis with 44% collecting data termly (p. 6). Forty-one per cent (n=2,431) of teachers felt that their school was collecting more data than are regularly able to use. Only 9% of respondents (n=2,431) thought that their school should collect more data. However, the FFT/datalab report stated:

> It is curious that one-in-five headteachers would collect data less frequently than they currently report doing, if left to make their own choice! This does suggest that some headteachers perceive some (real or imagined) outside pressures regarding data collection (p. 6).

According to the report, much of the data is used to predict the future performance of pupils with 43% of respondent (n=1,724) reporting that the data was given to students in year 7 or 8 to set GCSE grade targets. However, only 10% of teachers reported supporting such use with 4 in 10 stating it should not be shared until year 10 or later (p. 23). Of particular note, the report concludes that almost all teachers reported that the analysis of assessment data was central to their work as teachers, yet the majority of teachers in key stage 3 reported that such analysis does not play a regular part in informing their classroom practice:

> So, the contrast is quite interesting. On the one hand, the majority of teachers value the role of assessment data in helping them do their job. On the other hand, a minority seem to be routinely using test data to inform their classroom practice (p. 26).

However, the report does recognise that the analysis of the data is complicated by the language of assessment that can include teacher judgment of classwork as well as formal tests (ibid., p. 26).

These more recent changes challenge the role of teachers as expert professionals in high-stakes examinations with questions raised over their trustworthiness to administer, mark and develop assessments. Changes to the assessment regimes for National Curriculum assessments and qualifications have restricted the role played by teachers as they; "…can no longer be relied upon to produce valid and reliable outcome" (Opposs, 2016, p. 60). And even where assessment practices and data generation are not dictated by outside

agencies, the culture of performativity has become ingrained as the power/knowledge dimension has become woven into the very fabric of the English education system even to the point that:

> For teenagers in particular, school and education have, at their core, intense competition. Their contemporaries may be friends, but they are also rivals. And as education becomes increasingly important in determining people's lives, this rivalry becomes increasingly consequential (Wolf, 2008, p. 19).

4.6 Chapter 4 summary

Foucault's work does not provide a coherent methodological framework but presents a range of tools '…for intervening within contemporary discourses of power' (Ball, 2013. p. 4). These tools can provide new insights into the role of teachers as educational assessors over time. This approach does not overlay the period of time in question with any notion of progress towards an ultimate regime of educational assessment. Rather it seeks to explore the manifestations of change as contingent drawn from a complexity of interactions and options. The aim is to discover how the power-knowledge dimension as described by Foucault has influenced change in the role of teachers' assessment over a given period of time in order to make use of such information today. Educational assessment has a history and is therefore contingent; as such it is built around options and open to change (see Allen, 2012; Bailey, 2016)).

Foucault's discussion of the examination focuses on the impact on the individual:

> …it is the individual as he may be described, judged, measured, compared with others, in his very individuality: and it is the individual who has to be trained or corrected, classified, normalized etc. (1977a, p. 191).

However, since 1943, there have been distinct shifts in the education system. This chapter has presented three distinctive epochs in the way educational assessment has been used and the role of teachers in its application to high-stakes assessments. In the first period or episteme, there was a vision in which teachers would be the main source of determining the format and application of assessment in world where qualifications were seen as only for an elite few. Educational assessment as a means for determining the route through secondary education was to be in the hands of teachers. However, this goal as presented

by Norwood was not realised and instead there was a period of growth in what became known as the 11+ with the teaching profession showing no major opposition. However, as the second period illustrates, despite the role envisioned for teachers not being realised, there was a growth in the numbers of pupils taking examinations leading to alternative provision designed for a much wider group of students and increased control of examination content by teachers – what has been described a 'golden age' of teacher autonomy in terms of assessment and the curriculum. But the adoption of public accountability measures in the third period created a particular form of focus on schools, teachers, governors, government ministers, parents – and pupils - in a web of power relations. The arguably insidious growth of publicly available performance data has developed what Ball has described as the; "…new management panopticism and the new form of entrepreneurial control" (Ball, 2003. p. 219).

What does appear constant throughout these periods is the way Foucault's conception of power formed a diffusion of control. However, power has not been wholly used in a top-down structure as teachers have frequently rebuffed the assessment arrangements introduced by governments and spurned opportunities for wider control at various points in time. Yet teachers have used educational assessment for their own means and those of external agencies even when describing the ill effects of examinations, tests and accountability measures. Whilst teachers recognise the centrality of assessment in teaching and learning, workload impositions have frequently been raised in opposition to various approaches to assessment, but often at the expense of reduced teacher professionalism. Further, the diversification of teacher representative bodies has prevented a collective powerful and cohesive voice. The quote from Kenneth Baker earlier in this chapter illustrates the point. And since the early 1990s, the use of centrally derived curricula and curriculum content has undermined expertise and resulted in swathes of teachers who are products of the system in that they have been normalised by the regime of qualifications and accountability.

Over time, there have been important developments in the generation and quality of high-stakes testing instruments for example, a greater understanding of how performance standards operate over time. But the focus on externally set and administered examinations and tests has seen increased reliability, but at the expense of validity – and teacher expertise. Although teachers may not now quite recognise the reference to 'freedom', Lawton's words regarding examinations holds currency:

It is paradoxical that in a system where teachers are so proud of their freedom, they tolerate domination of the curriculum by examinations which are externally controlled (1980, p. 106).

As England and indeed other countries reflect on the contingency arrangements implemented in response to COVID-19, the place of teacher bases assessments will little doubt form part of the debate.

Chapter 5: ITT Provider questionnaire survey: data analysis

5.1 Introduction

This chapter presents an analysis of the data derived from the questionnaire survey of Initial Teacher Training (ITT) providers. A pdf copy of the questionnaire is attached at Appendix 3 for reference. The survey was designed with the aim of providing greater insight into the theoretical, technical and practical aspects of educational assessment covered in ITT courses.

A total of 49 responses were received from individuals working in ITT provider settings: the responses do not therefore necessarily constitute the views of their institution. Respondents were self-selecting and may have elected to respond for a variety of reasons including positive or negative views on their course content relating to educational assessment. It is important to note that not all respondents answered every question, hence percentages are calculated out of those for whom data was available (i.e., excluding missing responses by item).

The presentation of this chapter follows a similar format to that used for the survey and covers the background information on respondents' institutions, perceptions on the Carter Review of ITT (2015), the content of ITT courses offered by respondents' institutions and course development and delivery. A second survey was designed for teachers covering similar areas; the analysis of the teacher survey is presented in Chapter 6.

5.2 Respondents – background information

Question 1 of the survey asked respondents to identify the routes into teaching offered by their institutions. Figure 1 below summarises the responses. The majority of institutions were involved in offering School Direct (Fee) and HEI-led Postgraduate routes.[29] The responses broadly reflect the proportions of new entrants in initial teacher training by route 2016/17 (DfE, 2016b).

---

[29] Percentages calculated out of 49 total respondents to the survey.

Figure 5.1: Question1, routes into teaching offered by providers
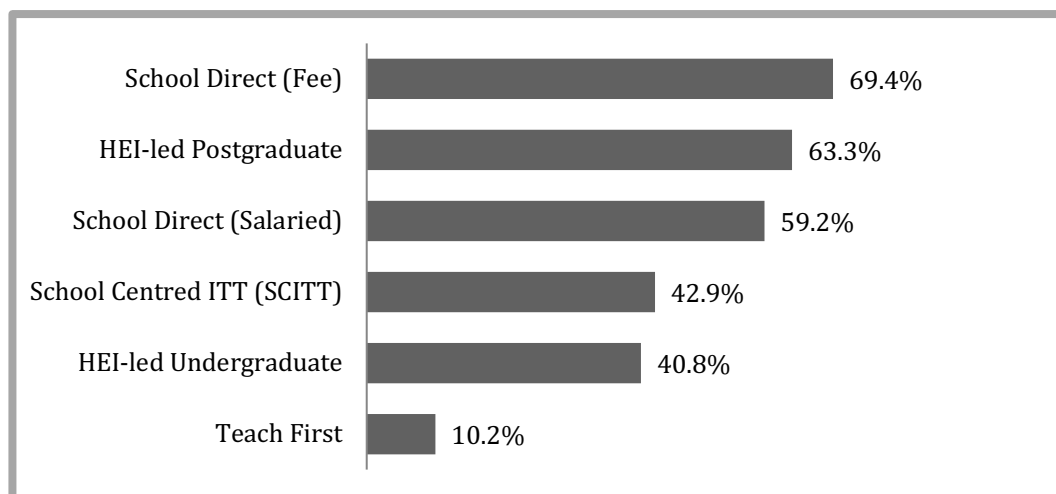
Question 2 asked school-led providers if they partnered with a Higher Education Institute of which 16 answered yes, 24 stated that it was not applicable, and 9 responses were left blank.

Question 3 asked respondents to identify which phases of education and/or subjects were covered in their ITT courses and the approximate number of students**.** 46.8% responded that their ITT courses covered just the primary phase of education and 17.0% just secondary, while 36.2% covered both.[30]

Of the 79.6% (n = 39) of respondents who covered primary education as part of their ITT provision, 41.2% had courses numbering under 100 students, 35.3% between 100-499, and 23.5% over 500 students.[31]

Of the 51.0% (n = 25) of respondents who covered secondary education as part of their ITT provision, 35.3% had courses numbering under 100, 52.9% between 100-499, and 11.8% over 500.[32] (According to the 2017 DfE Workforce census[33], secondary teachers made up 45% of the total number of teachers in all maintained schools in England.)

---

[30] Out of 47 total respondents to this question.

[31] 34 out of 39 primary (87.2%) respondents provided approximate numbers of current students.

[32] 17 out of 25 secondary (68.0%) respondents provided approximate numbers of current students.

[33] The total number of full time equivalent (FTE) teachers in state funded schools in 2017 was 451,900: the number of FTE nursery and primary teachers was 221,100. There were 204,200 FTE secondary teachers in 2017. Source DfE, 2018. School Workforce Data. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/719772/SWFC_MainText.pdf

The top two individual subjects covered in respondents' ITT provision were History (57.1%) and Mathematics (57.1%), with English, Geography and Modern Foreign languages occupying the next three spots each on 55.1%.[34] See Figure 5.2.

Figure 5.2: Question 3, subjects covered by respondents' ITT provision
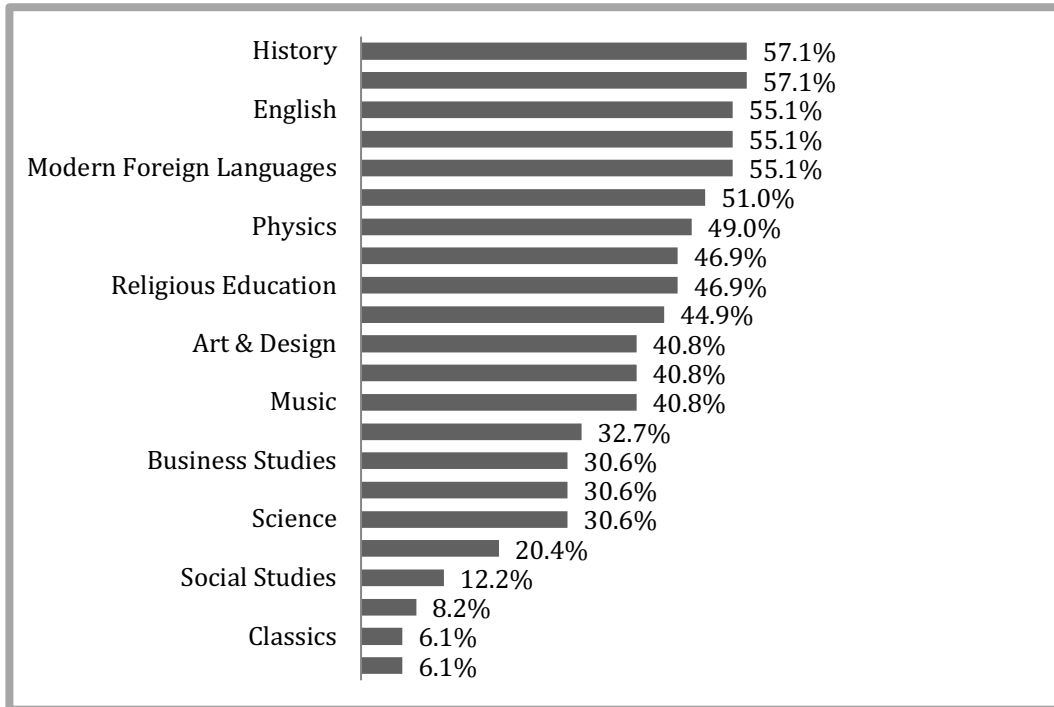
Question 4 asked respondents to identify the teaching qualifications offered by their institutions. As shown in the Figure 5.3 below, the majority of respondents' ITT offer covered the Post Graduate Certificate in Education (PGCE) and Qualified Teacher Status (QTS).

Figure 5.3: Question 4, qualifications offered by respondents' institutions

---
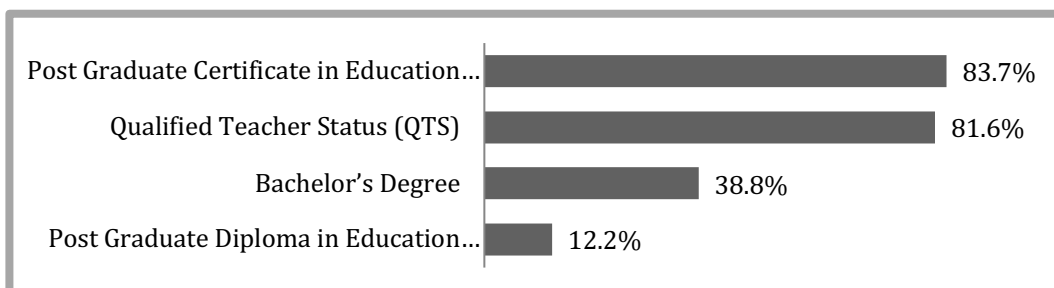
[34] All percentages calculated out of 49 total respondents to the survey.

5.3 The Carter Review of ITT

In the second section of the survey, question 5 asked respondents how strongly they agreed with the conclusion of the Carter Review of Initial Teacher Training that:

> Of all areas of ITT content, we believe the most significant improvements are needed for training in assessment (Carter, 2015, p. 9).

Question 6 also asked how much they agreed with Carter that there are significant gaps in the capacity of schools or ITT providers in the theoretical and technical aspects of assessment. Responses to questions 5 and 6 are shown in Figure 5.4.

Of the 49 responses to question 5, 59.2% agreed that in all areas of ITT content, the most significant improvements are needed for training in assessment. The highest percentage of respondents to question 6 agreed there are significant gaps in the capacity of schools regarding assessment (66.6%), with a much lower percentage of 26.5% agreeing that ITT providers had significant gaps in these areas.[35]

Figure 5.4: Questions 5 & 6, the extent of agreement with the Carter Review conclusions

---
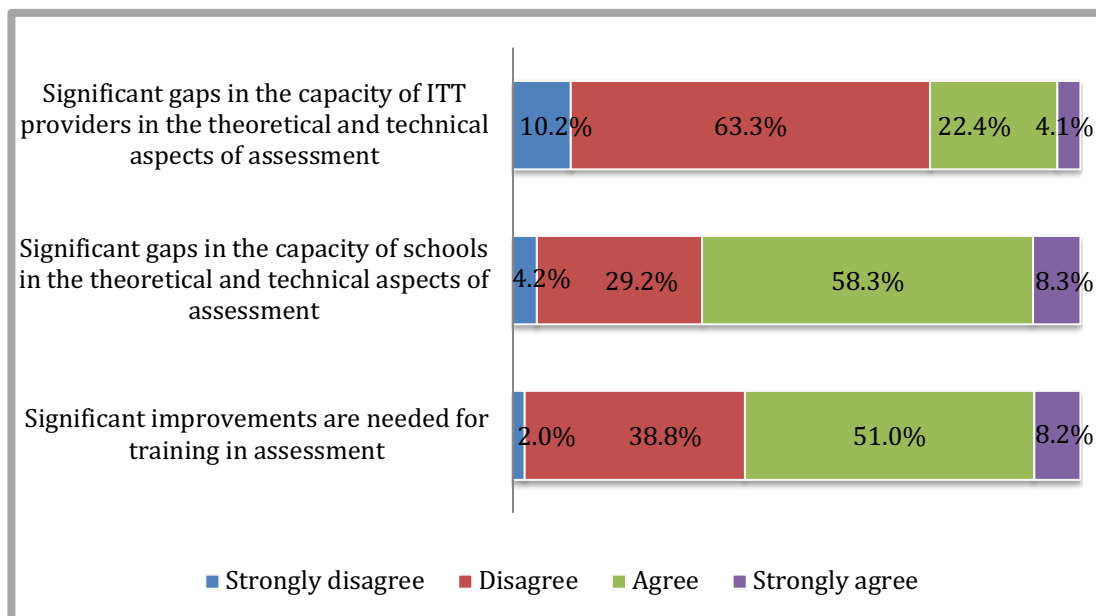
[35] There were 49 respondents to the significant improvements and ITT provider statements, and 48 to the schools statement.

5.4 Educational assessment content in ITT courses

Questions 7 to 12 of the survey asked respondents about whether specific aspects related to educational assessment were included in their ITT provision, and the extent to which they felt confident their courses fully prepare their students for their first years in teaching. The questions examined assessment aspects relating to:

- models of learning (question 7);
- understanding progression (question 8);
- principles of educational assessment (question 9);
- the uses of educational assessment outcomes (question 10);
- assessment design (question 11); and
- educational assessment and accountability (question 12).

Overall, respondents felt most confident (highest 'very confident' ratings) that their courses had prepared trainee teachers in:

- the differences between formative and summative uses of assessment (75% 'very confident');
- the use of assessment to give pupils regular feedback, both orally and through accurate marking, and encourage pupils to respond to the feedback (74.5%);
- the use of formative and summative assessment to secure pupils' progress (74.5%);
- the effective use of assessment to give effective and efficient oral and written feedback to pupils and parents (72.7%).

The areas in which respondents felt least confident (highest 'not at all confident' ratings) were:
- utility (11.1% 'not at all confident');
- the design of standardised tests (such as those that produce a reading age) (11.1%);
- standardisation – purpose and process (9.5%);
- criterion- and norm-referencing (9.1%);
- how nationally standardised summative assessment can be used to help teachers to understand national expectations and assess their own performance (8.3%);

- using externally set standards of educational performance for example Interim Teacher's Assessment Frameworks, exemplification of standards from STA or awarding bodies (8.3%).

Each of these areas is covered in more detail below.

5.5 Models of learning

Question 7 related to the provision of models of learning as part of course content and levels of confidence that the course fully prepares students for their first years in teaching. Over 90% of respondents stated that their ITT provision included 'knowledge and understanding of how pupils learn and how this impacts on teaching' (100%), 'stages of development within subjects' (95.7%), and 'approaches for strengthening pupil memory such as short tests and making effective use of questioning' (93.6%).[36]

Figure 5.5 below shows ITT provider respondents' confidence that they had fully prepared their students on these aspects; they were most confident (almost 70% selecting 'very confident') regarding 'knowledge and understanding of how pupils learn and how this impacts on teaching'.[37]

Figure 5.5: Question 7, ITT providers' confidence that their ITT course fully prepare trainee teachers regarding models of learning

---
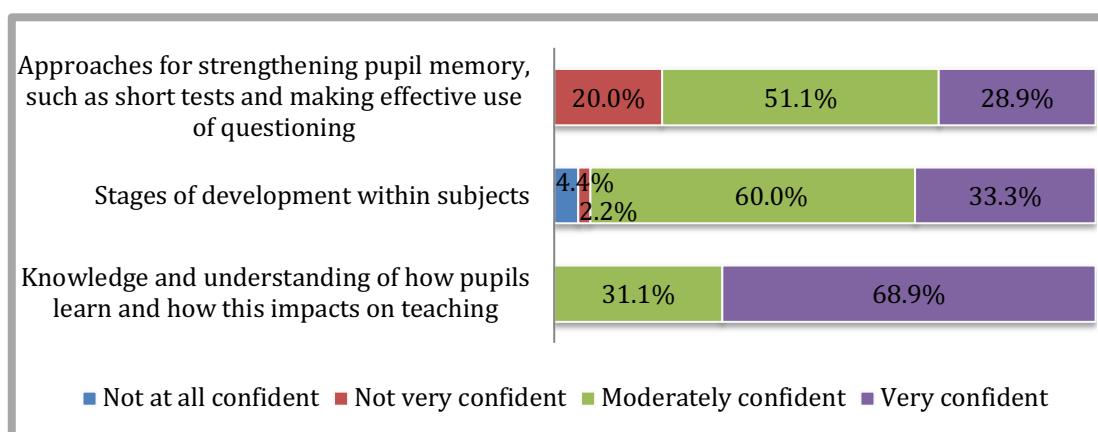
[36] Number of respondents for inclusion of aspects: 'knowledge and understanding…': 48; 'stages…': 46; 'approaches…': 47.

[37] 45 responded to each of these confidence questions.

5.6 Understanding progression

Question 8 related to the provision of the understanding of progression as part of course content and levels of confidence that the course fully prepares students for their first years in teaching. Over 90% of ITT providers who responded to this question stated that aspects relating to understanding progression were included in their ITT provision.[38] The highest percentage were 'very confident' regarding 'the use of assessment to give pupils regular feedback' and 'the use of formative and summative assessment to secure pupils' progress' (both at 74.5%), see Figure 5.6.

Figure 5.6: Question 8, respondents' confidence their ITT course fully prepares trainee teachers regarding understanding progression

---
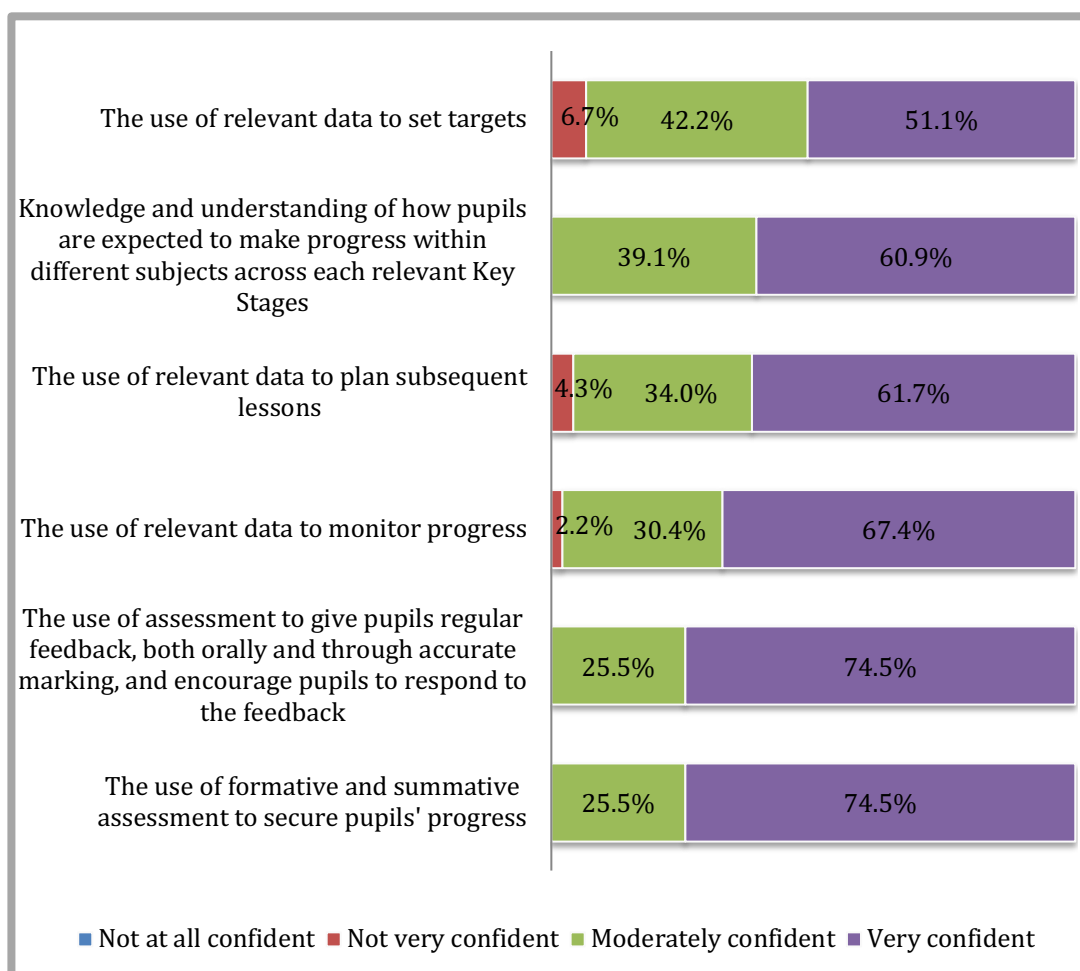
[38] 48 responded to the 'knowledge and understanding…' and 'the use of relevant data to set targets' items; 49 responded to all other items. Inclusion in ITT provision: 'knowledge and understanding…', 'the use of assessment…', and 'the use of formative and summative assessment…' were included in the ITT provision of 100% of respondents; 98% said their ITT provision included 'the use of relevant data to plan subsequent lessons'; 93.9% and 93.8% said their provision included 'the use of relevant data to monitor progress' and '…set targets', respectively.

5.7 Principles of educational assessment

Question 9 related to the provision of principles of educational assessment as part of course content and levels of confidence that the course fully prepares students for their first years in teaching. Over 95% of respondents' ITT provision includes 'the differences between formative and summative uses of assessment' and 'theory of assessment', the aspects where respondents felt most confident (75% and 50% 'very confident' ratings, respectively),[39] see Figures 5.7 and 5.8.

Figure 5.7: Question 9, inclusion of assessment aspects relating to principles of educational assessment on ITT courses[40]

Aspect	Aspects not included in our course	Aspects included in our course
criterion- and norm-referencing	42.6%	57.4%
The design of standardised tests (such as those that produce a reading age)	40.4%	59.6%
Using externally set standards*	40.4%	59.6%
utility	38.3%	61.7%
bias	31.9%	68.1%
reliability	29.8%	70.2%
How nationally standardised assessment can be used**	28.6%	71.4%
validity	27.7%	72.3%
standardisation – purpose and process	25.0%	75.0%
setting standards of educational performance	20.8%	79.2%
moderation - purpose and process	12.5%	87.5%
Theory of assessment – why, when, how to assess and report	4.1%	95.9%
The differences between formative and summative uses of assessment	0.0%	100.0%

---

[39] Total number of respondents for inclusion and confidence for all aspects in the following assessment aspect sections are included in a table in the Appendices: Principles of educational assessment; The uses of educational assessment outcomes; Assessment design; and Educational assessment and accountability.

[40] * The full text of this item is: Using externally set standards of educational performance for example Interim Teachers Assessment Frameworks, exemplification of standards from STA or awarding bodies. ** The full text of this item is: How nationally standardised summative assessment can be used to help teachers to understand national expectations and assess their own performance.

Figure 5.8: Question 9, respondents' confidence their ITT fully prepares trainee teachers regarding principles of educational assessment[41]
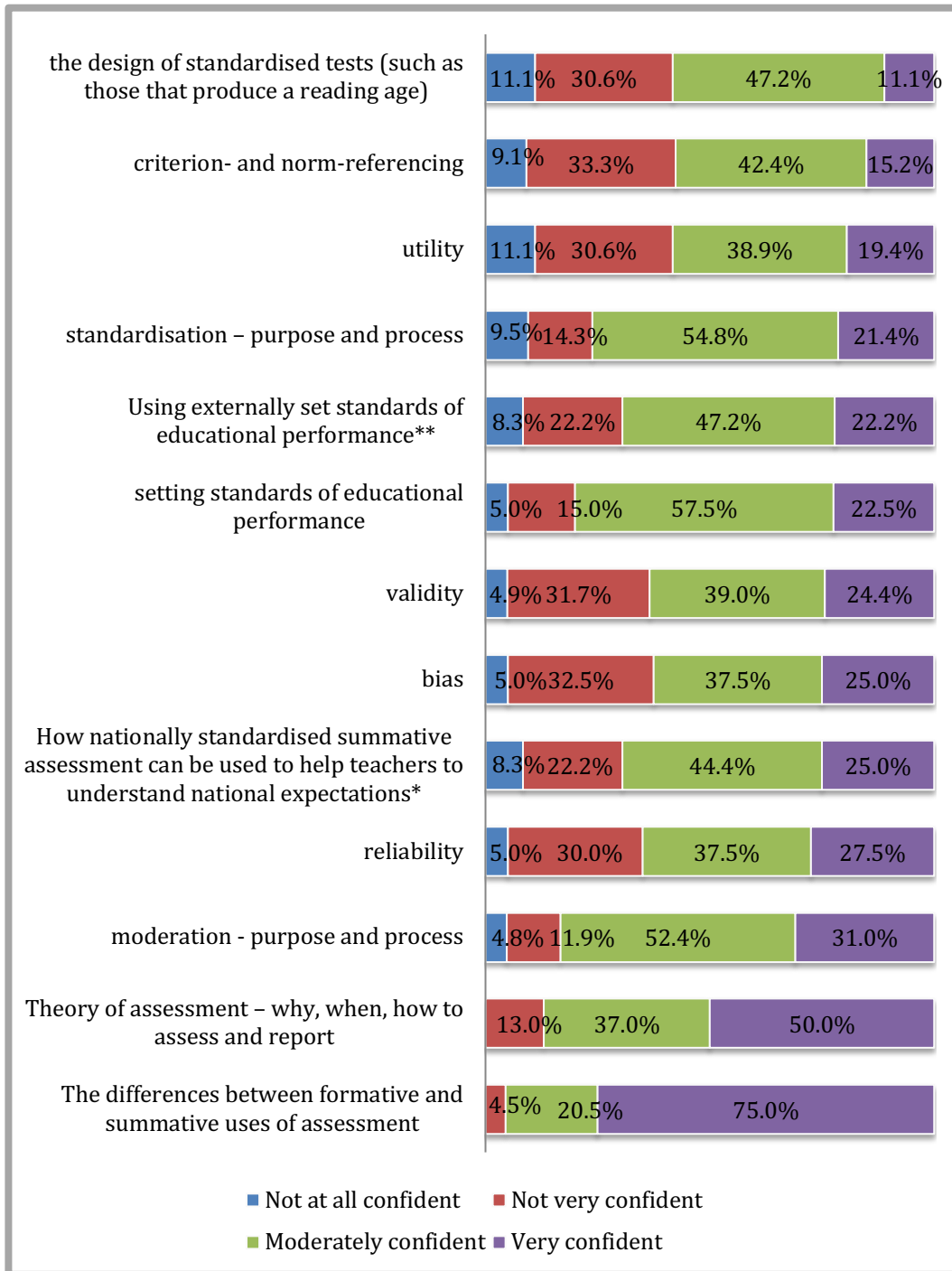
5.8 The uses of educational assessment outcomes

Question 10 related to the uses of educational assessment outcomes as an area included in ITT providers' course content and their levels of confidence that the course fully prepares students for their first years in teaching. All respondents reported the inclusion of all aspects shown in Figure 5.9 below in their course provision with the exception of 'using knowledge of how pupils make progress…' at 95.9%, with 89.8% including 'common inferences/uses of assessment' and 67.3% including 'the use and misuse of assessment outcomes and the ethics of assessment'.

Respondents were least confident in 'the use and misuse of assessment outcomes and the ethics of assessment' with 27.1% reporting that they were 'not very confident'. Alternatively they were most confident about 'the effective use of assessment to give effective and efficient oral and written feedback to pupils and parents' with 72.7% of respondents reporting that they were 'very confident' in this aspect: see Figure 5.9.

Figure 5.9: Question 10, respondents' confidence that their ITT fully prepares trainee teachers regarding the uses of educational assessment outcomes
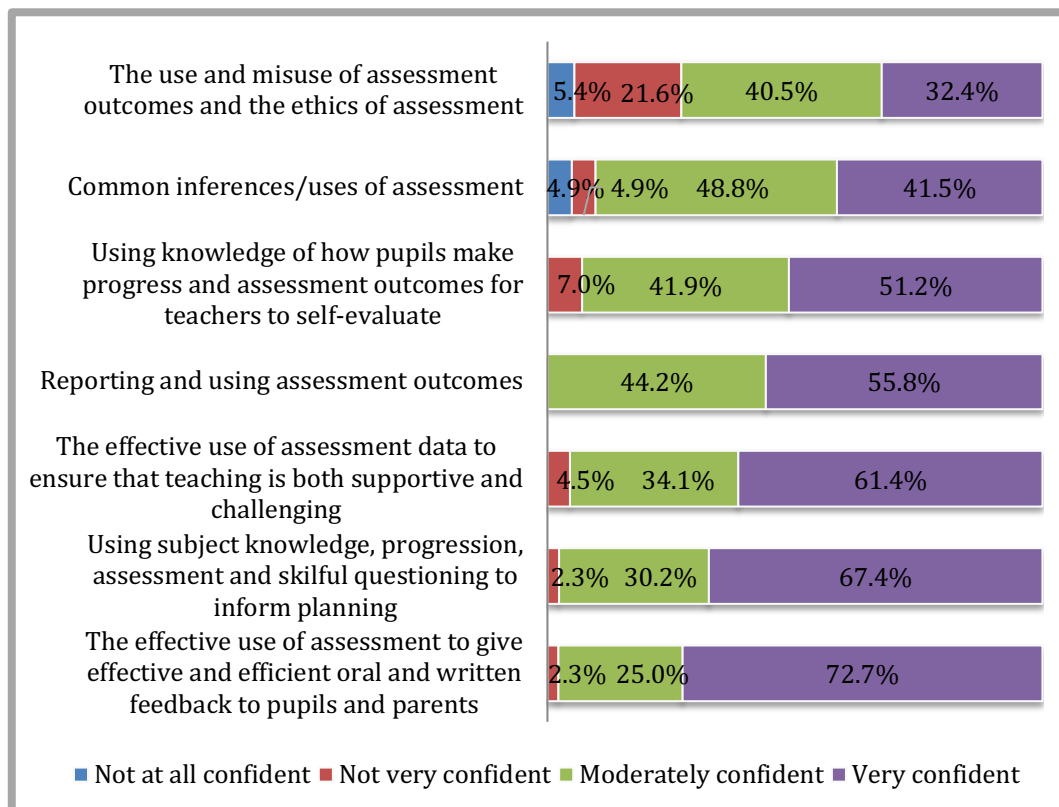
5.9 Assessment design

Question 11 of the survey related to assessment design as an area included in ITT providers' course content and their levels of confidence that the course fully prepares students for their first years in teaching. The one aspect of assessment design included on all respondents' ITT provision was the 'knowledge and understanding of how to assess the relevant subject and curriculum areas, including statutory assessment requirements': see Figure 5.10. This was also the area with the highest percentage of ITT provider respondents reporting that they were 'very confident' that they had fully prepared their teacher trainees for the first years of teaching. A smaller percentage (72.9%) of ITT providers' courses included 'how to identify key constructs and valid approaches to assessment'. See Figure 5.11.

Figure 5.10: Question 11, inclusion of assessment aspects relating to assessment design in ITT courses
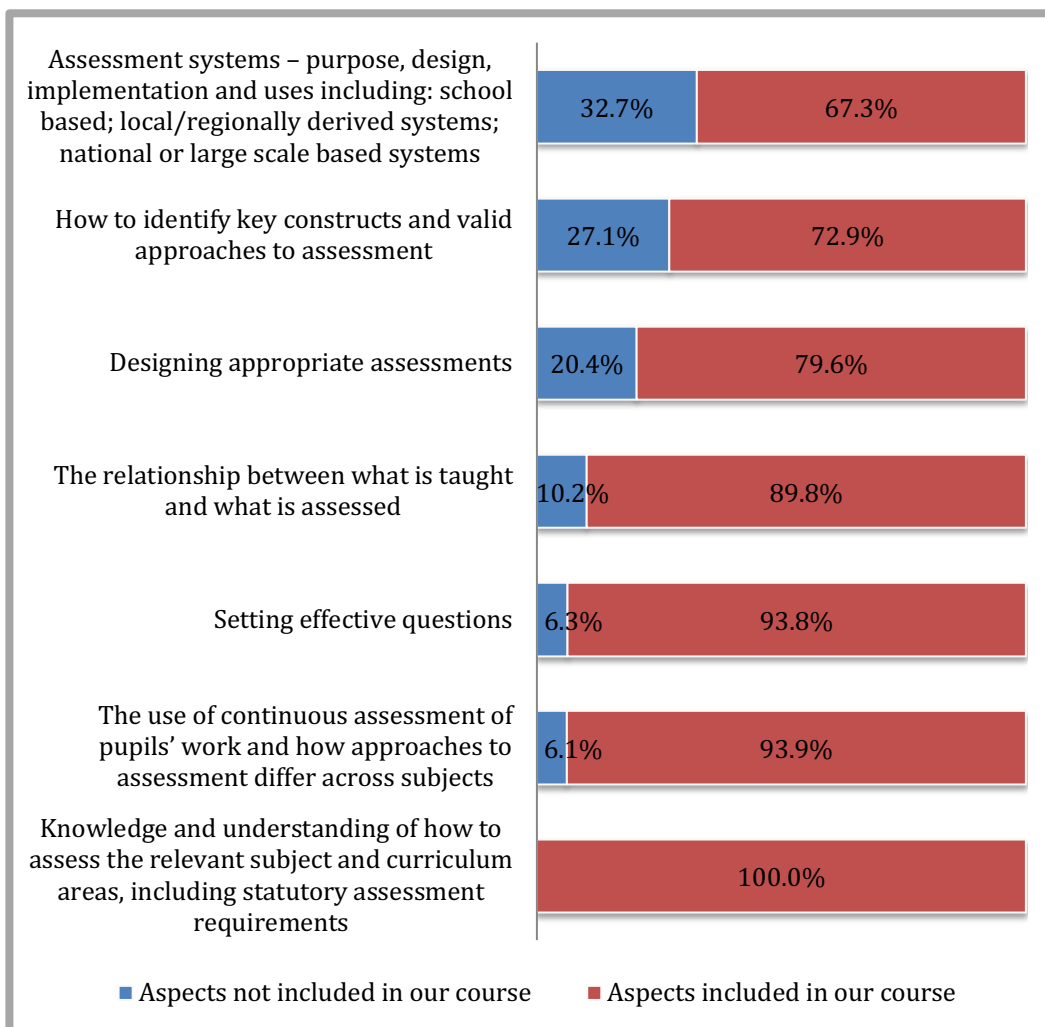
Figure 5.11: Question 11, respondents' confidence their ITT fully prepares trainee teachers regarding assessment design
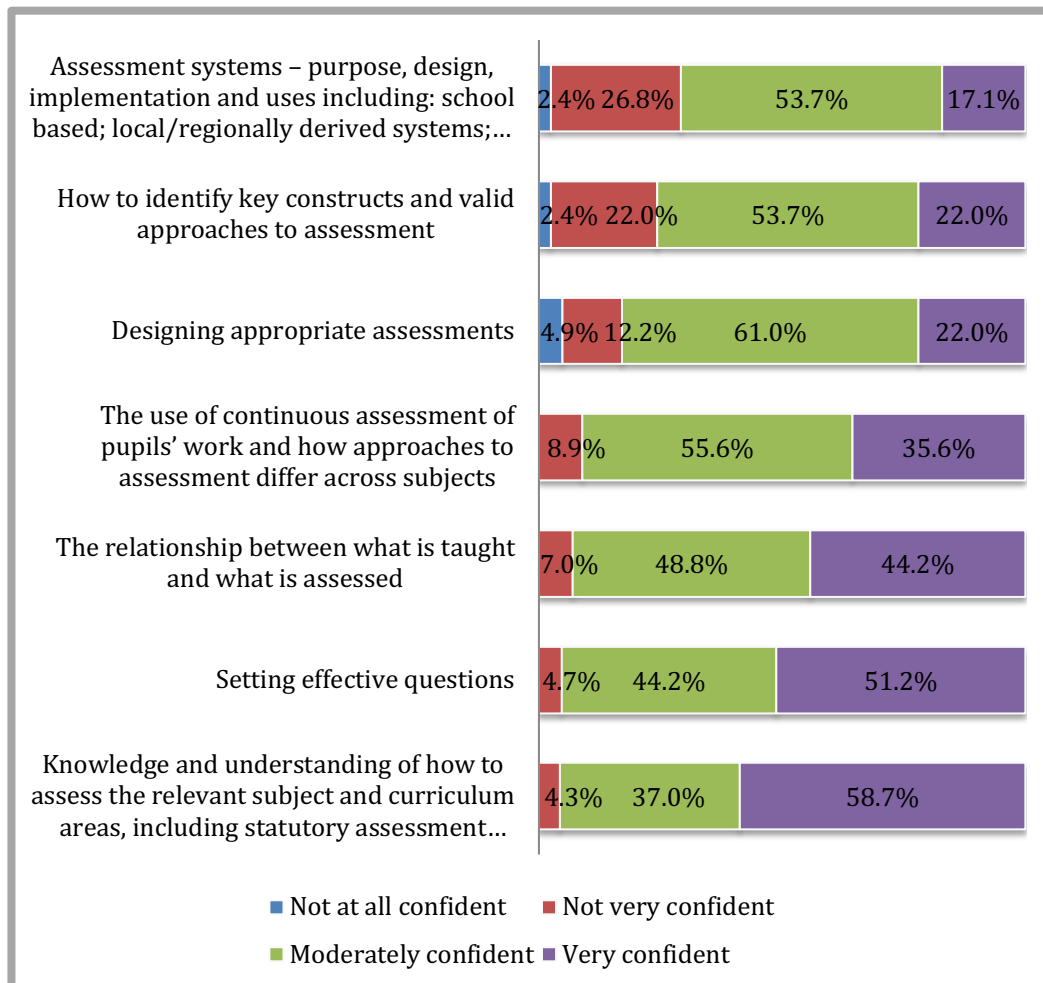
5.10 Educational assessment and accountability

Question 12 covered aspects of educational assessment and accountability covered in ITT courses and levels of confidence that students were adequately prepared for their first years in teaching. The aspects of educational assessment and accountability shown in Figure 5.12 below formed part of over 85% of ITT respondents' provision, with the exception of 'knowledge and understanding of the assessment system and administrative requirements for GCSE and GCE examination courses' (59.6%).[42] There was variation in

---

[42] Inclusion: 95.8% 'the requirements of the national curriculum, national Key Stage assessments and/or specifications for public examinations for the subject(s) and phase(s) they will be teaching'; 91.8% 'knowledge and understanding of the National Curriculum assessment system and administrative requirements'; and 85.7% 'knowledge and understanding of the statutory accountability measures'.

respondents' confidence across these aspects, with the lowest percentage of respondents 'very confident' their ITT fully prepares trainee teachers in 'knowledge and understanding of the statutory accountability measures' (30.2%).

Figure 5.12: Question 12, respondents' confidence their ITT fully prepares trainee teachers regarding educational assessment and accountability
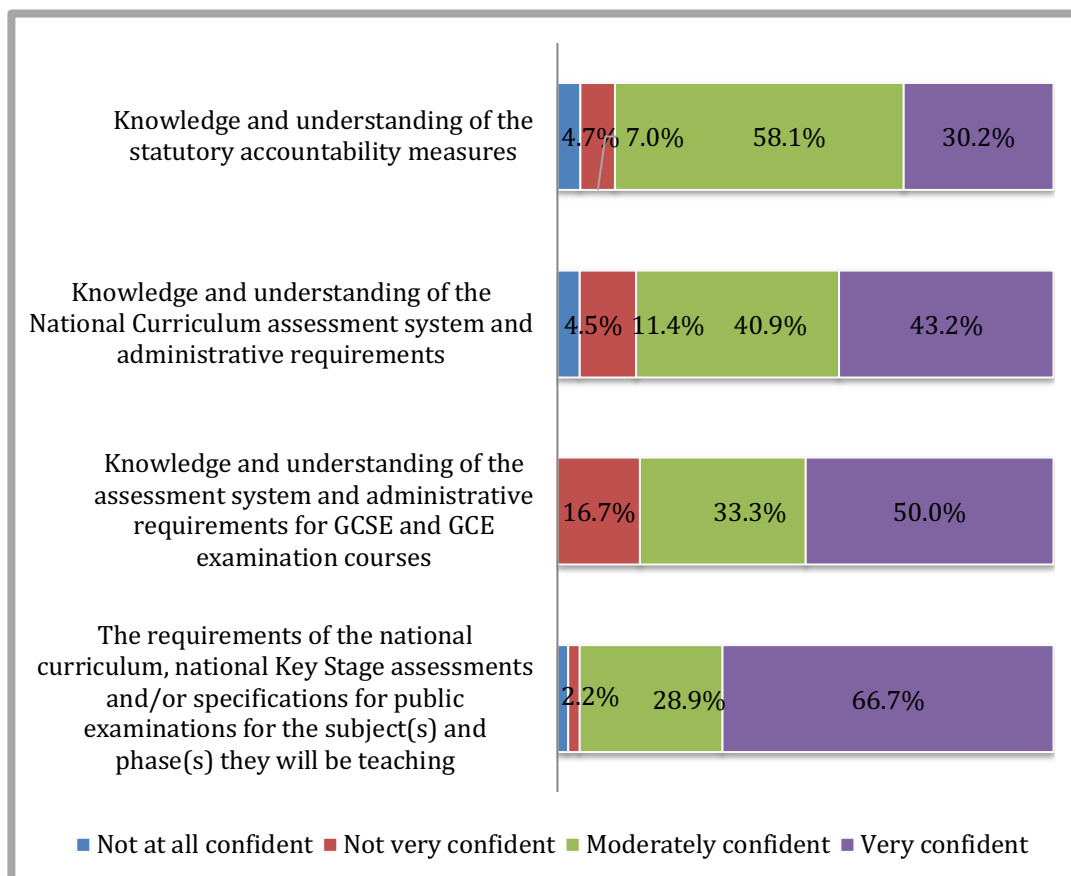
5.11 Course development

Question 13 of the survey asked respondents if they considered any of the assessment aspects identified earlier in the survey as unnecessary for a course in initial teacher training. Respondents were given free response boxes to name the aspects and further boxes to explain why.

As with the teacher survey, underlying most responses in this section was the limited timescale of the ITT course and how ITT must therefore focus on priorities. There was some concern amongst ITT provider respondents that assessment is a subject in its own right, therefore delving into the detail during the ITT year could be overwhelming and negatively impact on teacher trainee confidence as NQTs.

Specific aspects of assessment that were mentioned included: the administration or design of standardised tests as these were not considered a priority for the ITT year or for Primary focused courses; or even that they were "…best left until they need to do it" when they could be supported by a head of department; accountability measures; GCE (as some schools might not offer KS5); continuous assessment; and how assessment approaches may vary by subject (while an understanding of assessment approaches across subjects might not be a priority, one respondent did hope that trainees would be curious and consider this).

One other respondent underlined the importance of the term 'initial', which was echoed in others' suggestions that aspects such as assessment design could be covered in the first few years of teaching. However, there was some concern regarding the variation in quality of Continuous Professional Development after the ITT period.

Other responses covered broader issues regarding the inclusion of assessment in ITT: for example, understanding how one's assessments fit into the student's achievement across the curriculum, and being wary of 'feeding the data fire' (this response included mention of the DfE's "Reducing Unnecessary Teacher Workload" team).

Question 14 asked if there were aspects of educational assessment theory and practice not listed in the previous sections that should be added. Again, respondents were given free response boxes to name the aspects and further boxes to explain why. Respondents mentioned:

- the experience, including motivation and the emotional impact, of assessment (including summative) for pupils and parents;
- the ethics of "the structures, systems and approaches for assessment";
- drawing on evidence-based research;
- national and international assessment issues (for context and to underline that national approaches to assessment are decisions);
- diagnostic assessment and meeting the individual needs of the pupil;
- managing the associated workload (vis teacher well-being; and preparing trainee teachers to take a critical approach to the demands of school assessment systems);
- assessment and pedagogy (versus curriculum).

5.12 Educational assessment course delivery

Question 15 asked respondents if they would be prepared to share the content of their course relating to educational assessment theory and practice and if the details were already in the public domain.

Question 16 of the survey asked respondents about how confident they were that their institution had the right level of expertise to deliver all aspects of educational assessment theory and practice to the same depth and quality as all other aspects of their ITT course. 33.3% were 'very confident', 62.5% 'moderately confident' and 4.2% were 'not very confident'.[43]

Question 17 asked if there were any constraints that prevented respondents' institutions from delivering what they believe would be an ideal course of instruction in educational assessment theory and practice. Respondents were also given the opportunity to suggest how the constraints could be alleviated. The most commonly cited constraint was the limited time available in ITT courses with suggestions that teacher development should continue into the early years of practice. It was also suggested that instruction in some elements of educational assessment would be more productive once teachers had developed some experience of the classroom. It was also noted that more guidance to support schools and mentors would be beneficial.

Question 18 asked respondents to the ITT provider survey if their institution or any members of their institution engage in research on educational assessment theory and practice. Out of 28 respondents, 54% said individuals, 25% both individuals and the institution, and 21% just the institution.

Respondents were invited to provide any relevant supporting information, and those who did talked about who was involved in the research (e.g., an unspecified institution, specific individuals – not named here to retain respondent anonymity); the types of research (e.g., exam stress and anxiety); and details of the approach (e.g., via 40% of staff time allocated to research; via a research school; a working group on marking and assessment practice). At the end of the survey, respondents were given the opportunity to receive a copy of the final survey analysis or to take part in a follow-up interview. Twenty-five of the 48 respondents to this question stated that they would like a copy of the survey report and/or

---

[43] Out of 48 total responses to this question.

be prepared to take part in follow-up activities. These respondents will be contacted following the submission of this thesis.

5.13 Chapter 5 summary

Whilst just under 60% of respondents agreed with the Carter Review that the most significant improvements in ITT are needed in assessment training, almost three-quarters were of the view that ITT providers had no significant gaps in this area. However, two-thirds agreed there are significant gaps in the capacity of schools regarding assessment. The survey exposes some significant gaps in coverage of assessment principles, for example almost 43% do not include instruction on criterion- and norm-referencing with almost one third not including bias and assessment design. Gaps in provision related strongly to areas where respondents reported lower confidence that their courses fully prepare trainees. However, ITT providers were generally confident in their capacity to deliver aspects of educational assessment in their courses with over 70% of respondents disagreeing with the Carter Review conclusion that there are significant gaps in the capacity of ITT providers in the theoretical and technical aspects of assessment.

A key concern amongst ITT providers was the impact of time constraints on their facility to include more detailed study of educational assessment with some noting that assessment is a subject in its own right and that attempts to add more detailed study could overwhelm teacher trainees and result in a negative impact on their confidence. It was therefore suggested by some that training and development should continue into the early years of teaching through well-supported mentors.

Chapter 6: Teacher questionnaire survey: data analysis

6.1 Introduction

This chapter presents an analysis of the data derived from the questionnaire survey of teachers. A pdf copy of the questionnaire is attached at Appendix 6 for reference. The survey focused only on the theoretical, technical and practical aspects of educational assessment delivered as part of ITT courses taken by respondents along with more general questions regarding their ITT background, qualifications and levels of trust.

The following sections present an analysis of the 213 responses. It is important to note that not all respondents answered every question, hence percentages are calculated out of those for whom data was available (i.e., excluding missing responses). Furthermore, respondents self-selected to respond to the survey and may have participated due to strong feelings, either positive or negative, regarding the assessment content of their ITT provision.

The presentation of this chapter follows the format of the survey covering background information on respondents' routes into teaching including the timing of their ITT, perceptions on the Carter Review of ITT (2015), perceptions of trust in teachers' assessment, the content of ITT courses taken by respondents, and their levels of confidence in aspects of educational assessment, and reflections on course content and delivery.

6.2 Respondents – background information

Question 1 of the survey asked respondents to identify their routes into teaching. Figure 6.1 below illustrates the responses. As shown in the figure 6.1 below, the majority of respondents had entered teaching via HEI-led Postgraduate and undergraduate routes. Twenty respondents did not answer this question; hence percentages are calculated out of 193 (91% of the total respondents n = 213). There was some misunderstanding regarding this question, as shown by responses to the following item, where respondents could describe an 'other route' they may have taken: 2 respondents answered 'B.Ed.', 5 'PGCE', and a further 8 'Graduate Teacher Programme'. The remaining 6 responses included 'college cert ed', 'DTLLS', teacher training college, 'went straight from school', and 'assessment only'.

Figure 6.1: Question 1, respondents' routes into teaching
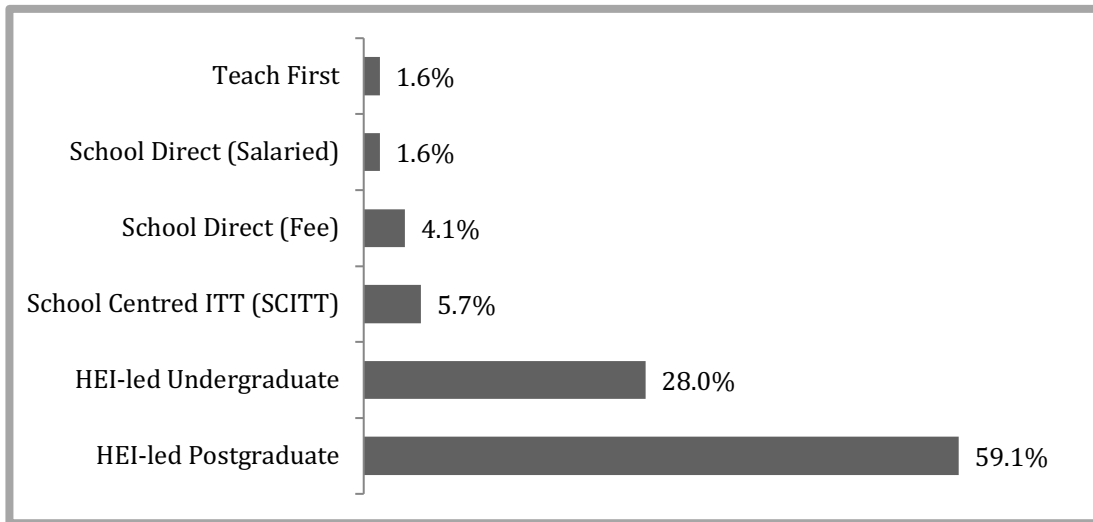
Question 2 asked respondents to identify the year in which they completed their initial training. Figure 6.2 below indicates the decades in which respondents completed their initial teacher training. 66.8% (out of 211 total respondents to this question) completed their initial teacher training since the year 2000, with approximately a third falling into the pre-2000s (33.2%). A further third trained between 2000 and 2009 (33.6%), and 33.2% trained post 2010.

Figure 6.2: Question 2, decade in which respondents completed their ITT
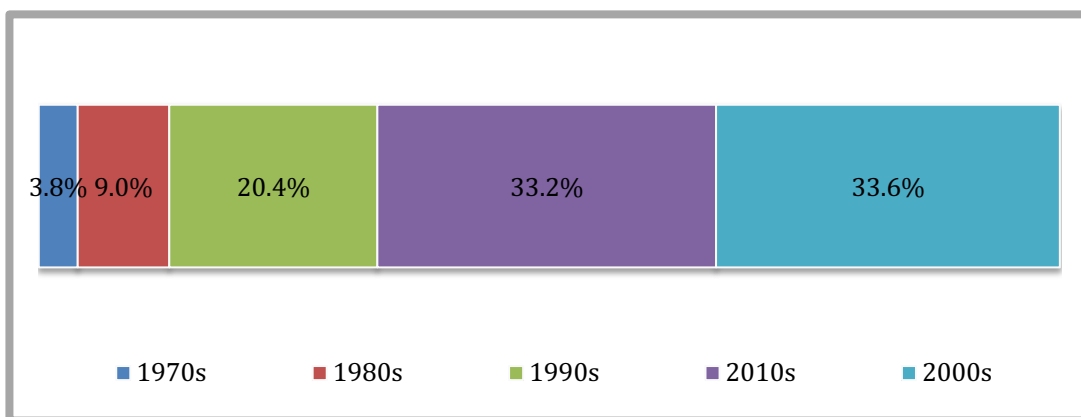
Question 3 of the survey asked respondents to name their ITT provider. 209 responses were received to this question of which 159 named a university with others providing the names of School-Centred Initial Teacher Training providers (SCITTs), individual schools and Multi-Academy Trusts (MATs).

Question 4 asked respondents to identify the phase of education for which they trained and their subject specialisms. 61.8% of respondents had been trained in the primary phase, 37.1% in secondary, and 1.1% had been trained in both primary and secondary.[44] The top three individual subject specialisms[45] in which respondents had trained were (out of a total of 213 survey respondents) English (12.2%), Mathematics (11.7%) and Science (9.9%). When subjects were categorised into clusters, e.g. science or humanities, as shown in figure 6.3 below, 22% (out of 127 total respondents to this question) had trained in science subjects.

Figure 6.3: Question 4, subjects respondents had trained in, by cluster
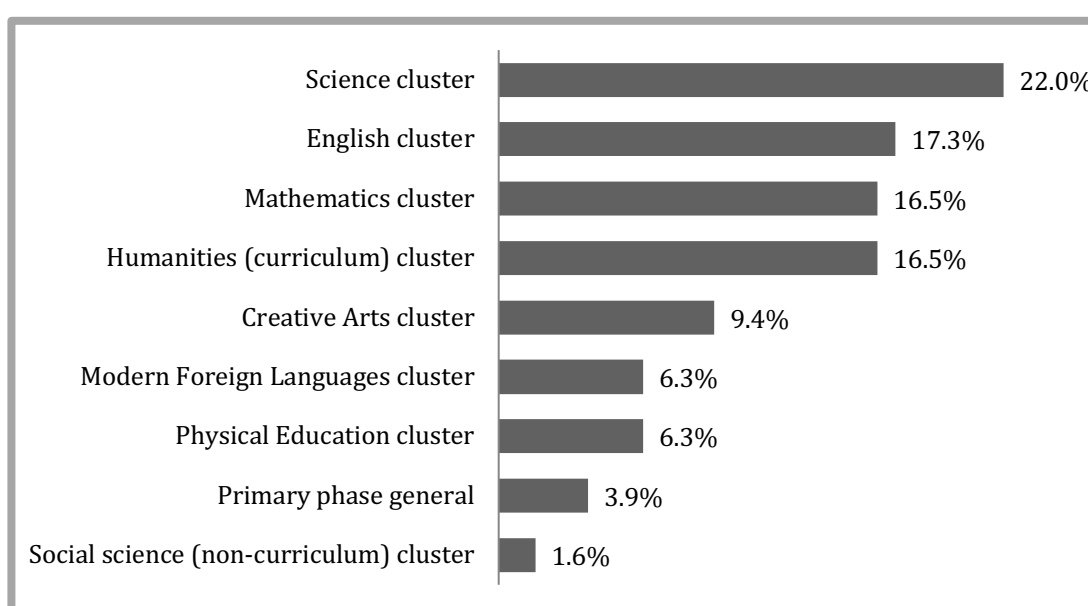
Question 5 asked respondents to identify their qualifications. As shown in the figure 6.4 below, the majority of respondents held a Bachelor's degree (90%, out of 213 total respondents), with 4% having a PhD.

---

[44] Out of 178 total respondents to this question.

[45] Respondents selected from subject options and/or provided additional detail in a comment box. Subject options offered were Art & Design, Biology, Business Studies, Chemistry, Classics, Computing, Dance, Design & Technology (incl. food), Drama, Economics, English, Geography, History, Mathematics, Modern Foreign Languages, Music, Physical Education, Physics, Psychology, Religious Education, Science, Social Studies.

Figure 6.4: Question 5, percentage of respondents holding each qualification
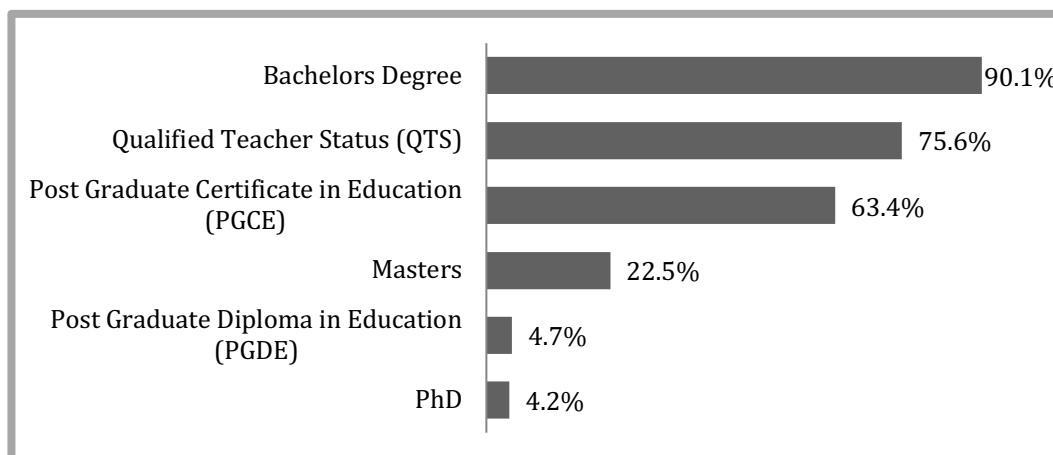
6.3 The Carter Review of ITT

In the second section of the survey, question 6 asked respondents how strongly they agreed with the conclusion of the Carter Review of Initial Teacher Training that:

> Of all areas of ITT content, we believe the most significant improvements are needed for training in assessment (Carter, 2015, p. 9).

Figure 6.5 below shows the overall agreement for all respondents answering this question. A further breakdown of responses is provided by the period in which respondents had completed their ITT. While the majority of respondents agreed with the conclusion (82.9%, out of 211 total respondents to this question), there were differences in the percentage strongly agreeing across the different ITT completion time periods: there was an 8.8 percentage point difference between those who had completed their ITT prior to 2000 and those completing in the 2000s, and a difference of 5.7 percentage points between the former and those completing from 2010 onwards.[46]

---

[46] The 211 total respondents to this question are made up of 70 pre-2000, 71 2000s, and 70 2010s.

Figure 6.5: Question 6, the extent of agreement with the Carter Review conclusion that significant improvements are needed for training in assessment
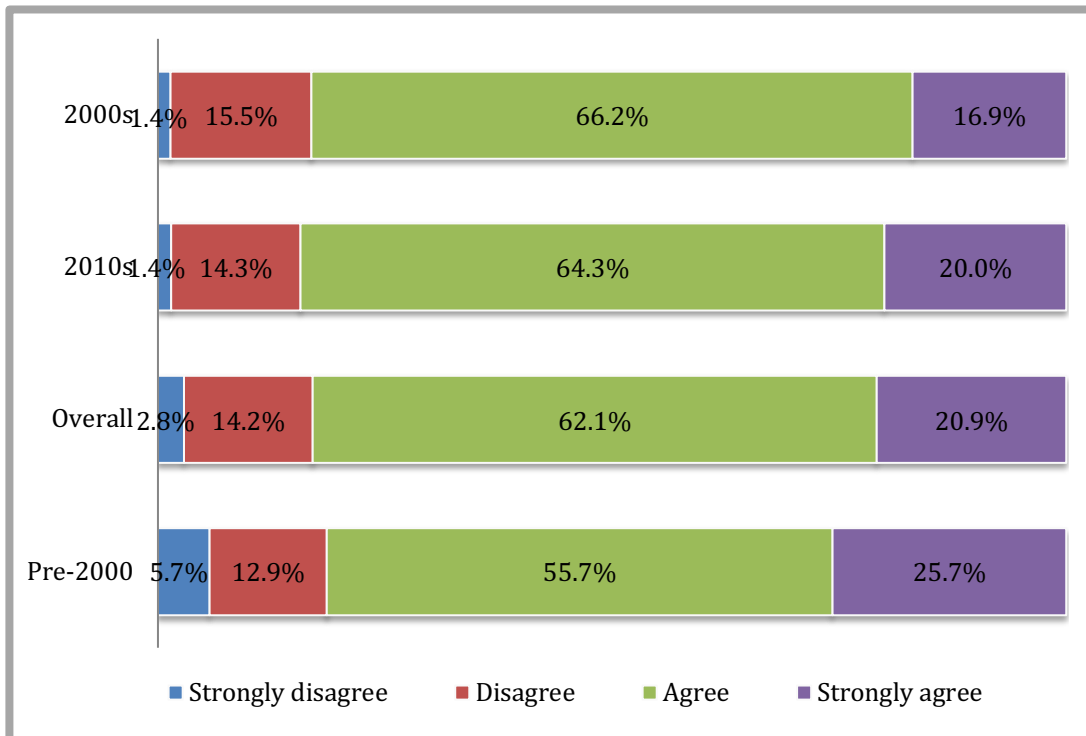
Question 7 asked respondents about the extent to which they agreed with Carter that there are significant gaps in the capacity of schools or ITT providers in the theoretical and technical aspects of assessment. Respondents largely agreed (80.8%, out of 211 total respondents to this question) that there are significant gaps in the capacity of schools in the theoretical and technical aspects of assessment, but here there were bigger differences across respondents completing their ITT in different time periods: see Figure 6.6. While 85.7% of the pre-2000 group either agreed or strongly agreed, this figure was lower for those completing from 2010 onwards at 68.6%, (17.1 percentage points lower). For those respondents who completed their ITT in the 2000s 60.6% either agreed or strongly agreed with Carter, 25.2 percentage points lower than those who trained in the previous decades.[47]

---

[47] Same number of respondents per training period as in previous footnote.

Figure 6.6: Question 7, the extent of agreement that there are significant gaps in the capacity of schools in theoretical and technical aspects of assessment
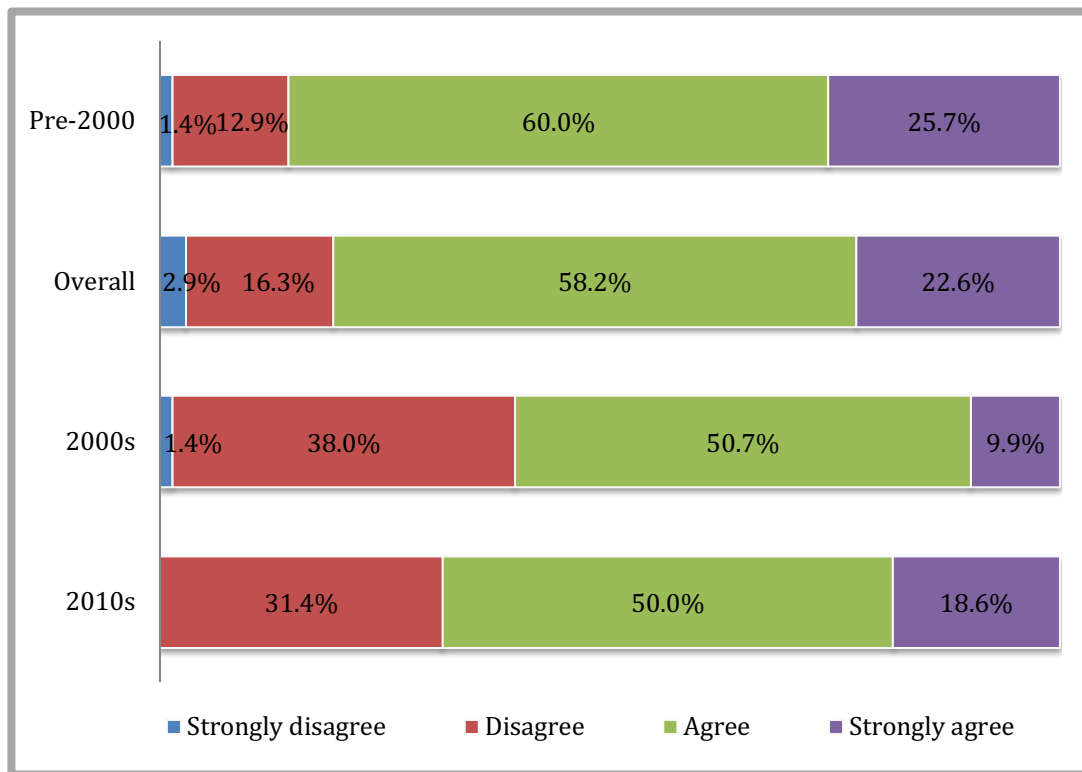
These differences were however smaller when respondents were asked about the capacity of ITT providers in the theoretical and technical aspects of assessments as shown in Figure 6.7. A smaller percentage agreed overall (71.6%, out of 208 total respondents to this question, compared to over 80% for both previous questions). Responses to this question showed a greater difference between the periods of training. The difference was greatest between those who had completed their ITT pre-2000 (84.3% selecting 'agree' or 'strongly agree') and those who completed their training from 2010 onwards at 74.6%, a difference of 9.7 percentage points.[48]

---

[48] The 208 total respondents to this question are made up of 70 pre-2000s, 71 2000s, and 67 2010s.

Figure 6.7: Question 7, the extent of agreement that there are significant gaps in the capacity of ITT providers in theoretical and technical aspects of assessment
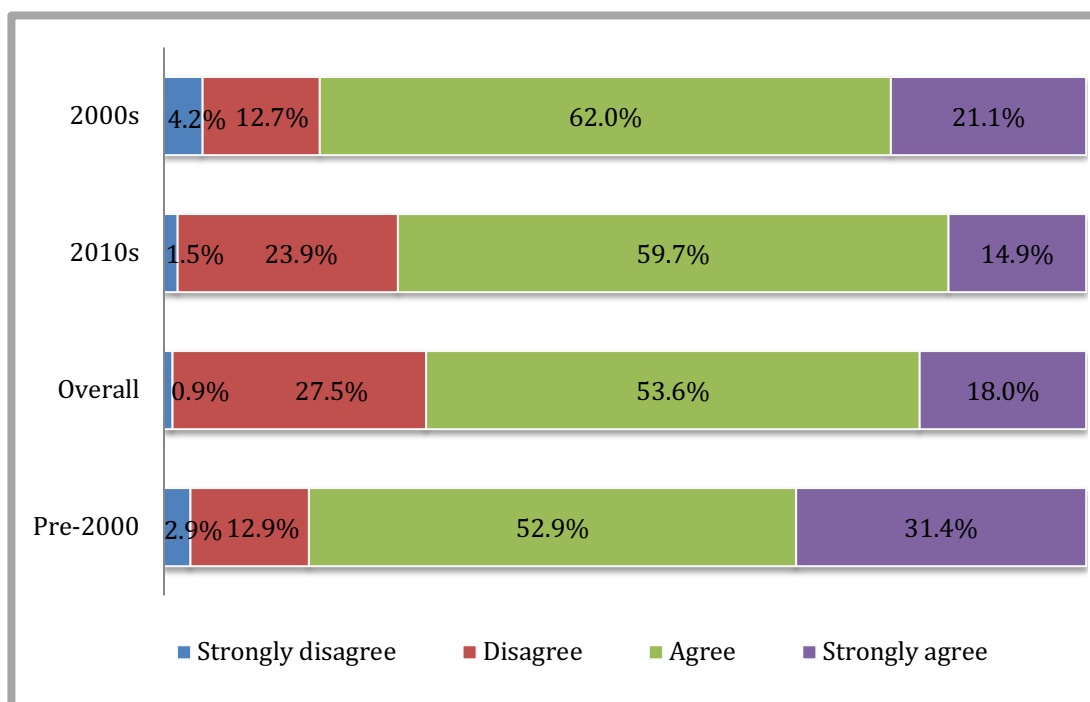
There were also differences in agreement with these statements between respondents who had been trained in primary versus secondary education.[49] A greater percentage of primary-trained respondents agreed (10.4 percentage points higher than secondary) with the Carter Review conclusion regarding training in assessment and regarding gaps in ITT provider capacity (8.7 percentage points higher than secondary), while a greater percentage of secondary-trained respondents agreed (7 percentage points) that there are gaps in the capacity of schools: see Figure 6.8.

---

[49] The 176 total respondents to the significant improvements and schools' capacity questions are made up of 110 primary and 66 secondary teachers; the 174 total respondents to the ITT providers' capacity question are made up of 109 primary and 65 secondary teachers.

Figure 6.8: Question 7, differences in agreement by phase of ITT

Category	Phase	Strongly disagree	Disagree	Agree	Strongly agree
Significant gaps in theoretical and technical aspects of assessment (ITT providers)	Secondary	1.5%	20.0%	52.3%	26.2%
	Primary	3.7%	9.2%	63.3%	23.9%
Significant gaps in theoretical and technical aspects of assessment (Schools)	Secondary		21.2%	48.5%	30.3%
	Primary	1.8%	26.4%	60.9%	10.9%
Significant improvements needed for training in assessments	Secondary	1.5%	19.7%	54.5%	24.2%
	Primary	4.5%	6.4%	68.2%	20.9%

■ Strongly disagree ■ Disagree ■ Agree ■ Strongly agree

6.4 Trust in teachers' assessment

In the third section of the survey, question 8 referred to the NAHT Commission on Assessment (2014) that stated that there is a lack of trust in assessments carried out by teachers. Respondents were asked if they, as teachers, think that their assessment judgments are trusted by different groups. As shown in Figure 6.9 below, 85.6% of respondents felt that their assessment judgments are not trusted by politicians with over 50% of respondents feeling that their judgments are not trusted by other external

organisations and regulators (i.e., STA, examination boards, Ofsted).[50]   44.6% felt untrusted by local authorities with 38.3% of respondents reporting that colleagues in other schools do not trust their assessments.

Figure 6.9: Question 8, differences in agreement by phase of ITT
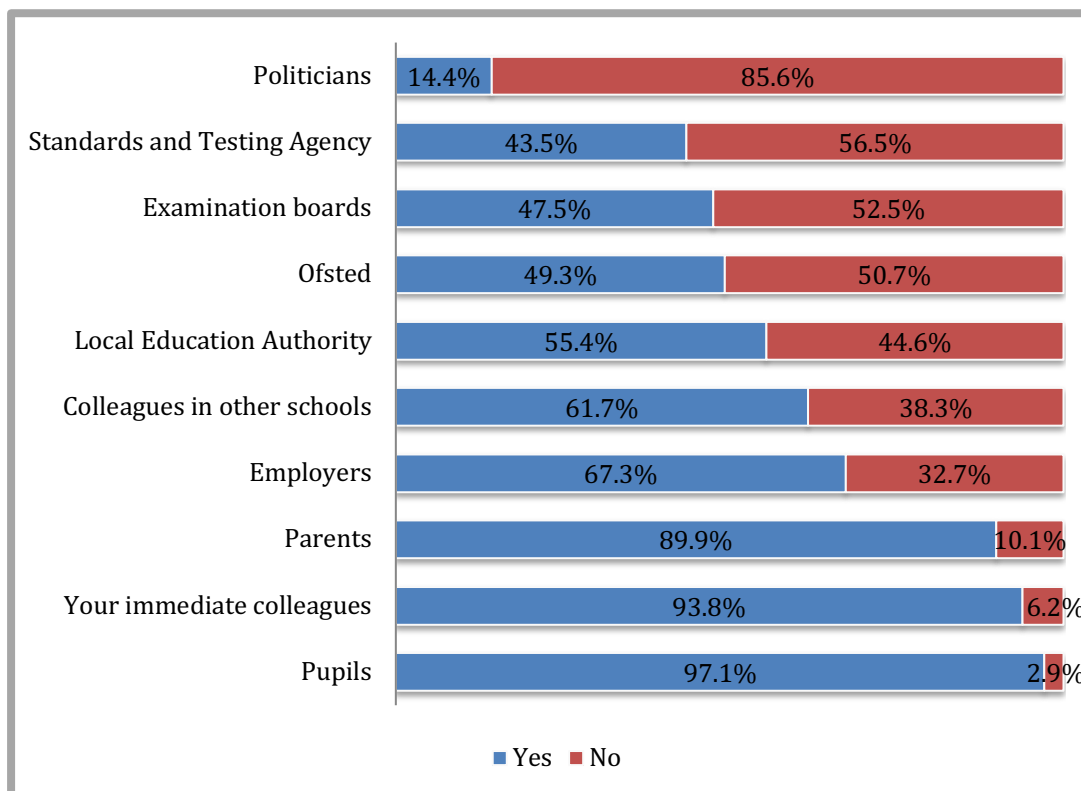
Figure 6.10 provides an analysis of responses to question 8 on the perceptions of the levels of trust shown by different groups when differentiated by the periods in which respondents completed their ITT.[51]   There are large differences (greater than ten percentage points) between respondents who completed their ITT prior to 2000 and those

---

[50] Respondents were asked about whether parents trust their assessment judgments twice, the analysis below illustrates responses to the first occurrence. Total respondents per statement were: 209 (Your immediate colleagues, Colleagues in other schools); 207 (Ofsted, Pupils, Parents); 202 (Local Education Authority, Politicians, Employers); 200 (Examination boards); 193 (Standards and Testing Agency).

[51] Respondents by ITT time period:
Your immediate colleagues (pre-2000: 68; 2000s: 71; 2010s: 70);
Colleagues in other schools (pre-2000: 68; 2000s: 71; 2010s: 70);
Ofsted (pre-2000: 66; 2000s: 71; 2010s: 70);
Local Education Authority (pre-2000: 65; 2000s: 69; 2010s: 68);
Examinations boards (pre-2000: 65; 2000s: 66; 2010s: 69);
Standards and Testing Agency (pre-2000: 59; 2000s: 67; 2010s: 67);
Pupils (pre-2000: 67; 2000s: 71; 2010s: 69);
Parents (pre-2000: 67; 2000s: 71; 2010s: 69);
Politicians (pre-2000: 65; 2000s: 69; 2010s: 68);
Employers (pre-2000: 65; 2000s: 68; 2010s: 69).

completing after 2010 in terms of the Standards and Testing Agency (15.2 percentage points), employers (12.7), Ofsted (12.0), and politicians (11.4). A greater percentage of more recently qualified respondents felt that parents do not trust their judgments (15.9 percentage points higher than respondents who qualified pre-2000).

Figure 6.10: Question 8, percentage of respondents answering 'no' to 'do you think that your assessment judgments are trusted by…[various groups]?' by when they completed their ITT
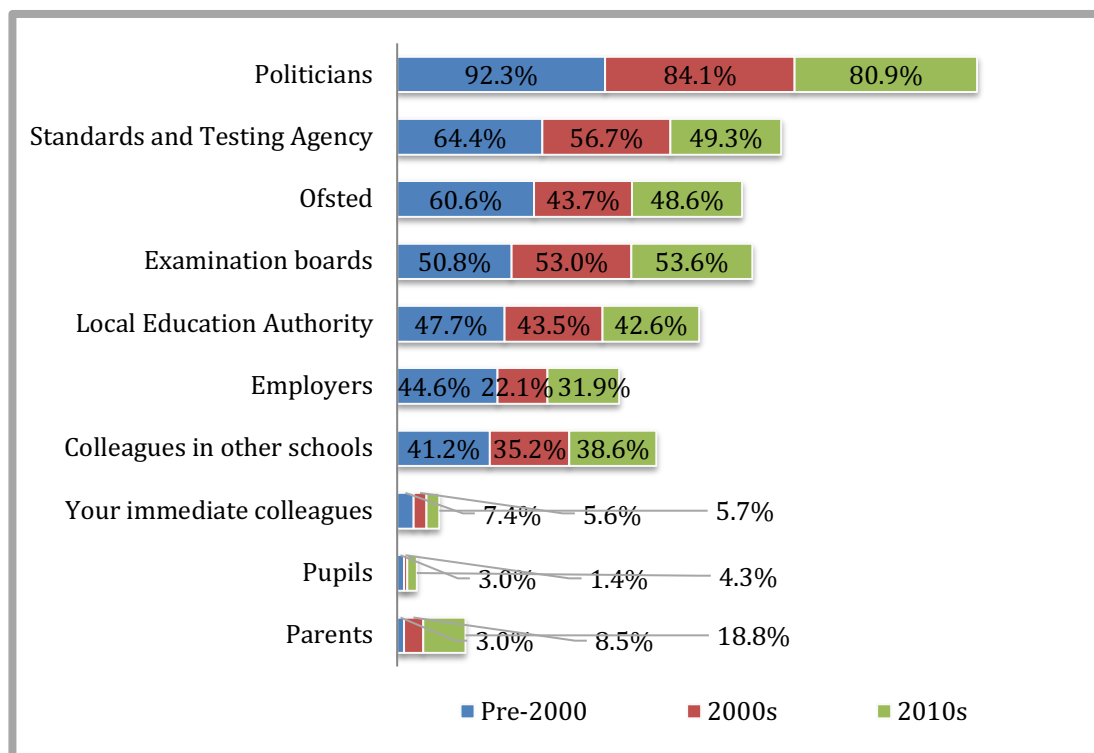
Figure 6.11 shows a further analysis of responses between respondents trained in primary and secondary phases of education.[52] A much larger percentage of primary-trained respondents felt their assessment judgments were not trusted by examination boards than secondary-trained respondents (a difference of 24.9 percentage points). A greater percentage of secondary-trained respondents felt their decisions were not trusted by employers than primary-trained respondents (12.1 percentage points).

---

52 Respondents by training phase:
Your immediate colleagues and Colleagues in other schools (primary: 110; secondary: 65);
Ofsted (primary: 109; secondary: 64);
Local Education Authority (primary: 109; secondary: 60);
Examinations boards (primary: 101; secondary: 65);
Standards and Testing Agency (primary: 107; secondary: 53);
Pupils and Parents (primary: 109; secondary: 65);
Politicians (primary: 108; secondary: 61);
Employers (primary: 104; secondary: 64).

Figure 6.11: Question 8, percentage of respondents answering 'no' to 'do you think that your assessment judgments are trusted by…[various groups]?' by phase of ITT
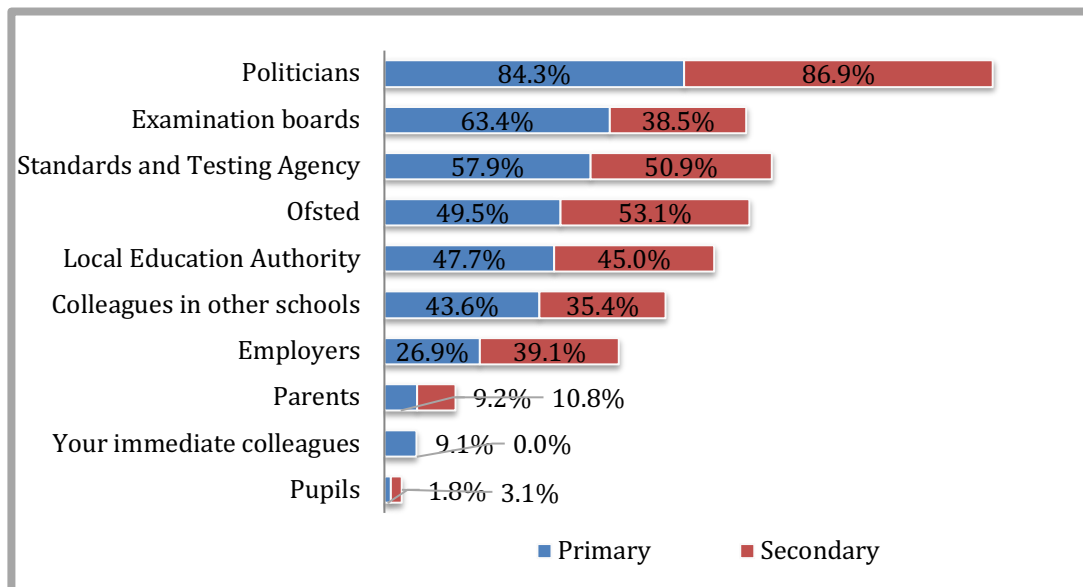
6.5 Educational assessment content of ITT courses

In the fourth section of the survey, questions 9 to 13 asked respondents to comment on whether specific aspects related to educational assessment were included in their ITT course, and the extent to which they felt confident their course fully prepared them for their first years in teaching. The sections below examine assessment aspects relating to:

- models of learning;
- understanding progression;
- principles of educational assessment;
- the uses of educational assessment outcomes;
- assessment design; and
- educational assessment and accountability.

Figures 6.12 to 6.25 present the data. Each figure contains finer details of the above categories on the y axis. For example Figure 6.12, models of learning, is subdivided into three sections. The first section, *Knowledge and understanding of how pupils learn and how this impacts on teaching,* shows where this aspect is included in respondents ITT courses broken down by overall responses then by year of ITT graduation and primary or secondary phase of training.

6.5.1 Models of learning

Question 9 of the survey focused on models of learning (see Figures 6.12 & 6.13). The 'Knowledge and understanding of how pupils learn and how this impacts on teaching' was included in 80% of all respondents' courses.[53] This figure was 90% for respondents who had completed their ITT after 2010 and also higher for those respondents trained in primary education rather than secondary education. 17% of respondents overall felt very confident they had been fully prepared on this aspect during their ITT compared to 21% of those trained in secondary education, and the lowest confidence (responses of 'not at all confident') were amongst respondents who had completed ITT pre-2000 (13%).

57% of overall respondents' ITT had included 'Stages of development within subjects'.[54] This figure was 60% for respondents who had completed their ITT after 2000 and also higher for respondents trained in primary education than secondary education. 16% of respondents overall felt very confident they had been fully prepared on this aspect during their ITT compared to 23% of those trained in secondary education, and the lowest confidence responses ('not at all confident') were amongst respondents who had completed their ITT pre-2000 (18%).

34% of overall respondents' ITT had included 'Approaches for strengthening pupil memory, such as short tests and making effective use of questioning'.[55] This figure was 56% for respondents who had completed their ITT after 2010 and also higher for respondents trained in primary education than secondary education. 12% of respondents overall felt very confident they had been fully prepared on this aspect during their ITT compared to 18% of those trained in secondary education, and the lowest confidence (responses of 'not at all confident') were amongst respondents who had completed ITT pre-2000 (37%).

---

[53] Out of a total of 196 respondents to this question. Confidence ratings out of a total of 155 respondents to this question. By time period (pre-2000: 66; 2000s: 67; 2010s: 63) and confidence ratings (pre-2000: 54; 2000s: 49; 2010s: 52). By phase of education (primary: 102; secondary: 60) and confidence ratings (primary: 80; secondary: 53).

[54] Out of a total of 196 respondents to this question. Confidence ratings out of a total of 146 respondents to this question. By time period (pre-2000: 66; 2000s: 68; 2010s: 62) and confidence ratings (pre-2000: 51; 2000s: 45; 2010s: 50). By phase of education (primary: 101; secondary: 61) and confidence ratings (primary: 79; secondary: 47).

[55] Out of a total of 197 respondents to this question. Confidence ratings out of a total of 139 respondents to this question. By time period (pre-2000: 65; 2000s: 69; 2010s: 63) and confidence ratings (pre-2000: 49; 2000s: 41; 2010s: 49). By phase of education (primary: 102; secondary: 61) and confidence ratings (primary: 74; secondary: 44).

Figure 6.12: Question 9, inclusion of assessment aspects relating to models of learning on respondents' ITT courses
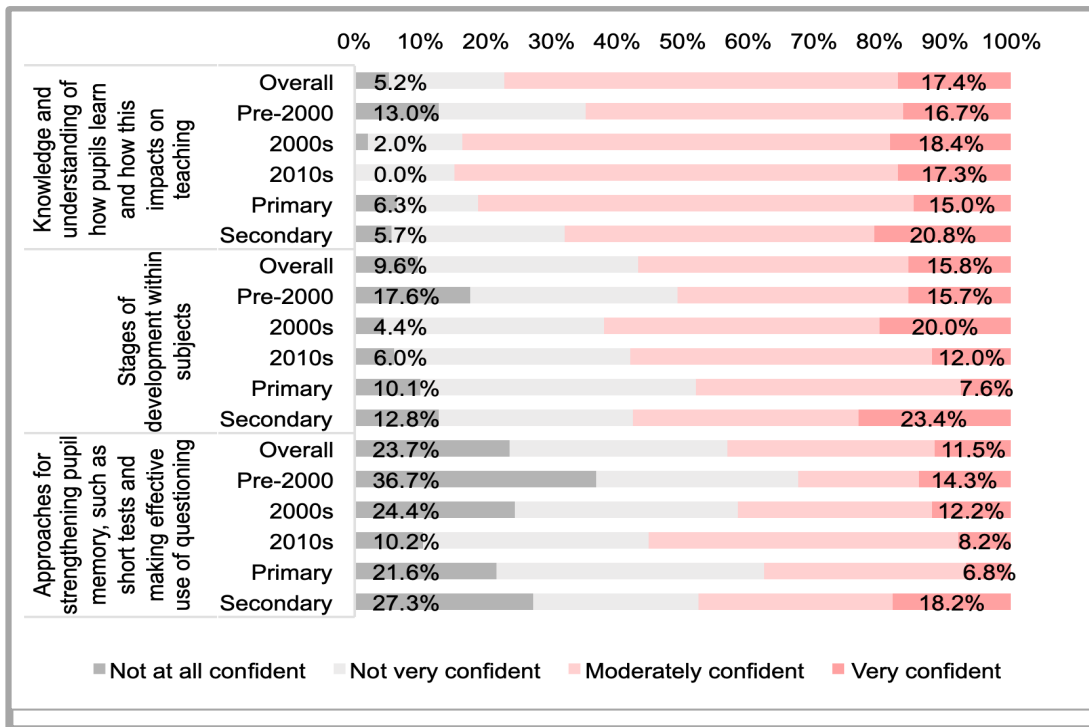
Figure 6.13: Question 9, respondents' confidence that their ITT had fully prepared them regarding models of learning
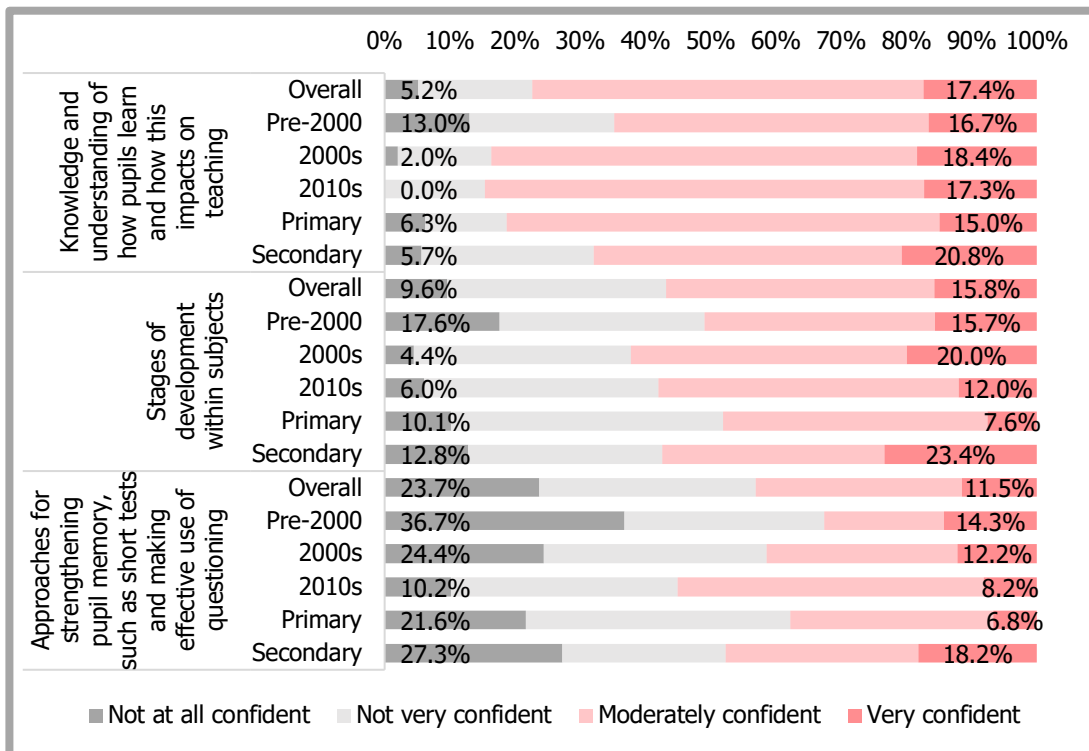
6.5.2 Understanding progression

Question 10 of the survey focused on understanding progression (see Figures 6.14 & 6.15). Over two thirds of respondents reported that their ITT had included 'the use of assessment to give pupils regular feedback, both orally and through accurate marking, and encourage pupils to respond to the feedback' and 'the use of formative assessment and summative assessment to secure pupils' progress'. These were also areas that respondents allocated the highest 'very confident' ratings. Inclusion in ITT provision for these two areas was also highest for respondents who had completed their ITT after 2010. Secondary phase-trained respondents also had a high 'very confident' rating regarding 'the use of relevant data to monitor progress' (23.8%).

Figure 6.14: Question 10, inclusion of assessment aspects relating to understanding progression in respondents' ITT courses

Knowledge and understanding of how pupils are expected to make progress within different subjects across each relevant Key Stages		
Overall	52.2%	47.8%
Pre-2000	67.2%	32.8%
2000s	47.8%	52.2%
2010s	41.5%	58.5%
Primary	46.7%	53.3%
Secondary	59.7%	40.3%

The use of relevant data to monitor progress		
Overall	59.6%	40.4%
Pre-2000	76.5%	23.5%
2000s	64.3%	35.7%
2010s	36.9%	63.1%
Primary	66.0%	34.0%
Secondary	61.9%	38.1%

The use of relevant data to set targets		
Overall	63.2%	36.8%
Pre-2000	83.8%	16.2%
2000s	66.2%	33.8%
2010s	38.5%	61.5%
Primary	66.7%	33.3%
Secondary	69.8%	30.2%

The use of relevant data to plan subsequent lessons		
Overall	50.5%	49.5%
Pre-2000	69.1%	30.9%
2000s	53.6%	46.4%
2010s	27.7%	72.3%
Primary	51.9%	48.1%
Secondary	57.1%	42.9%

The use of assessment to give pupils regular feedback, both orally and through accurate marking, and encourage pupils to respond to the feedback		
Overall	31.3%	68.7%
Pre-2000	44.1%	55.9%
2000s	33.8%	66.2%
2010s	15.4%	84.6%
Primary	31.4%	68.6%
Secondary	36.5%	63.5%

The use of formative and summative assessment to secure pupils' progress		
Overall	25.0%	75.0%
Pre-2000	42.6%	57.4%
2000s	22.4%	77.6%
2010s	9.2%	90.8%
Primary	25.7%	74.3%
Secondary	29.0%	71.0%

■ Aspects not included in my ITT course ■ Aspects included in my ITT course

Figure 6.15: Question 10, respondents' confidence that their ITT had fully prepared them regarding understanding progression

	Not at all confident	Not very confident	Moderately confident	Very confident
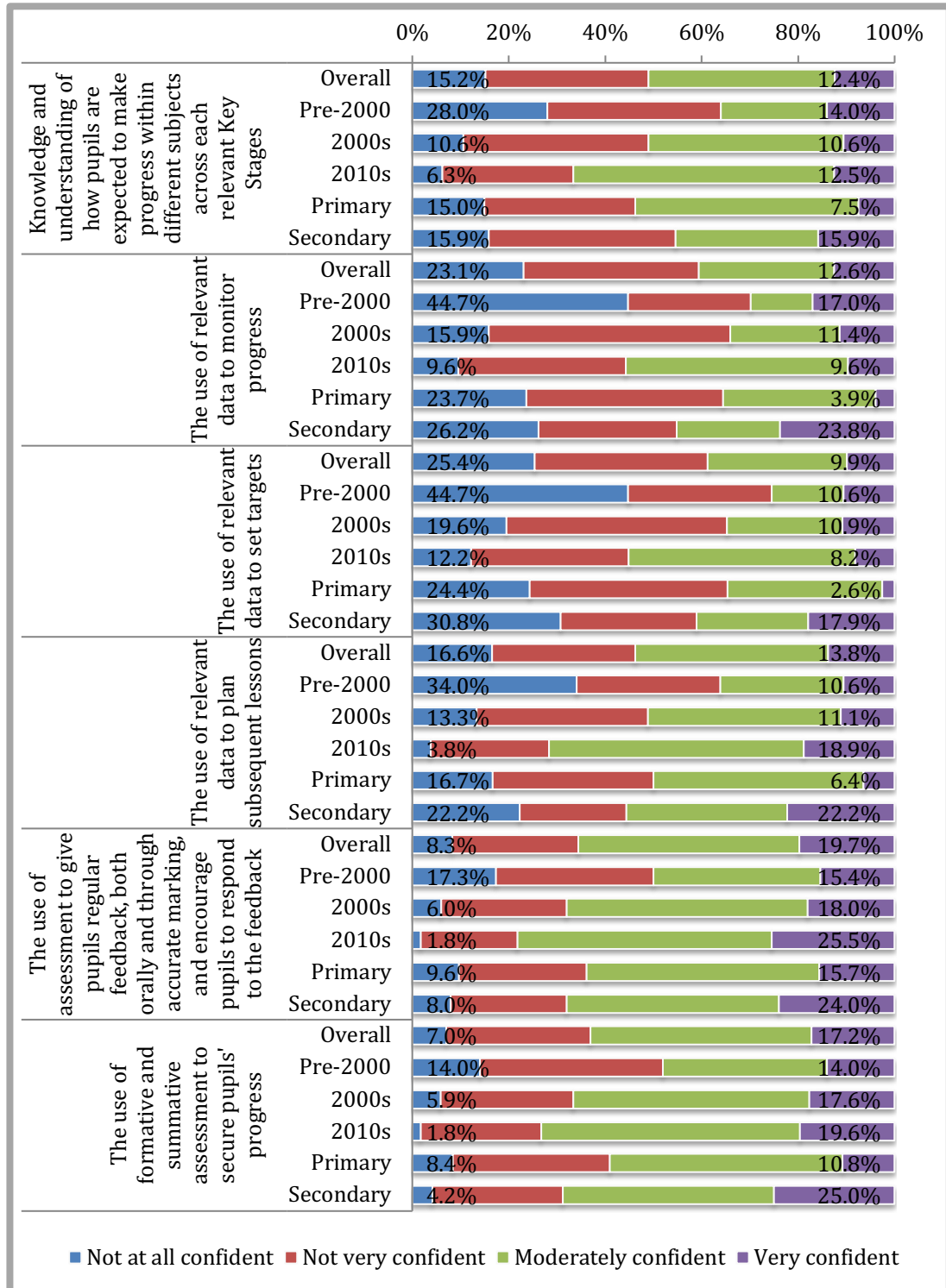
6.5.3 Principles of educational assessment

Question 11 of the survey focused on the principles of educational assessment (see Figures 6.16 to 6.19). 50.7% of respondents' ITT courses had included some 'theory of assessment', while 73.9% had included 'the fundamental principles of assessment' (Figure 6.16). These were areas that respondents allocated the highest 'very confident' ratings, 13.5% and 28.2%, respectively (Figure 6.18)). 40% of respondents completing ITT from 2010 onwards reported this included 'reliability', 'setting standards', and 'standardisation'; 40% of those completing ITT in the 2000s and those trained in secondary education reported this included 'moderation' (Figure 6.16 & 6.17).

In terms of key concepts used in educational assessment, 82.2% of respondents overall reported that bias was not included in their ITT, with 75.8% for validity and 73.2% for reliability. For those completing their ITT in the 2010s, the figures were 70.8%, 66.2% and 60.0% respectively (Figure 6.16). These figures were broadly reflected in the levels of confidence shown overall by respondents that their ITT had fully prepared them for teaching in these aspects of educational assessment. 39.2% reported being 'not at all confident' in respect of bias with a further 35% feeling 'not very confident' in this aspect. For validity, the figures were 40.2% and 28.7%; and for reliability 39% and 28.5% respectively. For those respondents completing their ITT in the 2010s, 17.5% stated that they were 'not at all confident' with regards to bias with 52.5% stating that they were 'not very confident'. The responses for validity were 22.5% and 45%: and for reliability, 20% and 42.5% respectively (Figure 6.18 & 6.19).

Figure 6.16: Question 11, inclusion of assessment aspects relating to principles of educational assessment on ITT courses (1)

		Aspects not included in my ITT course	Aspects included in my ITT course
Theory of assessment – why, when, how to assess and report	Overall	49.3%	50.7%
	Pre-2000	58.8%	41.2%
	2000s	57.7%	42.3%
	2010s	30.3%	69.7%
	Primary	45.8%	54.2%
	Secondary	56.3%	43.8%
The fundamental principles of assessment, including: the differences between formative and summative uses of assessment	Overall	26.1%	73.9%
	Pre-2000	41.5%	58.5%
	2000s	27.9%	72.1%
	2010s	9.1%	90.9%
	Primary	18.1%	81.9%
	Secondary	36.1%	63.9%
bias	Overall	82.2%	17.8%
	Pre-2000	88.9%	11.1%
	2000s	87.0%	13.0%
	2010s	70.8%	29.2%
	Primary	80.8%	19.2%
	Secondary	83.6%	16.4%
validity	Overall	75.8%	24.2%
	Pre-2000	81.3%	18.8%
	2000s	79.7%	20.3%
	2010s	66.2%	33.8%
	Primary	72.4%	27.6%
	Secondary	80.3%	19.7%
reliability	Overall	73.2%	26.8%
	Pre-2000	82.8%	17.2%
	2000s	76.8%	23.2%
	2010s	60.0%	40.0%
	Primary	71.4%	28.6%
	Secondary	75.4%	24.6%
utility	Overall	83.7%	16.3%
	Pre-2000	90.5%	9.5%
	2000s	85.5%	14.5%
	2010s	75.0%	25.0%
	Primary	83.7%	16.3%
	Secondary	85.0%	15.0%

■ Aspects not included in my ITT course   ■ Aspects included in my ITT course

Figure 6.17: Question 11, inclusion of assessment aspects relating to principles of educational assessment on respondents' ITT courses (2)

Category	Group	Aspects not included in my ITT course	Aspects included in my ITT course
criterion- and norm-referencing	Overall	80.4%	19.6%
	Pre-2000	76.9%	23.1%
	2000s	84.1%	15.9%
	2010s	80.0%	20.0%
	Primary	81.7%	18.3%
	Secondary	79.0%	21.0%
standardisation –setting standards of educational performance	Overall	73.7%	26.3%
	Pre-2000	85.9%	14.1%
	2000s	75.4%	24.6%
	2010s	60.0%	40.0%
	Primary	70.2%	29.8%
	Secondary	82.0%	18.0%
standardisation – purpose and process	Overall	70.4%	29.6%
	Pre-2000	76.9%	23.1%
	2000s	73.9%	26.1%
	2010s	60.0%	40.0%
	Primary	75.0%	25.0%
	Secondary	69.4%	30.6%
moderation - purpose and process	Overall	64.3%	35.7%
	Pre-2000	71.9%	28.1%
	2000s	59.4%	40.6%
	2010s	62.1%	37.9%
	Primary	68.3%	31.7%
	Secondary	59.7%	40.3%
the design of standardised tests (such as those that produce a reading age)	Overall	82.2%	17.8%
	Pre-2000	82.5%	17.5%
	2000s	88.2%	11.8%
	2010s	75.8%	24.2%
	Primary	77.9%	22.1%
	Secondary	85.2%	14.8%
How nationally standardised summative assessment can be used to help teachers to understand national expectations and assess their own performance	Overall	74.6%	25.4%
	Pre-2000	90.8%	9.2%
	2000s	72.9%	27.1%
	2010s	60.6%	39.4%
	Primary	72.9%	27.1%
	Secondary	77.4%	22.6%
Using externally set standards of educational performance for example Interim Teachers Assessment Frameworks, exemplification of standards from STA or awarding bodies	Overall	75.2%	24.8%
	Pre-2000	86.2%	13.8%
	2000s	73.2%	26.8%
	2010s	66.7%	33.3%
	Primary	80.2%	19.8%
	Secondary	79.4%	20.6%

■ Aspects not included in my ITT course   ■ Aspects included in my ITT course

Figure 6.18: Question 11, respondents' confidence that their ITT had fully prepared them regarding principles of educational assessment (1)

		Not at all confident	Not very confident	Moderately confident	Very confident
Theory of assessment – why, when, how to assess and report	Overall	21.3%		13.5%	
	Pre-2000	36.2%		12.8%	
	2000s	18.6%		11.6%	
	2010s	9.8%		15.7%	
	Primary	20.8%		10.4%	
	Secondary	24.4%		19.5%	
The fundamental principles of assessment, including: the differences between formative and summative uses of assessment	Overall	11.4%		28.2%	
	Pre-2000	22.4%		18.4%	
	2000s	8.2%		26.5%	
	2010s	3.9%		39.2%	
	Primary	10.5%		19.7%	
	Secondary	10.6%		38.3%	
bias	Overall	39.2%		8.3%	
	Pre-2000	50.0%		14.3%	
	2000s	50.0%		10.5%	
	2010s	17.5%		0.0%	
	Primary	38.5%		1.5%	
	Secondary	41.7%		11.1%	
validity	Overall	40.2%		9.0%	
	Pre-2000	48.9%		15.6%	
	2000s	48.6%		10.8%	
	2010s	22.5%		0.0%	
	Primary	37.3%		1.5%	
	Secondary	44.4%		13.9%	
reliability	Overall	39.0%		10.6%	
	Pre-2000	46.7%		15.6%	
	2000s	50.0%		13.2%	
	2010s	20.0%		2.5%	
	Primary	37.3%		1.5%	
	Secondary	40.5%		16.2%	
utility	Overall	40.8%		7.5%	
	Pre-2000	48.8%		11.6%	
	2000s	47.4%		7.9%	
	2010s	25.6%		2.6%	
	Primary	36.4%		1.5%	
	Secondary	48.6%		8.6%	

Figure 6.19: Question 11, respondents' confidence that their ITT had fully prepared them regarding principles of educational assessment (2)

		Not at all confident	Not very confident	Moderately confident	Very confident
criterion- and norm-referencing	Overall	42.5%		10.8%	
	Pre-2000	41.9%		16.3%	
	2000s	50.0%		13.2%	
	2010s	35.9%		2.6%	
	Primary	44.6%		3.1%	
	Secondary	40.5%		16.2%	
standardisation – setting standards of educational performance	Overall	37.8%		7.6%	
	Pre-2000	45.2%		14.3%	
	2000s	44.7%		7.9%	
	2010s	23.1%		0.0%	
	Primary	33.3%		1.5%	
	Secondary	44.1%		8.8%	
standardisation – purpose and process	Overall	30.0%		10.0%	
	Pre-2000	40.0%		17.8%	
	2000s	35.7%		4.8%	
	2010s	14.0%		7.0%	
	Primary	29.0%		1.4%	
	Secondary	35.9%		15.4%	
moderation - purpose and process	Overall	24.6%		12.3%	
	Pre-2000	31.9%		19.1%	
	2000s	27.5%		5.0%	
	2010s	14.0%		11.6%	
	Primary	24.6%		4.3%	
	Secondary	24.4%		17.1%	
the design of standardised tests (such as those that produce a reading age)	Overall	40.8%		6.4%	
	Pre-2000	54.3%		8.7%	
	2000s	44.7%		7.9%	
	2010s	22.0%		2.4%	
	Primary	32.8%		3.0%	
	Secondary	53.8%		2.6%	
How nationally standardised summative assessment can be used to help teachers to understand national expectations and assess their own performance	Overall	32.8%		10.2%	
	Pre-2000	51.2%		14.0%	
	2000s	36.6%		7.3%	
	2010s	11.4%		9.1%	
	Primary	31.4%		4.3%	
	Secondary	40.5%		10.8%	
Using externally set standards of educational performance for example Interim Teachers Assessment Frameworks, exemplification of standards from STA or awarding bodies	Overall	36.8%		8.8%	
	Pre-2000	47.7%		11.4%	
	2000s	47.4%		10.5%	
	2010s	16.3%		4.7%	
	Primary	36.2%		1.4%	
	Secondary	47.2%		11.1%	

6.5.4 The uses of educational assessment outcomes

Question 12 of the survey focused on the uses of educational assessment outcomes (see Figures 6.20 & 6.21). The ITT provision of over half of all respondents had included the aspects included in Figure 20 below relating to the uses of educational assessment outcomes, with the exception of 'common inferences/uses of assessment' (39.2%) and 'the use and misuse of assessment outcomes and the ethics of assessment' (12.7%). Figure 6. 21 shows that respondents were most confident regarding 'the effective use of assessment to give effective and efficient oral and written feedback to pupils and parents' (23.4% felt 'very confident', compared to 31.3% of respondents trained from 2010 onwards) and 'using subject knowledge, progression, assessment and skilful questioning to inform planning' (21.7%, compared to 26.0% for secondary-trained respondents).

Figure 6.20: Question 12, inclusion of assessment aspects relating to the uses of educational assessment outcomes on respondents' ITT courses

		Aspects not included in my ITT course	Aspects included in my ITT course
The use and misuse of assessment outcomes and the ethics of assessment	Overall	87.3%	12.7%
	Pre-2000	91.2%	8.8%
	2000s	87.1%	12.9%
	2010s	83.3%	16.7%
	Primary	89.7%	10.3%
	Secondary	90.5%	9.5%
Common inferences/uses of assessment	Overall	60.8%	39.2%
	Pre-2000	70.6%	29.4%
	2000s	61.4%	38.6%
	2010s	50.0%	50.0%
	Primary	65.4%	34.6%
	Secondary	65.1%	34.9%
Reporting and using assessment outcomes	Overall	42.6%	57.4%
	Pre-2000	56.7%	43.3%
	2000s	31.4%	68.6%
	2010s	40.0%	60.0%
	Primary	44.3%	55.7%
	Secondary	50.0%	50.0%
Using subject knowledge, progression, assessment and skilful questioning to inform planning	Overall	19.7%	80.3%
	Pre-2000	28.4%	71.6%
	2000s	21.4%	78.6%
	2010s	9.1%	90.9%
	Primary	18.9%	81.1%
	Secondary	28.6%	71.4%
The effective use of assessment data to ensure that teaching is both supportive and challenging	Overall	49.0%	51.0%
	Pre-2000	73.5%	26.5%
	2000s	47.1%	52.9%
	2010s	25.8%	74.2%
	Primary	48.1%	51.9%
	Secondary	50.0%	50.0%
The effective use of assessment to give effective and efficient oral and written feedback to pupils and parents	Overall	41.4%	58.6%
	Pre-2000	59.7%	40.3%
	2000s	34.3%	65.7%
	2010s	30.3%	69.7%
	Primary	39.3%	60.7%
	Secondary	44.4%	55.6%
Using knowledge of how pupils make progress and assessment outcomes for teachers to self-evaluate	Overall	46.6%	53.4%
	Pre-2000	69.1%	30.9%
	2000s	41.4%	58.6%
	2010s	28.8%	71.2%
	Primary	43.9%	56.1%
	Secondary	52.4%	47.6%

■ Aspects not included in my ITT course   ■ Aspects included in my ITT course

Figure 6.21: Question 12, respondents' confidence that their ITT had fully prepared them regarding the uses of educational assessment outcomes

	Not at all confident	Not very confident	Moderately confident	Very confident
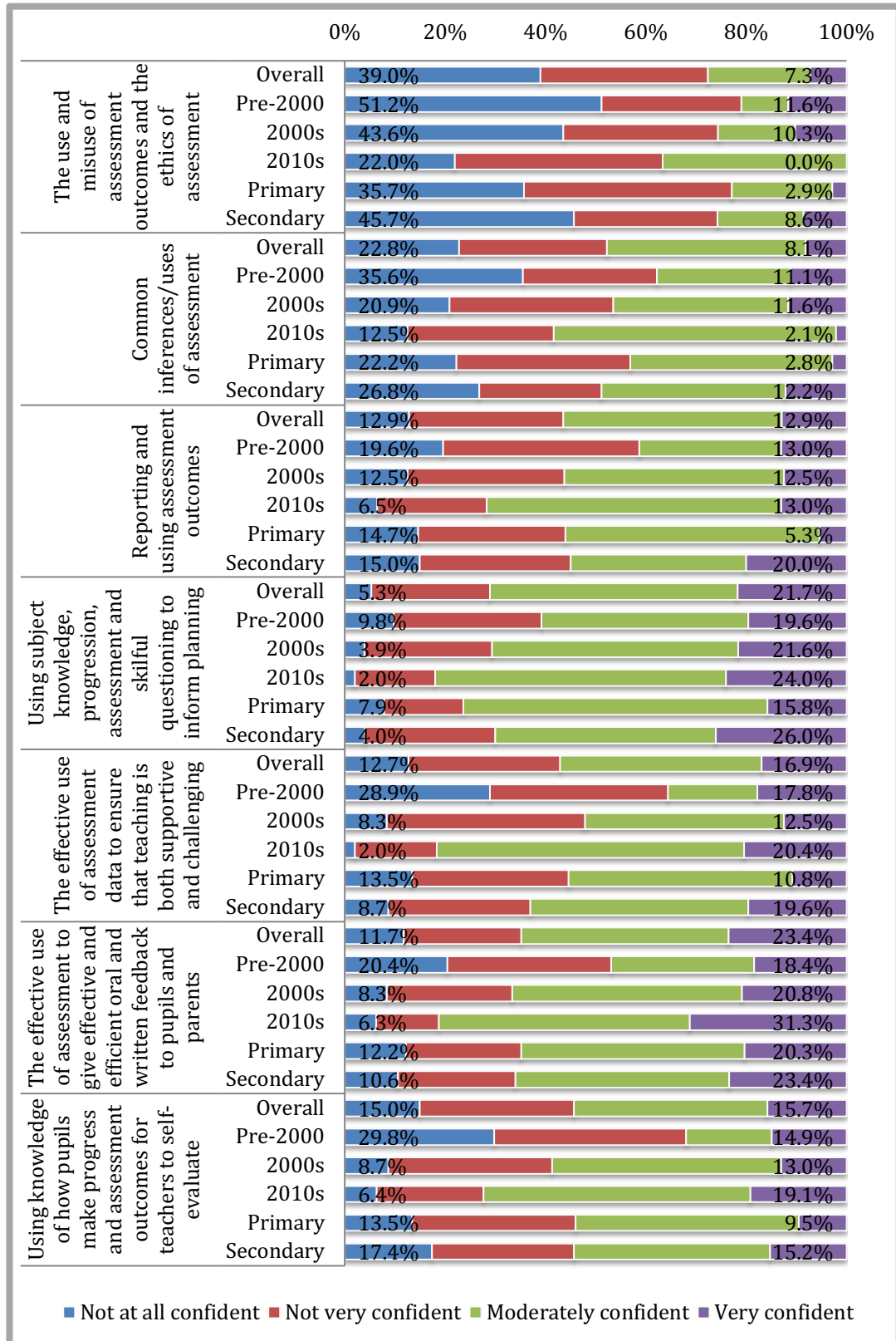
6.5.5 Assessment design

Question 13 of the survey focused on assessment design. Only 21.2% of respondents overall experienced ITT that included 'how to identify key constructs and valid approaches to assessment' compared to 61.2% whose ITT had included 'setting effective questions' (Figure 6.22), the area respondents felt most confident in with 20.7% 'very confident' (Figure 6.23). All aspects had been included in ITT for a greater percentage of those trained from 2010 onwards, with the exception of 'designing appropriate assessments' with 28.8% compared to 30.0% of those trained in the 2000s, and 30.9% pre-2000 (Figure 6.22). Those trained in secondary education had higher 'very confident' ratings than those trained in primary for all assessment design aspects (Figure 6.23).

Figure 6.22: Question 13, inclusion of assessment aspects relating to assessment design on respondents' ITT courses

	Aspects not included in my ITT course	Aspects included in my ITT course
The relationship between what is taught and what is assessed		
Overall	54.2%	45.8%
Pre-2000	64.2%	35.8%
2000s	52.9%	47.1%
2010s	45.5%	54.5%
Primary	50.5%	49.5%
Secondary	61.3%	38.7%
How to identify key constructs and valid approaches to assessment		
Overall	78.8%	21.2%
Pre-2000	82.1%	17.9%
2000s	81.4%	18.6%
2010s	72.7%	27.3%
Primary	80.4%	19.6%
Secondary	82.3%	17.7%
Knowledge and understanding of how to assess the relevant subject and curriculum areas, including statutory assessment requirements		
Overall	41.4%	58.6%
Pre-2000	58.8%	41.2%
2000s	35.7%	64.3%
2010s	29.2%	70.8%
Primary	45.8%	54.2%
Secondary	43.5%	56.5%
Designing appropriate assessments		
Overall	70.1%	29.9%
Pre-2000	69.1%	30.9%
2000s	70.0%	30.0%
2010s	71.2%	28.8%
Primary	77.6%	22.4%
Secondary	65.1%	34.9%
Setting effective questions		
Overall	38.8%	61.2%
Pre-2000	43.3%	56.7%
2000s	43.5%	56.5%
2010s	29.2%	70.8%
Primary	38.1%	61.9%
Secondary	40.3%	59.7%
The use of continuous assessment of pupils' work and how approaches to assessment differ across subjects		
Overall	49.5%	50.5%
Pre-2000	58.8%	41.2%
2000s	51.4%	48.6%
2010s	37.9%	62.1%
Primary	44.9%	55.1%
Secondary	58.7%	41.3%
Assessment systems – purpose, design, implementation and uses including: school based; local/regionally derived systems; national or large scale based systems		
Overall	70.1%	29.9%
Pre-2000	75.0%	25.0%
2000s	68.6%	31.4%
2010s	66.7%	33.3%
Primary	72.9%	27.1%
Secondary	71.4%	28.6%

■ Aspects not included in my ITT course ■ Aspects included in my ITT course

Figure 6.23: Question 13, respondents' *confidence* that their ITT had fully prepared them regarding assessment design

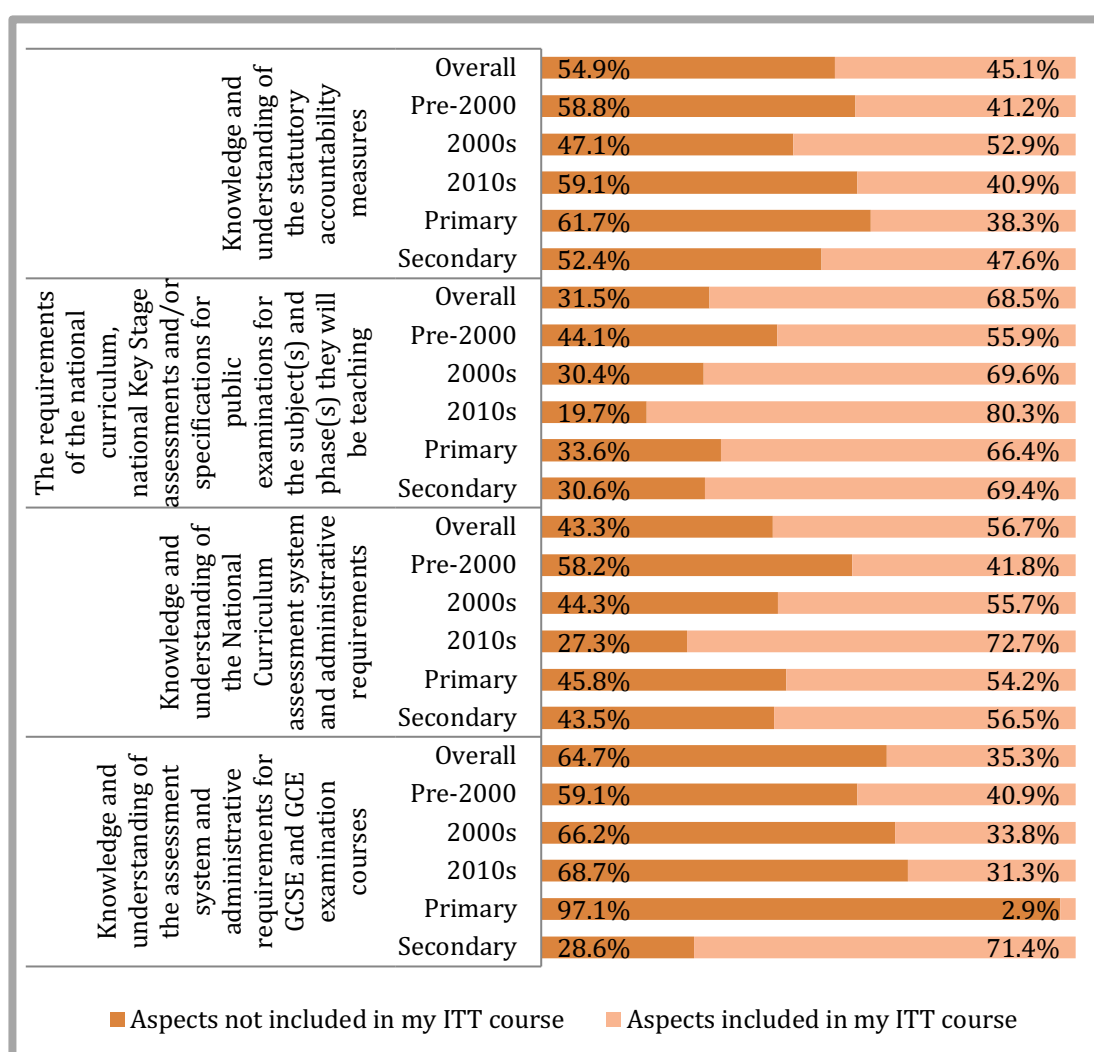|		Not at all confident	Not very confident	Moderately confident	Very confident
The relationship between what is taught and what is assessed	Overall	14.3%		12.0%	
	Pre-2000	16.7%		14.3%	
	2000s	17.8%		11.1%	
	2010s	8.7%		10.9%	
	Primary	16.4%		6.8%	
	Secondary	17.9%		15.4%	
How to identify key constructs and valid approaches to assessment	Overall	27.0%		7.9%	
	Pre-2000	36.6%		12.2%	
	2000s	31.7%		9.8%	
	2010s	13.6%		2.3%	
	Primary	28.6%		1.4%	
	Secondary	35.1%		10.8%	
Knowledge and understanding of how to assess the relevant subject and curriculum areas, including statutory assessment requirements	Overall	12.8%		15.6%	
	Pre-2000	19.6%		10.9%	
	2000s	13.0%		23.9%	
	2010s	6.1%		12.2%	
	Primary	14.9%		8.1%	
	Secondary	11.4%		20.5%	
Designing appropriate assessments	Overall	25.2%		11.8%	
	Pre-2000	21.4%		11.9%	
	2000s	32.6%		14.0%	
	2010s	21.4%		9.5%	
	Primary	31.8%		3.0%	
	Secondary	20.0%		17.5%	
Setting effective questions	Overall	12.1%		20.7%	
	Pre-2000	15.2%		19.6%	
	2000s	13.0%		17.4%	
	2010s	8.3%		25.0%	
	Primary	13.2%		14.5%	
	Secondary	11.4%		22.7%	
The use of continuous assessment of pupils' work and how approaches to assessment differ across subjects	Overall	17.0%		14.8%	
	Pre-2000	26.1%		17.4%	
	2000s	16.3%		16.3%	
	2010s	8.7%		10.9%	
	Primary	13.5%		9.5%	
	Secondary	26.2%		16.7%	
Assessment systems – purpose, design, implementation and uses including: school based; local/regionally derived systems; national or large scale based systems	Overall	24.0%		7.8%	
	Pre-2000	34.1%		14.6%	
	2000s	27.9%		7.0%	
	2010s	11.1%		2.2%	
	Primary	22.9%		1.4%	
	Secondary	25.6%		7.7%	

■ Not at all confident ■ Not very confident ■ Moderately confident ■ Very confident

6.5.6 Educational assessment and accountability

Question 14 of the survey focused on educational assessment and accountability. Given the different phases of education respondents had been trained in, some aspects of educational assessment and accountability applied less e.g., GCSE and GCE examination courses for primary phase respondents (Figure 6.24) and this impacted on reported confidence that their ITT had fully prepared them in these aspects (Figure 6.25).

Figure 6.24: Question 14, inclusion of assessment aspects relating to educational assessment and accountability on respondents' ITT courses

		Aspects not included in my ITT course	Aspects included in my ITT course
Knowledge and understanding of the statutory accountability measures	Overall	54.9%	45.1%
	Pre-2000	58.8%	41.2%
	2000s	47.1%	52.9%
	2010s	59.1%	40.9%
	Primary	61.7%	38.3%
	Secondary	52.4%	47.6%
The requirements of the national curriculum, national Key Stage assessments and/or specifications for public examinations for the subject(s) and phase(s) they will be teaching	Overall	31.5%	68.5%
	Pre-2000	44.1%	55.9%
	2000s	30.4%	69.6%
	2010s	19.7%	80.3%
	Primary	33.6%	66.4%
	Secondary	30.6%	69.4%
Knowledge and understanding of the National Curriculum assessment system and administrative requirements	Overall	43.3%	56.7%
	Pre-2000	58.2%	41.8%
	2000s	44.3%	55.7%
	2010s	27.3%	72.7%
	Primary	45.8%	54.2%
	Secondary	43.5%	56.5%
Knowledge and understanding of the assessment system and administrative requirements for GCSE and GCE examination courses	Overall	64.7%	35.3%
	Pre-2000	59.1%	40.9%
	2000s	66.2%	33.8%
	2010s	68.7%	31.3%
	Primary	97.1%	2.9%
	Secondary	28.6%	71.4%

Figure 6.25: Question 14, respondents' confidence that their ITT had fully prepared them regarding educational assessment and accountability
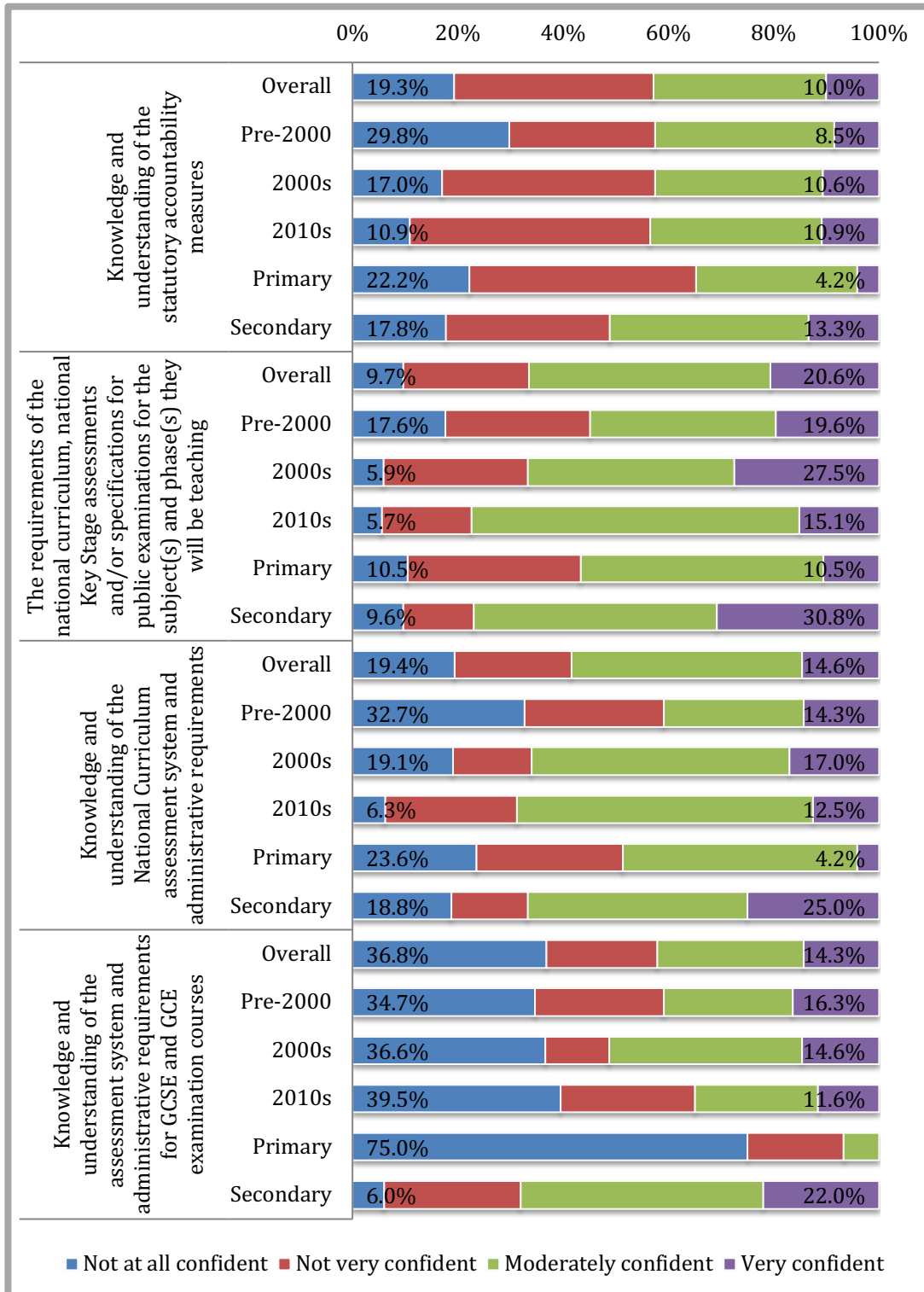
6.6 Reflections on ITT courses

Question 15 of the survey asked respondents if they considered any of the assessment aspects identified earlier in the survey as unnecessary for a course in initial teacher training. Respondents were given free response boxes to name the aspects and further boxes to explain why.

Several responses in this section focused on the short duration of the ITT course they had taken, and the challenge of fitting in aspects of assessment that are: very technical; for key stages they will not directly be involved in assessing; or likely to change often (e.g., administrative requirements, national curriculum assessments, national monitoring) or become out-dated quickly.

Some respondents felt that more in depth coverage could wait until a few years after they had become classroom teachers. Others felt that all the aspects mentioned in the survey so far should be included, even if they agreed that course duration is a constraint. Less common responses included the ethics of assessment, the design of standardised tests, and how assessment might be taught (e.g., "overemphasising assessment as a measure of a students [sic] ability, especially using data to profile students").

Question 16 asked respondents if there were aspects of educational assessment theory and practice that they would have liked to have covered that were not included in their ITT. Again, respondents were given free response boxes to name the aspects and further boxes to explain why. Respondents mentioned:

- wanting assessment to have received a bigger focus during their ITT;

- the links between curriculum and assessment (rather than adding assessment on at the end);

- expectations of stages of development at specific education levels (e.g., national testing, GCSE);

- how to assess across different subject areas;

- the need to base school systems and practices on theory (including for accountability purposes, and how to identify good research);

- using assessment to inform planning (e.g., writing lesson objectives);

- how to enable students to harness feedback from assessment (i.e., what happens after assessment);

- specific assessment topics (IRT, classical test theory, formative assessment, Bloom's Taxonomy, Stobart's 'uses and abuses of assessment', Ebbinghaus's theory regarding testing frequency (see Dempster, 1989), comparative judgement, effective questioning, effective marking, continuous assessment);

- specific theoretical issues in assessment (validity, reliability, accessibility, bias, equity, setting appropriate questions, identifying key constructs, standardisation);

- the ethics surrounding assessment (i.e., around the pressures to supress/inflate data);

- the administrative processes required (for statutory or GCSE assessment, online tracking systems).

One respondent summarised: "Teachers need to know when to assess and what and when to test and what for" while another said; "Teachers need to know that assessment should be used to inform the next steps in a child's learning or to help identify gaps in learning and not just to collect data." Some of the additional responses to this question mentioned change over time (e.g., introduction of key stages, league tables, 'gaming the system') and influences on attainment including where these relate to attainment below expected levels (e.g., SEN, mental health, date of birth).

6.7 Educational assessment course delivery

Question 17 asked respondents how confident they were that their ITT provider had the right level of expertise to deliver all aspects of educational assessment theory and practice, to the same depth and quality as all other aspects of their ITT course. Figure 6.26 below illustrates responses overall, by when respondents completed their ITT course, and the

phase of education in which they were trained.[56] 44.2% of respondents overall were 'not at all confident' or 'not very confident'. Those who had completed ITT after 2000 and those who had been trained in secondary education had the highest 'very confident' ratings.

Figure 6.26: Question 17, confidence in ITT provider expertise regarding delivery of educational assessment theory and practice
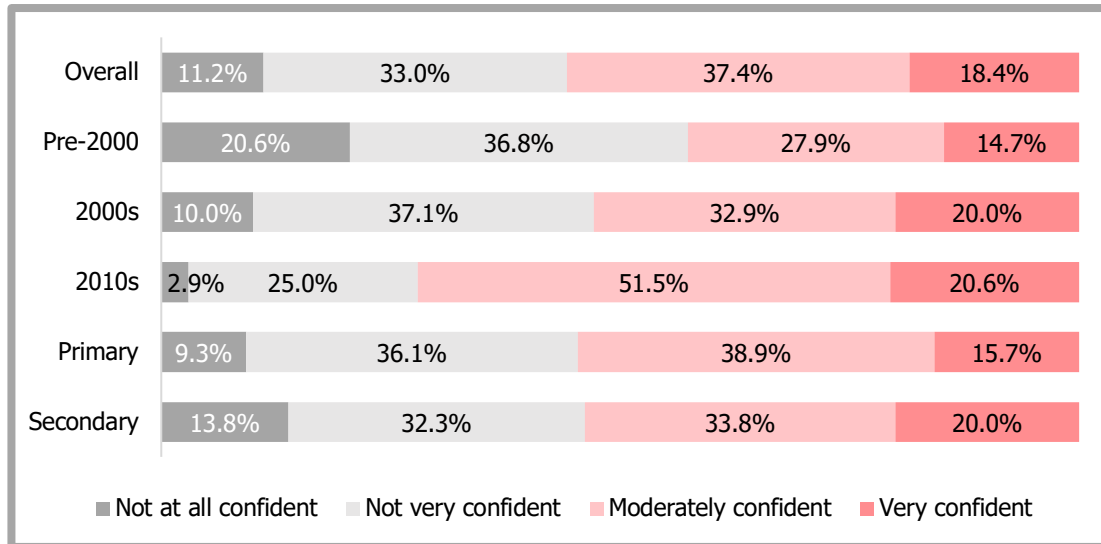

Question 18 asked respondents if they would be prepared to share copies of their ITT course curriculum or other documents that provided details of their course coverage, either in confidence or to state if the materials were already in the public domain. Of the 158 respondents completing this question, 59 answered 'yes' and 99 'no'.

Question 19 asked respondents if they had engaged in any research on educational assessment theory and practice since they completed their ITT. Since their ITT, 57.8% had engaged in research on educational assessment theory and practice (out of 206 total respondents to this question), compared to 72.5% of those who had completed ITT prior to 2000 and 64.1% of secondary-trained respondents.[57] See Figure 6.27.

---

[56] Overall respondents n = 206; Pre-2000s: 68; 2000s: 70; 2010s: 68; Primary: 108; Secondary: 65.

[57] Pre-2000s: 69; 2000s: 68; 2010s: 69; Primary: 108; Secondary: 64.

Figure 6.27: Percentage of respondents who had engaged in research on educational assessment theory and practice since their ITT
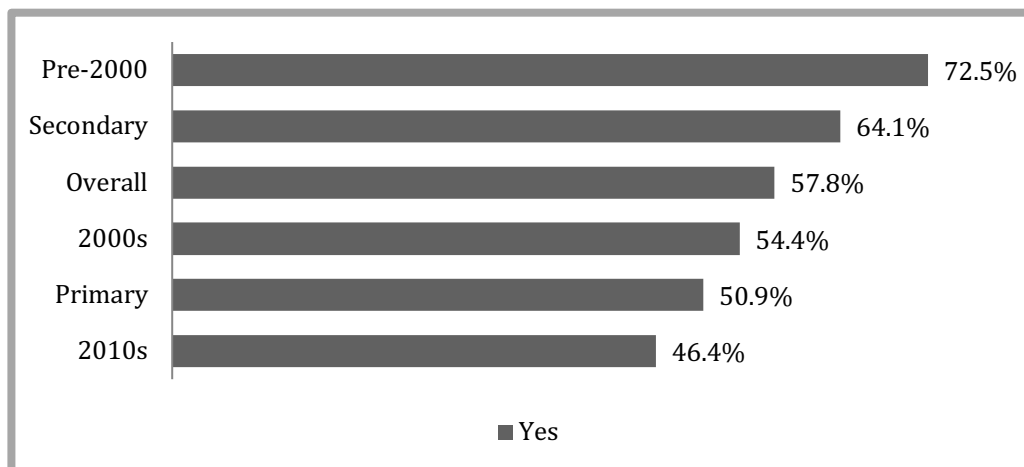
Questions 20 and 21 provided respondents with the opportunity to take part in a follow-up interview, receive a copy of the final analysis of the data or add any further comments; no further comments were made.

6.8 Chapter 6 summary

It is noteworthy that the highest 'very confident' rating was only 28.2% (regarding 'the fundamental principles of assessment, including: the differences between formative and summative uses of assessment'). Similarly, over 39% of respondents were 'not at all confident' that their ITT course had fully prepared them for their first years in teaching in terms of the key assessment principles of: criterion- and norm-referencing (42.5%); utility (40.8%); the design of standardised tests (such as those that produce a reading age) (40.8%); validity (40.2%); bias (39.2%); and reliability (39.0%).

Responses generally indicate that assessment aspects were included in ITT for a higher percentage of respondents completing their ITT post-2000 or 2010, and in the primary phase. At the same time, there is a trend for respondents who were trained in secondary education to feel most confident that they had been fully prepared in these areas with those who had completed ITT prior to 2000 feeling least confident. The data supports the NAHT conclusion that there is a worrying level of trust within the profession regarding assessment carried out by teachers with over one-third of respondents stating that they do not feel trusted by colleagues in other schools. However the data show that trust is more nuanced with respondents feeling highly trusted by their immediate colleagues, parents and pupils but much less so by national agencies and particularly by politicians.

Respondents supported the conclusions of the Carter Review of ITT that there are significant gaps in the capacity of schools and ITT providers in the theoretical and technical aspects of assessment.

Chapter 7: Interviews with Teachers: data analysis

7.1 Introduction

This chapter is presented in three sections one for each of the three teacher interviews (see Section 3.10 for an explanation of the approach and Table 3.3 for the interview schedule). Section 1 covers Interview 1 where the focus is on the participants' perceptions of their early experiences of educational assessment from their days as school pupils, through their time as students, their experiences of initial teacher preparation and their preparedness for entering the classroom post qualification. Early recollections of educational assessment were explored to see if the teachers were influenced by their own experiences as students. Section 2 reports on Interview 2, which examines the teachers' perspectives on their current experiences and views of educational assessment. The focus here is on the teachers' perceptions of the use of educational assessment as an instrument of accountability and as tool to aid teaching and learning. The final section covers Interview 3 where teacher participants discuss their views on the function of educational assessment now and in the future, and the extent to which they perceive the adequacy of their knowledge and understanding of educational assessment.

7.2 Participants

Table 7.2.1 provides a summary and pseudonyms of the teacher interview participants. A fuller account is presented in Chapter 3, Section 3.9.1.3.

Table 7.1: Teacher interviewee participants

Name and Reference	Phase	Current role	Background	Key reasons for Selection
Graham **GrahamP**	Primary	Deputy Headteacher	Completed initial teacher training in 1993.	Holds responsibility for assessment across the school. Moderator for National Curriculum Assessments.
Helen **HelenP**	Primary	Deputy Headteacher	Completed initial teacher training in 1996.	Holds responsibility for assessment across the school.
Tim **TimP**	Primary	Year 6 teacher	Completed initial teacher training in 2013.	Qualified for 5 years. Year Six teacher preparing children for National Curriculum Assessments.
Angela **AngelaS**	Secondary	Science teacher	Completed initial teacher training in 2001.	Holds senior management position – head of science. Teaches GCSE classes.
John **JohnS**	Secondary	Vice-Principal	Completed initial teacher training in 1999.	Senior Leadership position. Has experience as a marker and examiner for GCSE Assessments.
Sara **SaraS**	Secondary	Mathematics teacher	Completed initial teacher training in 2011.	Teaches GCSE classes. Works with feeder primary schools.

7.3 Section 1: Interview 1, Educational assessment and participants' early experiences of assessment

7.3.1 The primary years

Despite going through the primary phase of education as pupils at different times including pre and post National Curriculum years, all of the participants' first point of reference regarding assessment was to tests, examinations or formal assessment. For example JohnS left primary school just before National Curriculum assessments were introduced but noted:

> I know I'd been assessed but I don't remember any formal assessment - you'd get a bit of feedback from teachers, but nothing sticks in my mind.

And representative of those who attended primary school following the introduction of National Curriculum assessments, TimP stated:

> …at primary school I think we were one of the first year groups to do our SATs tests, so I remember doing SATs at primary school in a similar way to how they do them here, but we did a science one as well.

In all cases, assessment was viewed as being less prominent than it is today in what GrahamP described as; "…a very different era in terms of no SATs and no standardised stuff" (GrahamP). This viewpoint was reflected in AngelaS's comment: "I always say it must have been really lovely to teach back then because there appeared to be no accountability" (AngelaS). The idea of a different era in schooling was a repeated theme with references to an 'idyllic' and 'enjoyable' experience. However, in considering these perceptions it is important to note that the views are being expressed on the one hand through the lens of a pupil and the other as a practicing teacher. Recollections of the primary experience were often expressed through references to pressure and of particular interest, to a lack of perceived pressure, as exemplified by GrahamP:

> I mean primary just seemed like quite idyllic really and, yeah. There did seem to be very little kind of external pressure or, you know, outside pressure from exams or government.

7.3.1.1 Accountability

All interviewees reflected on how pressure is now a feature of current times, principally through the accountability system. For example JohnS illustrated the change in pressure from his time as a pupil with the current accountability system:

> I never felt that any of my teachers were under pressure to deliver results really. If they were, they hid it very well, and it makes me think about where we are now and the experience that I had at school and the experience here. I start to ask myself questions like, was it just my school that was like that or was it education in general at the time? Because I've noticed… that through my career, how things have changed.

7.3.1.2 Marking

Participants presented assessment as being less formalised with most comments related to the act of marking work. For example:

> …you went to the teacher's desk, you got them (books) marked, and then they told you whether to go on to the next page or whether to go and sit with the classroom support (HelenP).

Though in GrahamP's case he could recall little evidence of assessment, though quite clearly some marking did take place:

> …feedback that I can remember really from teachers was through marking. So I can still remember bits and pieces of work that I'd got, you know, A, A+s.

However, AngelaS highlighted the limited usefulness of the feedback she received:

> I don't remember doing any terminal papers at all in primary school of any year group.   The only time my work was assessed was when I did activities in class and they were marked, and I didn't have any feedback from that other than they're either right or they're wrong.

HelenP recalled differences between the primary and secondary phases of education in the way teachers interacted with classes through the marking process:

…at primary school, they did move around, but I do think it was more checking that you knew what you were doing, not moving you on. But high school it was very much the teacher at the front of the class, and not going round at all, just marking your books, getting your book back the next day and doing your corrections.

7.3.1.3 Ability setting

Five of the six teachers made references to the way classes were organised subject to assessments during the primary phase. For example HelenP noted that:

> …you'd have to get a certain amount right on the particular page you were doing before you could move on to the next page, and if you didn't get them all right, you then went on to a different table, and you had to stay on that table until you did understand it.

However, although mentioned by TimP, there was no recollection of grouping by perceived ability in his primary phase experience:

> …we were just sat mixed around, there didn't seem to be a higher or a lower table, so there can't have been separate tables for ability.

7.3.2 Participants' secondary, high school phase of education

7.3.2.1 Pressure on assessment outcomes

The teacher participants presented a similarly unified picture of their time as pupils in the secondary phase of education as an era with less pressure around assessment than found in today's practice. For example, SaraS noted:

> And then secondary school, this is just what I remember but I might have forgotten, but it seemed a lot simpler than now.

HelenP presented a similar viewpoint:

> I don't remember any kind of formal assessments (at High School), …it was just *this is the curriculum content, this is what you're learning in History*, and you did it. And then you moved on to the next thing.

JohnS, AngelaS and HelenP presented a similar viewpoint exemplified by HelenP's description of her time as a secondary pupil:

> …to be honest with you, through the whole of the high school period, it was pretty much the same really, I didn't feel any - unless it's just me - I didn't feel any particular pressure from the school. Obviously I put pressure on myself because with my own revision and things, but it didn't feel like we were being pushed to do more, there were no extra revision classes or anything like that, it just seemed very much the same – '*this is your content, you do it, you move on*' really.

Whilst none of the interviewees recalled excessive pressure during their time at secondary school, GrahamP noted a more obvious presence of assessment: but this was not universal as AngelaS noted:

> My books were not marked at all by anyone. I remember seeing in books a sprinkling of "good." The first teacher I remember marking my work at all, is my English teacher in year 10.

All six participants made references to an increase in pressure by comparing their experiences as pupils with the current situation in schools of which SaraS's comment was typical:

> I didn't feel that pressure that I know is here today because as a teacher I feel that pressure that my kids need to do well for me as a teacher not just for them, I don't remember teachers putting that sort of pressure on me really.

However, AngelaS and GrahamP made note of how they make attempts to minimise pressure on their pupils, but as GrahamP noted; "…as much as we try and take the external pressure off the kids, it is there".

7.3.2.2 Setting by ability

The use of assessment to define teaching groups was clearly more pronounced during the secondary phase of education, a theme commonly identified by interviewees:

> …secondary school differed, because then you were getting you're A's and your B's, or your percentages of your tests and things, and setting as well, we setted from straight away basically, for English, Maths and Languages...I recall quite a lot

of movement, it wasn't just at the start of the academic year where you moved, it was within the year, so I'm kind of assuming now that it was sort of termly, because I used to fluctuate between the top Maths set and the second Maths set (HelenP).

And despite AngelaS's earlier comments about the limited experience of her own work being marked, she recalled a similar experience to HelenP of setting:

I remember doing terminal papers at high school; we always did end of year exams. We always got them marked and we were always given a percentage…and then we were set – based on a ranked percentage in every subject.

In a number of the interviews, the teachers commented on the form of feedback in the marking of their work at secondary school, which mainly comprised of marks gained, right and wrong answers and comments about effort. For HelenP, this approach affected her confidence during her experience of secondary school:

…it was like you were given something back that you couldn't do, but you were not told how to do it, so you still couldn't do it, so that did …thinking back, that did knock my confidence in Maths quite a lot.

Conversely, TimP's recollection was more positive:

I had quite a positive experience at high school, I think the assessment that was used did spur me on, it did make me realise any areas that I wasn't sure on.

7.3.3 Key Stage 3 and the Sixth Form: a step change in the use of assessment

All participants reported a step change in the use of assessment and feedback in their sixth form days albeit a 'tiny bit' in JohnS's case. A common theme raised in the interviews around the use of assessment in the sixth form stage, was the way in which it was used to prepare for the final examinations. This was in contrast to some earlier experiences during key stage 3 reflected in the following two examples from AngelaS and JohnS:

I remember assessment in year 9 because I was the first year to do SATs…I remember my science teacher coming in saying, "We're going to do an SC1 investigation" and we had to do it this certain way. I don't remember what we did

with that. But I don't know what happened after that and I don't know what they did with it. (AngelaS).

We weren't prepared for it. We were just given it, and I remember the question – what I found really interesting is, I remember questions that I couldn't do (JohnS).

SaraS described a similar view of key stage 3 assessments as a low-key occasion: "I certainly mustn't have felt anxious about it or anything because I just don't remember doing them". And TimP described a similar experience of his time at primary school where the preparation for the tests was less evident than in current practice:

I remember doing SATs at primary school in a similar way to how they do them here…but I don't remember doing a great deal of revision, like we do now.

It is of note that the teachers involved in the interviews sat some of their National Curriculum assessments during the period in which they were introduced into the English education system, or not long after their introduction, which was described by GrahamP as a "different era". However, at the time of the introduction of National Curriculum assessments as a statutory requirement, there was considerable objection to the tests and tasks by teachers. Much has happened since that time which is reflected in the comments made by the interviewees when they compared their experiences as pupils to the way in which the outcomes of testing are used currently and their identification as being 'high-stakes; a term unlikely to be heard around the time of their introduction as discussed in Chapters 2 and 3. One of these features is the so-called concept of 'teaching to the test' – a process of directing teaching solely to the preparation for an examination or test. The notion of preparing for examinations or tests is clearly evident in the experiences of the teachers in their sixth form days. For example, GrahamP recalls his preparation for GCE A levels making reference to their 'high-stakes' nature:

I remember going into my maths A level and because again, it's high-stakes testing I suppose, and it was a really important exam. But we had practised; we'd done past papers and so on.

Interviewees in general reported an increase in the amount and quality of feedback on their work during their A level studies noting the possible influence of smaller class sizes. For example TimP stated:

I got lots of feedback actually, that was really good. And there was loads of support there as well if you needed it. I think because class sizes are smaller, so it was obviously more manageable for them to do that…it was done in class time. When we'd be working…they'd come around and give it you…only have two or three minutes with you, explain how they'd marked it and why they've given you what they've given you.

The notion of preparation for A level examinations was a common feature mentioned by interviewees, and in the main in a positive light. For example GrahamP noted in relation to his A level maths course that:

It was very much brilliant preparation. I mean this is somebody who knew the exam board inside out, – we must have got through pretty much the majority of the course work in the first year and then the second year was going back, revisiting and going through past papers and so on and really preparing for the exam.

7.3.3.1 Gaming the system

SaraS pointed out that the feedback and general support she received in her coursework was similarly helpful, but self-regulated her comments to make clear that this did not amount to undue assistance from her teachers in ways that are often described as 'gaming' or 'cheating'. However, SaraS referred to what she described as being able to "play the game" in finding marks in her coursework. This fine line between how teachers help and support their students with coursework as opposed to over-aiding or drilling for the exam was broadly evident in the interviewees' discussions.

7.3.3.2 Self-imposed pressure

References to pressure were made by all interviewees in relation to their GCE A level studies, but each pointed out this was self-imposed with the purpose of moving on to the next phase of education rather than emanating from their teachers or schools. Interviewees often referred to their own motivation or wanting to please others including their parents and teachers. TimP effectively summarises this viewpoint making reference to the pressure evident in the current accountability system:

I think the pressure came from me, because I wanted to do my parents proud, and I wanted to do it for me as well. And maybe to some extent, the teachers that had

> helped me, but I don't think the pressure was put on maybe in the same way as it might be now.

SaraS and JohnS contrasted their time as students with the current pressure felt as teachers. For example SaraS stated:

> I didn't feel that pressure that I know is here today because as a teacher I feel that pressure that my kids need to do well for me as a teacher not just for them, I don't remember teachers putting that sort of pressure on me really.

7.3.4 The university phase of education

7.3.4.1 Self-imposed pressure

The interviewees related their experience of assessment to their time at university with contrasting experiences. For example, SaraS raised again the concept of self-imposed pressure but noting that this in some way related to less pressure from her university until final examinations:

> I was really stressed (doing university finals), but that was just me personally, it wasn't from my lecturers or anything, but I was quite stressed that I wanted to do well.

For JohnS, the process of assessment at university was markedly different to that for GCE A level and GCSE that in his words opened up opportunities to 'play the system':

> I soon realised the difference between A-Level and GCSE, where it was externally set – what I realised with university maths is that it's set by the lecturer right? Yes, so this is when I start – university is where I learnt to play the system...it starts to become a bit more strategic...we just had to look at past papers, completely past papers, and it was all about looking at pattern spotting.

However, he was more complementary in terms of the level of access to useful feedback from his lecturers. GrahamP made reference to what he viewed as 'very little feedback' and that university lecturers constructed examination papers suggesting that their overall assessment was based on students' performance over the courses and not the final examination alone in what he described as a 'subjective' approach.

7.3.4 Prior experience impacting current practice

One aspect that came through at various points in the interviews was the extent to which some of the teachers' own experiences of assessment during their education influenced their current teaching and assessment practice. GrahamP reflected on the way his teacher prepared him for his A level maths examination and the feeling of elation by completing an examination well and how he uses this to drive his own teaching. In JohnS's case, he noted how he became fascinated by assessment during his secondary phase of schooling and at university:

> …my first memory was the Key Stage 3...and I just became a bit obsessed with assessment tests then, interestingly. So, I wrote off to the NEA or NEAB at the time to get the past exam papers. But assessment wise (at university), again, I was fascinated with past papers.

AngelaS developed a similar interest in examination specifications, though mainly driven by the lack of support available from her teachers:

> But the exams officer, he could order you specifications, but you had to buy them, but people did, so I didn't do that for GCSE and I didn't have any revision guides but my friends did. But by A level I were a bit more clued up.

7.3.5 Routes into teaching

The six interviewees entered teaching through a range of different routes. HelenP, TimP and AngelaS noted that teaching had been a long held ambition but only HelenP followed a four year BEd course based on advice received from her school's careers adviser that:

> …if you definitely knew you wanted to be a teacher, then that four year course was good because it would give you lots of experience, so that's why I went with that.

Conversely, AngelaS received advice that suggested limitations with the BEd route:

> ….my head of sixth form called me in and said that if I chose to do a Bachelor of Education and I didn't like teaching, I would be stuck. So he recommended that I

did a degree and did a PGCE if I still felt interested, which would still take four years. So I did – I followed his advice.

TimP began his career as a teaching assistant and like SaraS followed the Graduate Teaching Programme (GTP) who as mature students needed financial support as they trained. Financial incentive was also noted by JohnS who was initially attracted by the offer of a £4,000 bursary to train as a mathematics teacher. SaraS and TimP mentioned the need or expectation of experience of working in schools prior to application. As they were working in other fields, taking on any kind of voluntary experience would not be financially viable and so the GTP which provided an income was seen as the only viable option. Financial concerns when considering leaving employment to train as a teacher or taking part in any type of voluntary work prior to application were also mentioned by GrahamP.

7.3.6 Initial Teacher Education/Training stage

7.3.6.1 Course format

The six participants reported varying experiences of instruction in educational assessment during their period of training. The structure of instruction reflected the various routes taken into teaching. For example, HelenP, GrahamP, JohnS and AngelaS referred to lectures and assignments whereas for TimP and SaraS, the GTP route revolved around practical experience in their placement schools with instruction sessions usually held on one day or one half per week. These sessions brought trainees from a range of schools together to listen to speakers, share experiences and engage in practical exercises. Across all six participants, there were very few strong memories of any initial training relating to the theory of educational assessment, and where there was recollection, the frequency and depth of study was low. For example, JohnS stated in reference to his university PGCE course:

> Then one of the weeks - I remember them giving us a GCSE paper. It was a Friday afternoon, so you're not going to get the best out of it as well, and we had to look at this assessment and be critical – pull it apart and say, what's the point of this question,– how is this useful as an assessment item, etcetera? That's all I remember.

Each of the four participants who followed university based courses referred to assignments and lectures on assessment, but in HelenP's case, she noted:

I remember it was possibly one of the teaching and learning lectures, and it was just about questioning…it was just a part of good teaching and learning, that's what you do, but I don't remember anyone giving you any like tips or things.

AngelaS had to complete one assignment on assessment in her PGCE course that she described as being 'limited' to formative and summative assessment, but she also referred to lectures on what she described as the "mechanics of teaching" covering how to plan lessons, do a risk assessment, and how to check pupils understood through "…formative assessment. How to do summative assessment. What that meant, how to write a report. That's what lectures were about".

GrahamP was the only participant to mention having instruction in his PGCE course on what he described as the "theoretical underpinnings around assessment" but that "…It didn't really impact massively onto my teaching practice". And although GrahamP described his instruction on assessment as not being "a massive standout" he suggested that the timing of entry into teaching might have had an impact on the focus of his training course. Qualifying in 1993, he noted that his training focused on the 'new' National Curriculum and that when he began teaching he felt at an advantage; "…around knowing the curriculum and models for how to assess it" compared with practicing teachers who had to take on a new approach without having had the time to study the requirements and implications. AngelaS and SaraS made similar references to changes at the time of their qualification. AngelaS noted that; "levels were the big thing" because of changes in the use of attainment level criteria. By contrast, SaraS's view was that; "…behaviour management was absolutely massive and that probably overshadows a lot of other things". However, whatever was current didn't always apply as HelenP recalled:

The first time I really came across levelling was in my Year 3 placement, because the school that I was at…it was a great school, but they were not very friendly. And I was, as a third year student, given a pile of books, and said, 'Right, they all need levelling, because we have a planning meeting and we're going to say what level all our children were at', and I remember that because I couldn't do it, I didn't know what she meant by levels, we'd had hardly any input whatsoever. I remember stood at the bus stop crying with this bag of books thinking, 'I don't know how to do this', and I had to go back then to look at the level descriptors, but that was the only time in the four year course, and in to my first couple of years of teaching, that was the only time that I'd looked at levels.

The two participants who trained through the GTP could not recall instruction in assessment theory and practice. The majority of their training was through time in the classroom, most often through actual teaching. The one-day or one half day of meetings with other GTP trainees provided little on assessment theory and practice. SaraS recalled that she received; "No (formal instruction on assessment), not in my training no" and TimP recalled:

> …it doesn't stick in my memory at all. So any moderation, any assessment, anything like that has come from working here in the following years that I've obviously learned as I needed to learn it.

This was supported by SaraS who said:

> I think a lot of it depended on how good your school was or your mentor was. (Name of tutor) didn't give us any extra guidance from my memory.

TimP suggested that the weekly sessions didn't cover assessment theory and practice because the assumption was; "…trainees would cover it in school and we had to follow the procedures that the school had so maybe that's why it wasn't included".

Despite having now taught for several years, HelenP and AngelaS noted that new entrants to the profession show little if any signs of change in terms of their assessment knowledge. In stating this claim, HelenP referred to her experience of interviewing newly graduating teachers and the lack of any understanding of assessment issues shown by interviewees. However, HelenP was clear that the amount of change in the education system, particularly with regards to assessment, make it really challenging for ITT providers to adequately prepare new teachers.

7.3.6.2 The role of mentors

AngelaS and SaraS talked about the role and significance of mentors during their ITT school placements. For SaraS, the quality of one of her mentors was significant in terms of learning about educational assessment but noted that mentors would vary widely in quality. AngelaS spoke of her mentor helping her to take an interest in assessment and although she felt that she would have worked out some of the approaches on her own, her mentor accelerated the process. Expressing the view that not all schools have good quality

mentors, she compared this with her current school: "It's clear from my own faculty that some people know diddly-squat about assessment".

7.3.6.3 Point of transfer into teaching

The participants were generally ambivalent in their views on their preparedness for teaching in regard to their knowledge about educational assessment. For example, TimP declared he wasn't prepared, but as he had trained through the GTP course, his first teaching post was in the school he trained, so he was therefore familiar with the systems used in that school and that he had built a good relationship with other staff so he was confident about asking for help. However, had he moved to another school, he noted; "...it might not have been the same case, but I think because I was so comfortable here, that helped a lot". TimP also expressed the view that teachers in general have to adopt and fit in with the assessment system used by the school. JohnS noted that it was only when he began teaching that through his reflective nature elements of his initial preparation did help inform his practice, but that it was his own classroom experience that provided a greater influence:

> I'm a very reflective person. So, probably all the stuff that we did do through university, did help me in some ways prepare for teaching in general, but I felt it was more the classroom experience that gave me the greater preparation.

HelenP recalled that there were no assessment polices to guide her early career and that the curriculum was not highly structured leaving it to individual teachers to do their own planning in isolation. In terms of assessment, HelenP noted that she felt "totally unprepared" and "scared" to use the system of National Curriculum levels of attainments and that the focus of assessment was on writing a report at the end of the year. AngelaS was much more positive about her ability to use the system of levels and using commercially produced tests, but put this down to the influence and guidance of her mentor who at the time was completing a Master's project on assessment. For SaraS, it was only after qualifying and after she moved schools that her assessment knowledge developed:

> I learnt the most about assessment in any form when I came to (current school) 100% because I remember when I had my big observation from the lady that ran the GTP, and I will never forget her asking me this, you know when things stick with you? And she just said how do you know that they know? I was like, I don't know, and

she said, well you need to know. You need to find a way to assess it without just having a feeling about it, you need to know.

7.3.7 Perceptions of educational assessment over time

All of the participants were unequivocal about the amount of change in educational assessment practice when comparing their time as pupils to the present day. The consensus was that the amount of assessment has increased considerably though this was not seen as necessarily a negative thing. TimP broadly represented this view in saying:

> Assessment has increased massively…now, assessment for me from a teacher's point of view, we're assessing all the time, but it's useful, I need to assess all the time to know where the children are at and what I need to plan and what I need to teach.

However, JohnS was of the opinion that there is now a tendency to assess too often and much of it driven by a desire to generate data rather than for the benefit of teaching and learning. GrahamP agreed that there have been significant shifts in educational assessment and recollected the relative freedom of his early teaching career when assessment was less pressurised but nevertheless important.

All participants expressed a clear sense of growth in pressure on pupils, teachers and schools. The early use of National Curriculum tests was viewed as being less stressful with AngelaS describing them as; "…they weren't anything major at all'. However, SaraS noted that there is now more pressure on teachers than when she was a pupil and that pupils are aware of this to the extent that some pupils in her opinion believe that teachers' pay depends on examination results. She also put some of this down to basic competitiveness:

> Within the school…you want to get really good results in the department. So it makes you feel like one of the stronger teachers I guess, if I am being honest. But then obviously there is a collective thing because you want the school to do well so you would never like want your colleagues to do bad, you just want to do better.

This view was balanced with a desire to see all students achieve to their maximum capabilities.

7.3.8 Accountability

It was clear from each participant that the key driver of changes in educational assessment practice is the current system of accountability, although again there was an element of ambivalence, for example JohnS stated:

> Personally, I think the driver is the system that we're in. I think it's everything from Ofsted. Even though I believe in league tables. I think it's this constant drive for improvement in a game that you can only be average. At peak, there are going to be some people under average and some people over it. I think that's the biggest issue with education at the moment.

However, there was a strong view that some of the purposes of assessment had become distorted with an obsession on measurement and the loss of focus on the quality of teaching. JohnS again summed this up:

> I think we're our own worst enemy as a teaching profession because if we actually all just took a step back and thought, right, what we need to do now is actually just reflect on where we are as an education system and say, do we need to do these things all about marking, feedback, etcetera, etcetera...because I feel what happens is – we lose the focus on quality teaching. I don't think teaching has particularly improved really. I think – in fact, I would hazard a guess, I think maths teaching – the quality of maths teaching is decreasing.

AngelaS was more supportive of the impact of the accountability system:

> I think there's very much a basis for not trusting teachers. And back when teachers were not accountable in any way, shape or form, that was not the best education they could provide for that child - because it allows those people who are really lazy or not in it for the right reasons.

Nevertheless, AngelaS's view on the impact of pressurised examinations and tests on pupils and teachers raised concern relating to the wellbeing of pupils and teachers:

> But I'm pretty sure that the instances of self-harming in teenagers has gone up and that'll be exam pressure. We've got a higher level of mental health problems,

maybe it's because that's not been reported before because of the stigma. I certainly feel stressed all the time. I just put on a smile on my shop window and then ignore it. But even –you can see that the primary school kids are getting stressed out as well because of SATs, SATs, SATs.

7.4 Section 2: Interview 2, participants' early years of teaching

7.4.1 The level of preparation

The timing of entry into the profession appeared to influence the way some participants felt about their preparedness for entering the teaching profession. GrahamP related the time of his entry into the profession as a time of significant change in the English education system giving student teachers the opportunity to study the changes in a way that practicing teachers could not: he viewed this as advantageous. Although GrahamP felt there was a general level of familiarity with regards to discussions amongst teachers concerning the curriculum, this was not replicated in terms of educational assessment noting that schools did not have structures in place to support assessment practices introduced with the National Curriculum in the early 1990s. GrahamP did note that the Local Education Authority (LEA) offered support around assessment through courses and written resources aimed at providing examples of pupils' work at specified levels of attainment but that; "It tended just really to be for people in SATs years. So if you were in Year 2 and Year 6, you were able to go on courses". According to GrahamP, 'luck' played a part in accessing knowledge of formative assessment. As science co-ordinator, he was invited to take part in a project that introduced him to the concept of formative assessment for the first time.

SaraS felt less prepared by her training provider on entry into the profession noting the significance of and reliance on her first school for support and development on the curriculum and assessment. She noted that due to staffing issues, she was given a year 11 GCSE maths group in her NQT year and used past examination papers to support her teaching. However, in subsequent years, SaraS designed her own tests often drawing on a published database of test items but more closely related to the topics she was teaching: she stated: "I don't think I was great at general assessment for learning in my early years".

JohnS and HelenP were unequivocal about their level of preparedness when entering the classroom, for example, JohnS stated:

> Putting it bluntly, no. So, I would say that it's only when I started teaching and looking at the demands of Key Stage 3 and GCSE and A-Level…it's only then that I started really to appreciate the demands of assessment. I think I was prepared to teach maths but not to necessarily prepare children for assessment.

HelenP did use what she described as elements of assessment such as marking pupils' work, using lesson objectives and self-evaluation, but this was "definitely my own method, not a whole school method". However, there was a concerted focus on assessment by the school in which HelenP started her teaching career when it changed from an infant school to an all through primary school. At that point the National Curriculum statements of attainment became the focus of the school's assessment, but even here she recalls that when she submitted data on the children she believed had met particular statements of attainment, there was no further conversation or guidance. At the time, HelenP accepted this as a positive sign of confidence in her ability to assess the work of pupils, but on reflection questioned this in comparison to her later experience of teaching. Whilst TimP felt confident when entering the classroom, mainly due to his familiarity with the school as he had trained there, he was less confident in his educational assessment knowledge:

> My first day, when I got my class by myself, the teaching side, absolutely fine. But then in terms of assessment, I was a little bit less confident.

TimP was clear that as he had trained in the school in which he took up his first appointment, he was familiar with the approach used by the school. However, reflecting on his early training, TimP noted that his understanding was more to do with the processes or procedures used by the school rather than a more fundamental understanding of assessment theory and practice. For example, he noted that the on-going daily assessment used in classrooms like using effective questioning developed over time and was based on his own experience. He lamented that such knowledge gained through initial training would have been; "quite useful". AngelaS felt prepared, but put this down to the influence of her mentor rather than through her university course. She also noted that other trainees might not have benefited in a similar fashion. AngelaS stated that she wasn't taught anything explicit about educational assessment in her PGCE year and talked about learning what she described as the 'mechanics' of teaching. Conversely, she also noted that her university course made clear the distinction between the terms formative and summative assessment; but how these approaches manifest in the classroom was not made clear. She suggested that coverage was quite superficial. However, AngelaS did make references to receiving instruction in concepts that can be described as elements of

educational assessment, but because these elements were not explicitly related to assessment, the link was not perceived. She noted:

> We did open and closed questioning. I don't think that was referred to as assessment. It was referred to as questioning.

AngelaS explained that this may be down to what she described as the 'coming and going of crazes', or through the limitations of time in a one year course or indeed her own failure to ask or make the link. Time limitations and what she described as going through the mechanics of teaching was a recurring theme in AngelaS's recollections of her initial preparation. AngelaS defined 'the mechanics of teaching' to include:

> Things like planning and how to mark properly, how to teach properly, how to do the practicals, what instructions and how to give instructions to your class. You know, the classroom management, things like that, the things that you do in the classroom every day.

AngelaS recalled further details on her initial training in regard to educational assessment, but she suggested that the lack of clear terminology or direct reference to assessment theory and practice might have masked some elements of the course. AngelaS also noted that she lacked confidence in using assessment techniques as she embarked on her teaching career turning to past examination papers for support – though she explained this may also have been down to convenience: "I didn't trust myself to make assessments, probably because I'd never been shown how to do it".

7.4.2 Training and continuous professional development

SaraS, like GrahamP, reflected on the current provision of courses and training for teachers but was of the view that passing on learning from a particular course to other colleagues is problematic and that schools do not benefit more broadly through the formal reporting or feedback on training provided to individual teachers. JohnS, SaraS, TimP and AngelaS all noted that their use and understanding of educational assessment developed over their time as teachers, more often than not through their own reflection on practice rather than through courses or external support. In AngelaS's case, she noted how access to CPD had declined since she began teaching, citing the scarcity of financial resources in schools as the driver noting that she even financed her own Master's level study. AngelaS explained that all CPD is now internally run with no access to external speakers

or support, though she could not recall any of the training content being related to educational assessment. Training in educational assessment, particularly in their early careers was rarely reported by participants. Even though JohnS had acted as a marker for Key Stage 3 tests and GCSE examinations, it was not until he gained the post of GCSE reviser that he received any form of training in educational assessment other than the application of a mark scheme:

> I don't remember any training ever on assessment. The only training on assessment was when I became a reviser for GCSE. That's my only training for assessment, really.

Even here, the training was not extensive or formal and JohnS put this training down to the experience of working with two 'excellent' principal examiners:

> You had a bit of training, but it was more what you literally learnt sat in a room listening to their comments on why something is a good question, why something isn't a good question, etcetera.

However, JohnS's practice differed from other participants in that he makes a real point of passing on to his school colleagues what he has learnt from his experience as an examination marker and reviser: this practice is now common to other examiners working in JohnS's school, but this was not a practice noted by other participants. There were further references to possible educational assessment content in AngelaS's commentary where CPD or school training sessions may have covered relevant aspects, but as they were contained within other activities, they were not fully interpreted as being assessment related. Even here, AngelaS was dismissive of the value of the assessment techniques contained within broader topics stating that teachers used superficial techniques, such as asking children if they were confident in the learning demonstrated by the holding up of green cards – at which point teachers moved on.

7.4.3 Defining educational assessment

Most respondents talked about how they *used* educational assessment in their teaching rather than offering a definition. GrahamP presented the view that assessment is a positive tool for teachers and pupils. A key element of assessment for GrahamP is the facility to gain an understanding of what pupils have learnt from the teacher's instruction, and for those pupils who have not internalised the learning, as a tool to help teachers to

understand why not. He also noted how he used assessment as a means of self-evaluation. JohnS supported the use of assessment as a tool for teachers to reflect on the impact of their teaching, but was keen to emphasise that teachers should reflect equally on things that go well, something that is often overlooked. For TimP, assessment is a day-to-day activity by which the teacher keeps a constant eye on children's understanding and if necessary, refines the teaching sequence. TimP explained further that termly and half-termly assessments played an important part in informing teaching and the planning of intervention strategies. These assessments, particularly end of year tests at the end of years three, four and five, play an important part in passing information to the next class teacher so that they; "…can then decide where they want children to sit or what they can expect them to do".

HelenP's response was clearly a description of what she called the purpose of assessment. This response reflected the descriptions offered by the two other primary teachers interviewed for this study:

> …the purpose of assessment is that you know what children can and can't do, and if they can't do it, you then have to do something about it, that's the whole point.

However SaraS's immediate response centred on assessment as a means of certification: "…as a way to open a door to a career path. To get a good grade". JohnS and AngelaS differentiated between summative and formative uses of assessment. Although AngelaS defined the formative uses of assessment in similar terms to her primary colleagues, she noted that it was not necessarily an everyday occurrence. However, JohnS was more reflective in his answer clearly differentiating between assessment as a means of certificating qualifications and the day-to-day uses at classroom level.

7.4.4 Lack of clarity on the purpose of educational assessment

Drawing on his experience of visiting other primary schools as a National Curriculum assessment moderator, GrahamP reflected more broadly on how assessment is perceived within the teaching profession, noting what he described as differing opinions and 'confusion' as to what constitutes good assessment practice citing the accountability system as a key source of mixed messages:

> I think there is an element in schools where assessment is not co-ordinated or managed well where it does become this overbearing emphasis on external

accountability and being done unto you — at the end of the day, assessment should empower teachers.

7.4.5 Accountability

With the one exception of JohnS, all of the teacher participants referred to the purpose of assessment as being at least in part driven by the accountability system, either internally for generating performance data for senior managers, or externally through for example Ofsted or DfE performance tables. For example, HelenP noted:

> I think it (accountability) impacts kind of massively really…I think at a leadership level we do feel quite a lot of pressure in terms of performance and performance data, and league tables and all of that, an awful lot of pressure.

TimP also linked the purpose of assessment to national targets set by the DfE, but noted that the headteacher and deputy-headteacher drive performance to meet these targets and monitor progress. This was echoed by all three participant teachers in the secondary phase of education.

TimP and SaraS also highlighted the focus on key stage tests and GCSE examinations. TimP talked about a growing "panic" as the tests became ever closer and described the use of; "…a huge range (of practice tests) because we don't know what's going to come up in the SATs". His view was conditioned by what he saw as undue influence of what he referred to as a "one-day judgement". SaraS described the practice of teaching pupils 'tricks' to get them through the examinations rather than focusing on understanding of the subject:

> I'm going to teach you a trick. Learn this trick, learn it off by heart and it becomes drilled memory to get them through this test.

GrahamP was alone in making reference to the Teachers' Standards as a contributing factor to viewing the purpose of assessment as being a tool of the accountability system rather than its wider use as an element of the teaching and learning process:

> I think there's elements within the Teachers' Standards that probably drive that kind of accountability thing to the top of the agenda, whereas I would much rather that it was a much broader picture around assessment.

7.4.6 Wanting children to do well

Despite references to the pressures exerted by the accountability system, TimP was clear that for him, part of the pressure was self-imposed by his desire to see children achieve to their best. This was a sentiment shared by SaraS:

> ...to get them (pupils) the best grade possible. And not only for the teaching accountability but I think for the students.

7.4.7 Assessment in the English education system

7.4.7.1 Reliance on summative assessment and data

A recurring theme in this section of the interview was the references to educational assessment being driven by data, particularly from summative assessments such as key stage tests of GCSEs. JohnS and SaraS have roles in a Maths Hub that involves working extensively across a range of schools. Both of these participants were highly critical of the quality of the formative use of assessment as used day-by-day by teachers in order to pick up on pupils' misconceptions or to ensure their understanding of teaching content. JohnS described most of what he sees as "pretty poor" whilst SaraS described practice as; "…terrible – on my school visits, assessment for learning in the classroom is practically non-existent in most schools; it's crazy; it really is". JohnS put some of this down to his view that it is easier for teachers to direct the subject content without what he called the 'dangerous territory' of checking for misconceptions that require a level of subject confidence that may well be lacking in some teachers. JohnS related his concerns directly to the use of assessment techniques in the classroom. For example, he spoke of some teachers who he termed as doing too much – not in terms of workload, but by providing answers when children say they can't find an answer. This, he suggested, does not promote learning. According to JohnS, teachers are generally aware of effective questioning, but that they do not get any opportunities to see such approaches modelled. He put this down to the selection of mentors for newly qualified teachers and supporting initial teacher training as being viewed as a low priority.

A central concern for JohnS was his view that expertise held by experienced teachers is being lost as they look to leave the profession. AngelaS was of the view that given recent changes to the way schools are organised, some teachers are effectively working from a set of teaching instructions. She stated:

> If you look at how education has changed, like you go in certain types of academies
> and everything's done for you, so why do they need to? They don't need to plan
> an assessment. It's there, it's already done, they just use it.

Drawing on his experience as a key stage two assessment moderator, GrahamP noted
that in terms of the quality of assessment in schools; "it is a bit of a mixed economy at the
moment". HelenP put some of this down to the removal of National Curriculum Levels in
2014 noting that schools have since generated a range of differing systems. She spoke of
two issues coming out of this. Firstly, that the removal of levels gave assessment a higher
profile, but conversely, it resulted in schools working in isolation to generate their own
models that she described as a 'waste of time'. Further, it raised issues around confidence
and worry that other schools were designing better systems. Secondly, HelenP raised
concern about understanding the different terminology used by other schools. The system
of levels had provided what HelenP described as a common approach and as such was
broadly understood by schools. She noted that the loss of a common approach has
created particular difficulties when children move between schools and schools struggle
to interpret differing assessment approaches and nomenclature. AngelaS noted that there
are differences within schools in the secondary phase of education. AngelaS described
how the various faculties worked in isolation with regards to assessment and how some
of her own faculty members are not involved in making school based, formal assessments,
a move she believed may have hindered their development:

> Because they haven't had the training or because they don't have the same levels
> of understanding as us, we don't allow them to make the assessments. So they
> do their own formative assessment.

HelenP spoke positively about the increased reflection on assessment systems in schools.
She felt supported by attending meetings arranged by the Local Authority where a group
of headteachers and deputy-headteachers could discuss assessment matters noting that
the assessment co-ordinators meetings are particularly well attended. However, in
explaining this further, the agendas for these meetings tend to be dominated by the
provision of information such as the analysis of national and local authority test data, the
latest news from the STA and arrangements for moderation. HelenP noted the group spent
time looking at areas of issue in the performance of schools in accountability measures
and applying for bids to support intervention strategies with the goal of improving results
rather than any discussions about the theoretical side of educational assessment. AngelaS
also raised concern about the quality of the moderation of GCSE coursework and how

interpretations of what is expected is influenced by the moderators personal experience rather than applying the agreed standards.

GrahamP was supportive of the idea that every school should have a designated assessment co-ordinator which he described as a huge role given the accountability requirements currently in place and the focus of many schools on tracking pupil's progress. However, he raised concern that one danger in this role is that it could be become an administrative role rather than one that focused on the use of assessment as being integral to the quality of teaching and learning. SaraS raised similar concerns.

7.4.8 Newly qualified teachers

HelenP and SaraS discussed how initial training and teachers new to a school are areas requiring support. SaraS noted that newly qualified teachers enter the system with a set view on what is expected around assessment and that once that has been accepted as an approach it is difficult to change. For more experienced teachers joining a new school, there is a tendency in some to stick to familiar routines. For HelenP, the removal of levels again posed problems as new teachers joining her school were now more likely to have experienced a different approach to assessment. This requires more of a focus in inducting new teachers into the system used by the school.

7.4.9 Potential changes to the system

7.4.9.1 Accountability

GrahamP was clear that the current system of holding schools to account through a focus on test results is driving the system too much, and in particular how a set of poor results can impact on a school. TimP took a similar view. He was clear that tests in themselves were not a bad thing, but he did raise the same concern that a school is effectively judged by the performance of pupils on one single day, more so because the performance has considerable weighting in the way schools are judged externally. AngelaS's response was more focused on concerns with the way accountability measures are used by schools in the management and evaluation of the performance of teachers which in turn impacts on their pay. She noted that this was not an approach used when she began teaching but was a more recent development. She described the process as a 'minefield' for teachers that creates intense pressure on them to make sure that their pupils achieve highly in external examinations. She was however clear that:

…teachers who are not performing and teachers who are not doing their best for the students, they need to go, simple as.

GrahamP did however support more use of what he described as 'internal accountability' as a means of giving schools more control over pupils' learning without the externally imposed accountability measures.

7.4.9.2 Examinations and tests

None of the participants raised any issues of principle with the continued use of externally set and marked public examinations and National Curriculum assessments. For example TimP noted that he would like to see National Curriculum tests continue, but that they should be combined with other elements when the performance of schools was being judged for accountability purposes. GrahamP supported this view describing the model used at key stage one as a way forward. His view was that externally set tests should be used to supplement externally moderated teacher assessments. However, in terms of the secondary phase of education, JohnS was clear that the recent changes to GCSE examinations had been a backward step driven by a political rather than an educational agenda. Whilst of the opinion that some change in the design and format of GCSE and GCE examinations was probably necessary, he was critical of the reforms. He put the changes down to a political ideal rather than what he described as a pragmatic or realistic response that should have been gradually introduced into the current system. He was critical of a system driven by what he described as being; "…one man's personal opinion. I think it's one man's ideological vision of what education should be like" and "…that too much has happened too quickly". He suggested that there is little point in reacting angrily to things you cannot change and that it is better to focus on those things where teachers can hold some influence.

SaraS took a different view on some of the recent changes to the GCSE, particularly around the removal of modular examinations and the introduction of linear examinations and the limits on re-sits which she believed had previously been more open to gaming or playing the system by schools that focused on cramming students in the final year, or re-taking examinations in a bid to improve results rather than working on subject understanding across each of the previous years of study, GrahamP raised similar concerns to those expressed by JohnS about the pace and direction of change in reference to key stage tests over the last four years. He noted that agencies such as the

STA are equally challenged, for example in developing and publishing timely guidance materials for schools.

However, not all recently introduced assessment reform has been unwelcome amongst participants. HelenP spoke positively about the removal of National Curriculum attainment levels as a catalyst for raising the profile of educational assessment in schools. Even here, there was an element of concern in that it did have a negative impact on the workload of schools as they responded to the change by developing their own individual approaches to assessment, and even then, there remained doubt in the efficacy and transferability of adopted methods.

7.4.9.3 Reluctance to change

SaraS and AngelaS expressed the view that some teachers are reluctant to change, either through what AngelaS referred to as the challenge of making significant change because of negative attitudes around the "black hole" of workload or in SaraS's case that teachers tend to stick to what they know through their initial training and they do not like changes in their routines.

7.4.9.4 Teacher based assessment

TimP and HelenP stated that they would like to see more teacher assessment in high-stakes assessments at key stage 2. Whilst supporting the continuation of externally set tests, TimP considered a more balanced approach to be desirable with teachers assessments being given equal weighting for accountability purposes stating that: "I'd like the teacher assessment to have the same weighting as the tests". As with GrahamP, HelenP also favoured the approach used at key stage 1 with external tests informing teachers' overall assessments based on work produced across the year rather than limited to a short final test. All three primary phase participants advocated a system of external moderation of teachers' assessments. AngelaS also supported an increased reliance on teachers' assessments, but with two caveats. Firstly, she noted that some teachers have the required skill set to apply marking standards consistently though this is not universal adding that this could be addressed through appropriate training. She thought such a move would be 'powerful' in supporting what she described as; "…the ultimate (in) teacher trust because we do know what we're doing". Secondly, she was of the view that marking schemes need to be more precise to aid common understanding and interpretation. However, this was balanced against her view that: "I think there's a lot of cheating in

coursework. I'm glad coursework has gone". She explained that in her view the widespread cheating was a consequence of the pressure on teachers to produce data rather than a real desire to be devious.

As a result, AngelaS was quite clear that she would not advocate more teacher assessment in public examinations. JohnS described himself as "sitting on the fence" with regards to coursework. On the one hand he talked of concerns with an over-reliance on terminal examinations, which is why he favours a modular approach. On the other hand, he raised concern that schools game the system in their manipulation of coursework marks. He explained that by understanding the tolerances allowed on the marking of coursework, schools can 'inflate' marks to the top of the tolerance range knowing that the marks, though not exactly correct, sit within the tolerance limits so would not be changed by an external moderator.

Concerns over coursework were also raised by SaraS who talked about teachers over-aiding pupils. However, she was also clear that if teachers' assessment at primary level was isolated from the perverse incentives resulting from school accountability performance tables, the information derived from teacher assessment would be of great benefit to the receiving secondary school.

7.4.10 Teaching and learning

GrahamP shared AngelaS's view that increased involvement in assessment would empower teachers but for him the reference was towards aiding the teaching and learning process rather than supporting examinations or tests. This view informs GrahamP's quest within his own school in what he described as an attempt to 'inculcate' the idea:

> …that assessment is a force for good that's going to improve you as a teacher and improve the outcomes that you achieve with the kids that you're working with.

7.4.11 Trust in teachers' assessment

7.4.11.1 Perceived levels of trust between schools

All of the participants raised concerns over a lack of trust in teachers' assessment. Some of this was related to practice across the system, for example in relation to cheating in public examinations, with other concerns being related to the level of trust between and within schools. However, there were also expressions of real trust within schools and that

with appropriate training and modifications to the accountability system, trust could be restored across the teaching profession. JohnS was very direct in his analysis of the level of trust between schools and with previous teachers:

> Yes, I think there's a lack of trust amongst teachers, previous teachers – not just previous schools.

SaraS noted that in the transition between the primary and secondary phases, there is little information passed on between teachers or schools. However, she was clear that she did not trust the end of key stage 2 test results based on her experience of teaching year 7 pupils. HelenP reinforced the idea that there is a paucity of educational assessment information exchanged between some schools. TimP expressed the view that he would like to think that other schools in the local pyramid of primary and secondary schools trust his judgment and noted that he had no evidence that his assessments were not trusted. However he did reflect on the fact that the secondary school re-tested year seven pupils on entry and that his own junior school re-tested pupils on entry from the infant feeder school to provide a baseline from the start of their time in the school. HelenP spoke of the same issue and shared TimP's view that children's performances in the May national tests could differ from that in tests set at the beginning of the following September when they joined a new school.

Both TimP and HelenP's analysis stemmed from concerns that schools feel at least a level of unease in the assessment information derived from previous schools. However, SaraS was more critical of the veracity of assessment practice in her comparison between the primary and secondary phases of education describing how primary colleagues are more likely to inflate their assessments than secondary teachers drawing on what she had heard from her colleagues. SaraS expanded on this to include concern over the alleged misconduct of primary teachers during National Curriculum tests, a situation that was not repeated in GCSE examinations because of the stricter administration rules. This further explained SaraS's lack of trust in national curriculum tests:

> I hear a lot about the teachers going round and giving the look or giving the tap; check that one, and things like that unfortunately. So I don't fully trust the things that come through.

However, JohnS did not share the same level of confidence in the administration of the GCSE raising his concern that teachers 'cheat' or 'game' the system, at least in the

coursework element. This type of alleged activity was put down to the pressure of performance measures on schools and a suggestion that it is predicated on the view that "other schools do it". GrahamP also spoke at length on the incentives to 'game' the system in the primary phase of education and the real need for teachers to demonstrate that they can be trusted. However, he raised concerns that the accountability system of performance tables and school inspections is driving miscreant behaviour. Looking to the future, GrahamP raised further concerns about the introduction of the revised baseline assessment to be introduced in 2020 driven by the high-stakes school accountability measures:

> I think the accountability and the high-stakes is driving the system too much. I mean, for example, reception baseline fills me with absolute horror because schools will game the system on that one, if I'm quite honest.

AngelaS talked of a similar example in her secondary school leading to what she described as a lack of trust in some of her senior colleagues. She related an example of a senior colleague asking for changes in the distribution of marks in an internal test because her teaching group had not performed as well as expected. However, TimP and HelenP were of the opinion that based on their own experience, teachers are generally honest in their assessments, in the main due to robust internal monitoring systems. TimP added that the external set and marked tests used for key stage 2 also act as a comparator, which he believes reflects a level of honesty found in other schools. HelenP did however describe her feelings of being mistrusted, particularly by the STA. She described it as feeling under 'threat' of being labelled a cheat even for making what might be honest administrative errors. She was very clear that she did not want to give her children an un-fair advantage because of her concern that it would then provide a false baseline used by secondary schools to measure progress against GCSE results. Despite TimP's broader optimism, he also reported felling mistrusted by an external agency, in his case Ofsted. AngelaS reported a wider feeling of not being trusted relating at least some of this to the excessive demands made on the teaching profession.

7.4.11.2 Subject knowledge

SaraS suggested that to some degree, the shortcomings in trust in teachers' assessment may well be due more to their lack of subject knowledge and understanding of standards rather than deliberate attempts to mislead. AngelaS related this to experience in her school where she compiles tests for her faculty because of her lack of 'trust' in her colleagues'

ability to construct valid assessments. SaraS was of the opinion that there needs to be a system of communication to rectify this shortfall in knowledge, a view supported by AngelaS who stated that by working across schools she had seen that teachers can correctly apply assessment standards, but that they need that level of support and training. And importantly, she felt that flaws in the system were more to do with the pressure exerted by schools to achieve in performance tables rather than the ability of teachers to assess to required standards. As a result she concluded that across the UK, there is cause for concern. However the notion of working collaboratively across schools was challenged by JohnS and SaraS who both described schools as being 'in competition' leading schools to what JohnS called "…pushing practice to the limits in an attempt to beat the school down the road". SaraS summed up the viewpoint:

> …that's exactly how I feel with GCSE. It's just a massive competition. You've just got to be better than the school next door, so how are you going to truly collaborate?

GrahamP spoke with passion about wanting to improve the level of trust in teachers' assessments driven by a desire to improve teaching and learning as "…using good assessment materials and helping teachers to get kids towards the outcomes that they would be expected to achieve".

7.4.12 The role of external agencies

7.4.12.1 Ofsted

The theme of trust re-emerged in relation to national bodies, for example by HelenP who noted:

> …my experience of Ofsted, they just seem to come in with a pre-conceived idea based on data and they don't almost trust, don't believe, what you are trying to say.

GrahamP was equally critical of the Ofsted inspection process noting that the organisation does not provide any level of support for schools.

7.4.12.2 Standards and Testing Agency (STA)

GrahamP noted that despite his earlier comments that STA face a difficult task in keeping up with the pace of change, he was critical of deficiencies in their earlier exemplification

materials. He also raised concern about the training and consistency of moderators noting that in 2016 over half of the selected moderators failed the first standardisation exercise. This was reported by STA, but GrahamP drew on his experience of working for one local authority as a moderator and teaching in another to illustrate differences in interpretation of the expected standards of moderated work. TimP was complimentary about the exemplification materials published by the STA stating that he believed it had helped to raise the standards of performance in schools through setting higher expectations. TimP also expressed his confidence in the ability of the STA to mark key stage 2 test papers noting: "It's a tough system so I know they've got to do it right if they're marking it. So, I trust the marking of those tests". TimP concluded that all schools would benefit from having a member of staff who is also a test marker so that they could; "…share their marking and how they're marked and why they get a mark or why they wouldn't for certain questions. That would be useful".

JohnS reflected his interest in tests and examinations in talking about the data STA holds on tests noting for example the item level data, but lamented that although schools may download the information, he was unsure if they use it effectively to help improve assessment and help improve teaching and learning.

7.4.12.3 Awarding bodies

JohnS acknowledged that awarding bodies are making some attempts to support teachers' understanding of how the general qualifications system operates including the writing of examinations. He expressed the view that this information may not in reality be accessed by many teachers because of other demands on their time, but he suggested that the materials made available by examination boards:

> …could be part of PGCE now, the actual process of examination. Ultimately, most schools want you to deliver good results. So, actually having a good understanding about the exam process and how it works would be a really useful thing.

7.4.12.4 Ofqual

Of the six participants taking part in the interviews, only JohnS referenced the qualifications regulator Ofqual, but was of the view that they appear removed from the day-to-day realities faced by schools and that they did not recognise fully the views expressed by teachers in their consultations. He raised concern that Ofqual does not present

sufficient challenge to policy makers citing the recent changes to GCSE qualifications by the then Secretary of State Michael Gove as an example. However he concluded that Ofqual has an important role.

7.4.12. 5 Local Education Authorities

GrahamP was particularly concerned about what he described as an almost overnight decline in support from Local Education Authorities due to the fall in government funding leading to a decrease in staff offering support to schools. According to GrahamP, most of those leaving the local authority are returning to schools, setting up as consultants or joining newly formed Academy Chains. This is despite the fact that local authorities have to manage the process of moderation. He concluded that it; "…seems like it's not been very well thought through and it's not very joined up thinking". SaraS was also concerned about the impact of reduced funding for education, but took the view that this appears be driven by the government's political ideology and the drive for data and not by the day-to-day needs of schools in what she described as; "…fighting a battle that you just can't solve".

7.4.13 Changes to educational assessment practice

The most frequently cited changes in educational assessment were in connection with the growth in importance of the system of school accountability. In this regard, most references related to the use of tests and examinations in generating competition between schools and in driving performance management within schools. GrahamP, JohnS and SaraS expressed the view that the level of competition between schools has increased over their time in education through the focus on school accountability. This was also reflected in responses regarding the performance management of teachers. For example, AngelaS talked about how the outcomes of examinations have grown in importance in terms of performance management and of their potential impact on employment and career development. GrahamP was of the view that the system at primary school level is much more driven by National Curriculum assessments. But like AngelaS, he was clear that there had been benefits to the quality of teaching and learning through the use of tests and examinations. He spoke of how his experience of using of standardised tests in years 2 and 6 can hold benefit in developing the capability of teachers and should be used in other school years to improve the outcomes for pupils.

However, GrahamP added the caveat that the purpose must be clear: to make sure that pupil's learning remains the focus and not the generation of data. TimP held a similar viewpoint. He was of the view that since he began teaching five years ago, the system had made what he described as; "quite a lot of progress" in that assessment is being used more judiciously citing examples including a clearer focus on the purpose of marking and discussing information on pupils' progress in internal assessment meetings.

HelenP noted more recent changes in her school's approach to marking that was implemented by the senior leadership team with a view to reducing teacher workload caused by excessive time spent marking without clear evidence that it made a positive impact on the learning of pupils. The focus is now more on picking up early on pupil's misconceptions and reducing the time spent by teachers in marking books. SaraS spoke of the increased focus on preparing students for examinations, not just in terms of a response to the accountability measures, but also for the benefit of students, particularly in the final year of the GCSE course, and sometimes at the expense of; "…good teaching and learning". SaraS also likened this approach to that used in primary schools with regards to preparing pupils for National Curriculum tests. She also welcomed the change to stop the use of National Curriculum levels of attainments.

7.4.14 Initial Teacher Education

SaraS was alone in stating that new entrants to the profession are better prepared in comparison to her own experience. However, she pointed out this is still to some degree variable and dependent on the training provider and the impact of placement schools. By contrast, she was less positive about the level of understanding of new teachers in terms of what she described as the more formal assessments such as written tests. Some of this she suggested was because such expertise is not generally available as there are few people involved in the development of formal assessments.

7.5 Section 3: Interview 3, the function and future of assessment and adequacy of teacher knowledge

7.5.1 Assessment as an aid to teaching and learning

All six participants voiced the opinion that the prime us of educational assessment should be for the benefit of pupils. However, the wide range of uses to which assessment can be applied was acknowledged by all participants. For example JohnS was clear that the ideal

use of educational assessment should be for the benefit of pupils *and* teachers by providing both with:

> …a sense of where they are at different stages. I feel that the purpose of assessment is for teachers to check where children are at different points really and check that they understand what they're doing.

For AngelaS, the ultimate use of assessment was seen as a means of improving the educational experience of children by making; "…sure that they're learning what they should be learning". As a part of this goal, AngelaS encourages her pupils to use peer and self-assessment approaches to involve them in the process. She noted that this approach requires a strong support mechanism to train pupils to use the procedure, particularly in stating positive aspects of their own work, a task they often find difficult due to low self-esteem, and identifying improvements. However, AngelaS spoke of other challenges in involving and monitoring pupils in their use of assessment to aid learning such as the impact on teacher workload, often exacerbated by a high rate of staff absence. She stressed the need for assessment outcomes to be used as instruments to provide feedback to pupils and stressed the need to train pupils to understand and use the feedback to inform their learning. AngelaS has developed a range of proforma to be used by staff and pupils to provide a system of feedback that aids pupil's understanding and reduces workload for staff. She related the approach to her 'value system':

> My kids are trained very well because it's part of my value system so I do it. So this is also to help the staff. Because I'm trying to reduce workload, I'm trying to make what I value to be workable in a classroom, for both pupils and teachers.

7.5.2 Formative and Summative assessment

In discussing the function of educational assessment, all participants frequently used the terms 'formative' and 'summative assessment'. In using these terms, formative assessment was used to describe the process of identifying gaps in pupil's knowledge; this was a recurring theme throughout the interviews. All participants noted that the identification of gaps in knowledge was used to inform future planning and instruction. In support of her view that testing is a useful approach, HelenP noted that commercially produced standardised tests are helpful to teachers, parents and pupils in measuring progress and identifying gaps in teaching and learning. AngelaS was the only participant to refer to drawing on recent professional development in giving her knowledge and

confidence in applying educational assessment as a formative tool. GrahamP and TimP noted the use of summative assessments to measure the progress of pupils. Summative assessment was also recognised as a useful factor to inform teacher's planning. GrahamP noted the use of summative assessment as a way for teachers to reflect on the effectiveness of their instruction and aid future approaches to teaching subject content. There was however an element of dual use of the term 'summative assessment' in that it was put to use as a measure of pupils' progress *and* as a way of informing teaching in both the long and short term, a process that might equally be described as formative assessment.

JohnS expressed the view that across the system there was too much emphasis on assessment at the end of set periods of time rather than more on-going use built in to the programme of teaching. HelenP and TimP also supported the use of assessment as an on-going process and that the outcomes should be used as part of the accountability system rather than reliance on an end of key stage test.

7.5.3 The use of educational assessment data

JohnS reported other issues with the way in which schools use data derived from assessments. He spoke of data being used to drive targets set by senior school leaders and shortcomings in the general understanding of how data work:

> I think there'll be teachers who don't engage with the data, who don't understand data, who just put data into a system and then it churns out something and so therefore they don't understand it. And to be honest – and this is quite controversial, but beyond Maths teachers – I think as you get into other subjects teachers struggle to understand data less and less and what good data looks like. I think teachers just see it as marks and I think teachers see assessment as marks on a test.

HelenP was critical of the way the data from end of key stage tests is used, in that the cohort has already moved on by the time an analysis of the results can be made which provides little opportunity to rectify shortcomings for that group of children. However, she did talk of how the item analysis data made available through the STA can be used to understand strengths and weaknesses in teaching or if one particular cohort was not in-line with previous cohorts.

7.5.4 External tests and school accountability

There was strong recognition across all interviewees that externally set and marked tests provided a measure of performance for those outside of the teaching profession. However, whilst all six participants espoused the benefits of educational assessment to teaching and learning, concerns about the current use of external tests and examinations as measures of school performance were raised. The notion of a nationally set and externally marked assessment used as a school accountability measure was questioned by all participants. HelenP made very clear that she supports the use of externally set tests, but echoed GrahamP's viewpoint in stressing that the tests should be used as but one element to inform teachers' judgements. However, unlike GrahamP, HelenP was of the view that the model used for key stage one statutory assessments has 'gone too far' in the direction of being totally schools based. She also noted that key stage 2 tests are used by secondary schools as predictors of future performance in GCSE examinations, often determining which ability sets children are located on entering secondary education. Although continuing her support for external tests, HelenP raised concern about the over-reliance and consequences of what she described as short tests administered to pupils as young as ten as determinants of future opportunities. This same concern led HelenP to conclude that if there was a possibility of removing the reliance on testing at key stage 2, she would be supportive. JohnS also raised concerns about age appropriate testing but in the context of the secondary phase of education questioning politicians' appetite for such a debate around the pre-occupation with testing at age 16 rather than when pupils are ready to move on to the next phase of their education. TimP was less critical of external tests and their use in the school accountability system noting that it provided him with a comparative measure on which he could judge his own performance.

JohnS and SarahS showed more concern for the structure of tests both favouring a criterion-based system above a normative based system. Both were of the view that if a pupil meets the set criteria for a particular standard, that should be acceptable rather than designing a system to differentiate or rank schools: JohnS described this as reducing educational assessment to a competition.

7.5.5 Teaching to the test

JohnS typified the comments of all other participants in evaluating the impact of accountability measures on the way educational assessment is used across the education system in driving schools to 'teach to the test' rather than a focus on measuring genuine

learning and understanding. None of the participants argued against the use of tests, set either internally or external to schools as measures of educational achievement, and all recognised the value of externally set standardised tests as a means of evaluation at a broader level than a single school. JohnS echoed a view expressed by HelenP that more testing should be used, but at different points rather than end of a phase, a view he believed to be controversial but one that should be subject to more debate. JohnS justified part of his argument on the basis that schools in general test pupils on at least an annual basis, so this would not in his view add to the teacher workload. But of more importance to JohnS, was further consideration of the purpose of educational assessment. But in line with other participants, JohnS was clear that such a testing regime would not be used to formulate national performance tables. For JohnS, the issue raised questions about the way teachers view the purpose of assessment and its relationship with the purpose of schooling:

> I think teachers see assessment as end of year tests and end of term tests. I think assessment should be constant though. I suppose it's what the goal of schooling is. I think it goes back to the question about, "What do you see the point of school being?" Do you see the point of school being that at 16 and 18 they do exams and great, whatever they come out with? Or do you see it as genuinely being giving them the knowledge that they need to succeed?

7.5.6 Internal accountability

For AngelaS, internal or school based assessment was seen as a tool for measuring her own impact as a teacher, which she described as 'self-accountability' and a feature of being a 'good teacher': She also explained how she used her system of monitoring assessments across the faculty as a means of checking important knowledge is delivered to pupils. AngelaS was very clear that assessment provided information that would identify gaps in pupil's knowledge, either through a lack of learning or understanding, or through deficiencies in teaching. However, the end goal of performance in summative assessments was acknowledged:

> I would say the whole purpose of this it's to identify gaps to plug them wherever they're coming from. You know, teaching, learning – it's to identify the gaps and plug them so the kids know more. Because ultimately, the more they know the better they're going to do at their end of assessment. There's no tricks to doing assessment.

7.5.7 Communicating with other schools

The communication of educational assessment outcomes between the main feeder infant school and TimP's junior school was noted as an area in which he had seen some improvement, especially in light of the way in which progress is measured between key stages 1 and 2. However, HelenP pointed to limitations in the process of passing on educational assessment outcomes to the local secondary schools as the focus narrowed to core subjects. As with the exchange of information between TimP's infant/junior schools transition, the motivation behind the transfer of information was driven by progress measures as required by the system of school accountability.

7.5.8 Teachers' assessments and the accountability system

All participants expressed the view that teachers' assessments should have a place in the school accountability system. HelenP's view that tests should be used to inform teacher's assessment was typical:

> I would test. I've not got an issue with testing. But I think the test should be used to inform the teacher's judgments. I think it's another tool.

Notwithstanding this remark, HelenP made clear she could equally support the total removal of tests in favour of a teacher assessment based system provided the approach was universal, that is, applicable to all schools. Despite a consensus that teacher assessment should be included as a measure in the school accountability system, the exact form of involvement of teachers varied between participants. For example TimP and SaraS shared HelenP's view that teachers should be involved in educational assessments used as a part of the accountability system. However, neither believed that any such assessment should be totally school based and that teachers' assessments should be subject to some form of moderation. For TimP, this would be through a moderation model similar to that used in current key stage 2 tests. SaraS spoke of the use of a final externally set and marked module or using national data as a means of aligning teachers' marking of the GCSE.

7.5.9 Increased testing

TimP did expand on the use of moderation stating that all schools should be moderated instead of the current sampling method used at key stage 2. He also suggested that the

number of assessment points should be increased rather than the current focus of end of key stage tests. Such tests would be of a smaller scale and include sampling of teachers' assessments of children's class work. The increase in assessment points was a common theme raised by all participants. However, JohnS was not enthusiastic about the role of coursework as part of formal examinations but did back some element of teacher assessment in the process. However, he shared TimP's view that it would need to be valued as being of lower status than externally set and marked tests. In line with all other participants, JohnS suggested an approach with an increase in low stakes testing, and in line with GrahamP's view, with a focus on improving learning.

SaraS suggested that the marking of tests and school-based assessments should be against a set of performance criteria rather than through the annual setting of grade boundaries. This was in her view to reduce the competition between schools that may challenge the reliability of teacher's assessment. Despite this, and along with TimP, she favoured the retention of school performance tables.

7.5.10 Extending the current regime used for key stage 1 assessment

GrahamP held the view that the current system used for National Curriculum assessments at key stage 1 provides a way forward. However, despite favouring teacher assessment supported by externally set tests, HelenP made clear that there would need to be careful consideration of how best to ensure that teachers were setting and applying common standards. HelenP was ambivalent about whether the system would benefit from adopting the current key stage 1 model where tests inform teacher's assessment judgments, or a key stage 2 model using externally set and marked tests. If a move was to be made to allow teachers to mark externally set tests, HelenP was clear that this would require a programme of training for teachers and a system of moderation to ensure standards were being correctly applied. She also expressed the view that to some degree the current key stage 1 model whereby teachers marked the externally set tests without any moderation; "…had gone a step to far the other way" and on reflection, suggested marking conducted by an external agency reduces workload for teachers and offers more assurance on the rigour of marking quality across the education system.

7.5.11 Assessment as a means of improving teaching and learning

GrahamP presented his strongly held view that assessments should be focused directly on improving teaching and learning rather than as part of the system of school

accountability. He was also clear that the teacher's overall judgment, though informed by externally set tests, should over-ride test results should teachers view that as being appropriate, a viewpoint supported by HelenP.   He cited examples of pupils over-performing as well as under-performing in one-off tests as a justification for teachers to over-ride test results in the interest of validity. AngelaS expressed a similar view in relation to what she described as the 'fairness' of examination results noting that pupils may receive GCSE examination results that in the teacher's view are not fully deserved. AngelaS's argument was based on the view that teachers can make their assessments on the performance of pupils over a longer period of time. However, she concluded that where the results were advantageous to schools with regards to school accountability measures, teachers are unlikely to complain. HelenP supported the view that teachers are in a better place to gauge the performance of children over time rather than through a one-off test at the end of a key stage. In practice, her view was that the current system objectifies children as; "a number or a commodity". HelenP's view was compounded by the Government's announcement in 2018 that the collection of teachers' assessments was to be removed from the statutory assessment arrangements, justifying the move as a response to concerns raised by teachers over excessive workload. She described this as a total disregard of teachers' assessment stating that it undermined the whole concept of teacher assessment. For AngelaS, the utilisation of teacher's assessments should inform a more localised form of accountability forming part of a system she described as self-accountability.

7.5.12 Trust in teachers' assessments

All participants made further references to the issue of trust in regard to teacher assessment and the use of tests. SaraS expressed the view that commercial or national produced standardised tests offer more assurance of their quality compared with those produced by class teachers. She noted this was particularly relevant since the removal of National Curriculum levels expressing the view that; "…in this current climate I'd be a bit dubious because I wouldn't feel certain that teacher knows what is age expected for year eight". However, she explained this as a training issue rather than any deviousness in teachers' actions.

AngelaS expressed the view that teachers do not deliberately set out to cheat and that she would support more trust in their assessments because their daily contact with pupils makes them best placed to make a more comprehensive judgment of their performance. For TimP, trust in teachers' assessment should be assigned to teachers with several years'

experience in the classroom – to the extent that it should not be questioned. TimP balanced this view referencing the pressure on teachers and schools to 'look good' and he aligned with the view expressed by HelenP that there is a need for a system of external checks to verify the application of procedures and standards of performance. GrahamP also expressed an ambition for greater trust in teachers' assessment:

> I want to make sure that teachers are the ones that are trusted; that teachers' practice around assessment is such that they feel confident and that they're empowered by summative assessment, which is produced thoroughly and produced well, whether that's from central government or wherever else it comes from.

However, for JohnS, any form of teacher assessment must be kept 'low stakes' because of concerns over the trustworthiness of their judgments.

7.5.13 School accountability - impact on schools

GrahamP expressed the view that schools are in effect at what he described as the "receiving end" of the accountability system, a system of which he has "issues". For example, in his view the accountability system ignores the impact on attainment caused by the age range of children within a designated school year – often referred to as the 'summer born' children issue that describes concerns that children born towards the end of the school year enter school at a younger age than some of their peers causing an adverse impact on attainment as measured by accountability measures. HelenP took the view that primary school teachers are more likely to feel a connection with their pupils stating that they have more time invested in the children. This provides primary teachers with a broad understanding of their pupils, but this is over-looked by an accountability system with a focus on a single test at the end of year six. This regime results in schools focusing on test readiness. The narrow focus on testing overlooked what HelenP described as broader measures of progress demonstrated by children such as in their confidence or willingness to participate in a class discussion, none of which is recognised by testing or by Ofsted inspection, which again HelenP characterised as an inspection of data than evaluating teaching and learning. JohnS expressed a similar view to that of HelenP's but from the perspective of a secondary teacher:

I just think education just seems at the moment, particularly when you get to exam years more about an exam factory rather than actually about what I'd call just enjoying school and education.

7.5.14 The impact of school accountability on teachers' practice

For TimP and SaraS, the system of external testing provides a measure of a school's performance that can be compared across the education system. TimP found this of benefit to those he described as 'outsiders' and as a year six teacher, despite the added pressure, it provides him with a target. GrahamP also described himself as being 'competitive', however, he reflected that despite the impact of the accountability system, schools still hold a tendency to share their practice, but concluded that the negative effect of that is that schools endeavour to find and replicate so called 'good practice' which to GrahamP can be at the expense of innovation:

> And that does worry me. I think what the accountability system is driving down is innovation within schools and schools daring to be different.

SaraS stated that competition in itself is 'not a bad thing' but for her the competitive goal should be to raise standards for all through co-operation. HelenP raised concern about how accountability measures relay pressure from one school to the next and on children as pupils move from primary to the secondary phase of education. JohnS partially agreed with HelenP's viewpoint about the pressure on pupils, but was of the opinion that teachers are under the greater pressure.

7.5.15 The impact of school accountability on learners

TimP described his pupils as being used to being tested, but that they were aware of the difference between high and low stakes uses of the outcomes, and the way pressure builds when children have the time to think about upcoming assessments. This resulted, in his view, that testing should be more frequent and delivered in a less formal way, what he described as being administered 'nonchalantly' as part of a lesson. TimP further described how the perceived importance of key stage 2 tests leads to pressure on children emanating from parents and teachers. Noting that not all parents exert such pressure, he acknowledged the pressure simply came from parents wanting their children to do well and from children themselves who equally want to do well. However for teachers, he described how the pressure built because; "…there's a lot riding on the outcomes". HelenP

took a broader view on children's reactions to testing and schooling which in her view was a result of more complex issues citing the possible reaction to the way schools deliver a pressurised and more demanding curriculum that "turns children off learning". She expressed the view that changing the way children are assessed, such as through a combination of teacher and externally set assessments would not change this as a key fault of the current accountability system is the focus on academic subjects stating: "This is the other thing, it doesn't seem to be acceptable to not be academic in the current system". HelenP described this as having an impact on some children throughout their time in schools. Of particular concern to HelenP was that the publication of results as seen in the system of accountability and that telling a child that they have failed to meet a standard is more punitive than it is helpful. She included the impact of Ofsted inspections, also an element of the school accountability system, having a similar impact on the pressure placed on teachers and children.

7.5.16 The impact of school accountability on teacher workload

GrahamP raised concern over the workload implications of assessment stating that in reality it should be an integral part of the teaching and learning process. However, over time he has witnessed a growth in the collection of assessment data for accountability purposes that was far less evident ten years ago. AngelaS in part adopted a strategy of involving pupils in the assessment process, partly to aid their understanding, but also as a mechanism to reduce workload on teaching staff. HelenP presented a view that changes announced by Government to stop the collection of data on teachers' assessments at key stage 2 other than that for English to reduce teacher workload was a ploy to undermine its importance. However she acknowledged that at least one advantage of externally set and marked tests was that it saved teachers from having to do it, thus keeping workload down. JohnS looked to a future where artificial intelligence could play a significant role in assessment and lead to an increase in low stakes testing and a reduction in teacher workload. However, he presented a view that teachers often react negatively to what might be perceived as an increase in assessment not recognising that they already use assessment extensively in their day-to-day practice.

7.5.17 The expected future of educational assessment in England

SaraS raised an issue with the appropriateness of GCSE examinations for pupils who are unlikely to go on to study GCE A levels. Her comments were more focused on the content of GCSEs rather than the qualification itself. She expressed the view that the content of

GCSE mathematics should be geared to more practical aspects, real life situations for example securing mortgages. SaraS was clear that for those students going on to study A levels, the current relationship between GCSE and GCE A level works well. She was also supportive of retaining the GCSE qualification, but would not support an exclusive use of coursework and would favour a modular approach with examinations set over three tiers to provide better access for students with differing abilities. SaraS was equally clear that a more practical alternative to the GCSE would not and should not be presented as an 'easier' option; rather just different in terms of content. It could also be marked by teachers in combination with an externally marked element that would act as a comparator and external check on teacher marked components. However, she would prefer a system that used more low stakes testing used by teachers as a means of aiding the memory of pupils. Such use, she explained, should be made clear to pupils so that they are aware of the purpose of assessment.

TimP expressed a view that any further significant change was unlikely but any change should be demonstrably beneficial. He acknowledged the challenges faced by national agencies in writing valid tests, but given his way he would like to see some consideration of practical assessments, for example in science, design and technology and physical education as a way of recognising the broader and different attributes children possess. Along with SaraS, TimP was critical of the speed and frequency of change stating that this leads to faults in the design of assessments. Acknowledging that the current system is not without fault, on balance TimP preferred a period of stability in the way assessments are designed and administered. However, he did note that the content of the curriculum should be more dynamic in order to keep the content up to date; this TimP suggested should be supported by training for teachers. GrahamP made a link between developments in assessment practice and teacher retention issues, particularly in reference to the growing number of experienced teachers leaving the profession, a phenomenon he referred to as; "haemorrhaging expertise". He proffered that the loss of expertise in local authorities and schools will result in wide variation in educational assessment practice. This, he suggested would be exacerbated by competition between multi-academy trusts that will result in a reluctance to share knowledge and best practice. JohnS suggested a growing opportunity for the use of artificial intelligence geared to the needs of individual pupils, through for example, adaptive tests and quizzes – but not replacing teachers. However, he concluded that there is an element of negativity from teachers around such use of technology, seeing it as a threat to their role. JohnS presented a very clear view that the technology would support rather than replace the teacher by reducing teacher workload and producing a more focused view on the learning needs of individual pupils.

7.5.17.1 The approach of government to educational assessment

SaraS supported the current Government's production and use of performance tables stating it provided parents with useful information. She also expressed the view that teachers in general follow rather than challenge government instigated change stating that she has never been involved in discussions with other teachers about the possibilities of a future assessment and accountability system. SaraS proposed that the energy of senior leaders was more focused on how to meet the expectations of the accountability system rather than how it could differ. TimP raised concern over the Government's planned introduction of a baseline assessment in 2020, although he noted this was dependent on how the outcomes would be used and the implications of assessing children at such a young age.

GrahamP raised concern over the lack of exemplification available to teachers from the STA stating there is a 'desperate need' to exemplify expected standards of performance: this, he suggested, should be supported by materials produced by schools to ensure standards are understood and applied consistently. For AngelaS, wider political considerations were of little interest expressing disappointment in politicians of all party persuasions stating that her focus should be on areas where she can effect change - in her classroom. JohnS expressed an equally despondent point of view:

> I don't think it's working at the moment; I don't think the system of training teachers is working, it's too all over the place it's too confusing. So I think the whole education system at the moment to me feels a bit disjointed. It needs to take some bold people to make some bold decisions about the future.

7.5.17.2 Professionalisation of assessment

TimP and GrahamP supported a system of teachers accredited for their expertise in educational assessment, such as Chartered Educational Assessors, who could support a system less dependent on external moderators. However, TimP tempered this by stating that such a system would require extensive training leading to financial implications at a time when schools lack funding for professional development and that local authority support had also reduced. TimP suggested that a cascade system of training could be initiated whereby one teacher receives external accreditation and then passes on their knowledge to teachers within their own school. GrahamP noted that access to such

expertise would have broader benefits to the curriculum and pedagogy, especially in relation to his view that newly qualified teachers are lacking in their knowledge of these areas. For GrahamP, the role of Chartered Educational Assessor would also provide a pathway for career development, and possibly reduce the loss of expertise from the teaching profession.

7.5.17.3 Teachers' knowledge and expertise in educational assessment

SaraS expressed the view that on the whole, teacher assessment knowledge within her own school was "pretty good". HelenP presented a similar view of her own school but noted that when she attended meetings with other assessment co-ordinators in the local authority their range of approaches raised an element of doubt in her own system. However, on reflection, she was confident in the veracity of her own approach. SaraS also contrasted practice in her own school with what she found in her work with other schools: this she noted was very evident in teachers' use of questioning used in lessons where questions were pitched at too high a demand or where teachers failed to probe pupil's answers. JohnS expressed a similar viewpoint noting that although teachers are able to talk about assessment for learning, he believed they were in effect going through the motions and questioned their understanding and effective use of the approach:

> I think teachers will be able to talk about that. Whether or not it's effective is another question. I see very, very little effective assessment in a classroom. I don't think it's very, very, strong really.

TimP was clear that he and other teachers in his school need more support around assessment expertise, especially across the range of subjects taught in the primary setting. He noted his confidence in teachers' capability to develop their educational assessment knowledge and skills, but felt this was contingent on having access to guidance from experts. GrahamP supported the notion of training, for example in terms of moderation, and believed that this would add a professional pathway for teachers. As with TimP, GrahamP noted that teachers need specified time to develop their expertise, noting being shocked by the lack of expertise and confidence shown by senior leaders in his own school regarding educational assessment. AngelaS was equally unequivocal about the lack of expertise in her own school noting that she had to teach her faculty the difference between standardisation and moderation.

7.5.17.4 Access to professional development

HelenP talked about how new approaches to assessment, or new administrative tasks around assessment are not supported by training. SaraS's experience contrasted with that of HelenP noting that CPD was well supported in her own school with access to external courses and weekly in-house sessions at whole school or department level. However drawing on her work with other schools, she reflected that such an approach was not widespread. HelenP noted further that a lack of certainty that the system or any part of it will not be changed creates a lack of security. This feeling was exacerbated for HelenP by frequent changes of government, government policy or even the appointment of a new Secretary of State for education, with questionable knowledge of the education system. TimP had difficulty recalling any CPD outside of the school environment citing the lack of money to support training and the decline in provision from the local education authority – which he also connected with cuts in funding: this has resulted in reliance on internally sourced training in TimP's school, although he did express full confidence in the assessment leader's expertise. GrahamP also noted how in-house CPD is now the norm and how assessment is built in as a common feature. AngelaS talked about how her self-funded Master's course aided her understanding and confidence in her knowledge of educational assessment. She felt that the school would support her attendance at an external course if she pressed for it, but noted she was attending a weekend course in the near future with a colleague – but both were self-funding the course. AngelaS's view was that there is "real thirst" amongst staff for professional development but concluded:

> …at the end of the day we're falling short because we're not being developed as staff but it's not through the headteachers' fault – there's no money.

All participants repeatedly mentioned the lack of money in the system as the real barrier to CPD.

7.5.17.5 Use of nationally available resources

HelenP described how she used the item analysis facility provided by the STA to engage staff in discussions about the wording of test questions and the performance of their pupils on particular questions as a way of engaging staff with educational assessment practice and familiarity with the way tests are constructed. HelenP also provided a talk to other local assessment leaders on her school's assessment system post the removal of levels. However, this was not reciprocated by other assessment leads. In reference to the

Government's decision to remove attainment levels, HelenP felt that the lack of support for schools in developing their own systems resulted in additional workload repeated across the entire education system; she described this as a "complete waste of time" that could have been avoided by the provision of centrally produced support and guidance. HelenP recalled how support materials were at one time produced by the Qualifications and Curriculum Authority lamenting its demise and how this moved the onus onto schools to develop their own approaches. JohnS doubted the capacity of schools in general to provide adequate CPD and questioned if there was any real incentive so to do.

7.5.17.6 What do teachers really need to know about assessment on entry into the profession?

In terms of initial teacher preparation, SaraS, TimP and HelenP were very clear that the purposes and use of testing should be a strong feature of ITT noting that assessment was undervalued, especially in the use of standardised tests. For TimP, understanding the key stage 2 testing regime and expected standards of performance would have been beneficial at the start of his teaching career. He added that this should be the case for all new teachers and not those working in the final year of the primary phase of education stating that teachers working across all year groups should take an equal interest in educational assessment. SaraS and HelenP would also like to see less of a focus on the generation of data stating most lacked any real use and placed a burden on teachers through the time spent entering data into school tracking systems. Spending more time on understanding the purposes and uses of assessment was also recommended by TimP and that assessment was integral to teaching and learning rather being an "add-on". TimP was not critical of the amount of time spent on assessment in the current education system as it provided teachers and parents with insight into educational performance and pupil's needs. He also recommended that new teachers should spend time on familiarisation with the approach to assessment taken by their prospective school. TimP suggested this aspect in view of the different systems now used by schools since the removal of National Curriculum levels. However, he noted that in essence, schools were dealing with broadly the same curriculum and aiming for the same standards of performance, so common features could be the focus of training. This he argued should be supported by more practice in using, setting and evaluating assessments. Recalling some of his own experience as a newly qualified teacher, he found discussing children's work with teachers of great help in developing his own understanding of performance standards.

Assessment as an aid to planning was also stressed by TimP using outcomes to modify planning in line with assessment outcomes, be they formal tests, day-to-day marking or from talking to children as lessons are in progress. TimP noted this was a key learning point in his first year of teaching. For GrahamP, developing a personal philosophy around educational assessment was of central importance. AngelaS presented a different view to GrahamP's stating that in a recent interview situation, she didn't even consider asking any questions on assessment as appointees would be expected to follow the school's model. However, for HelenP, assessment knowledge and understanding are essential attributes. JohnS felt there was a significant gap in teachers' knowledge about how children learn and working memory stating this was a significant gap in ITT programmes and of central importance to an understanding of the educational assessment process.

7.5.17.7 CPD in early career development

TimP was of the view that newly appointed teachers need support in understanding the assessment approach and system used by the school. Given that schools differ in their approach, this becomes of paramount importance. GrahamP spoke of the benefits of experiencing and understanding statutory assessment used in year 2 and 6, if not in their first year of teaching within the first four or five years. For those he described as not being "lucky" enough to gain such experience early in their careers, knowledge of how summative assessment fits into the work of the school is essential. JohnS also noted the importance of the first five years of teaching stressing the need for CPD and practice noting Gladwell's 10,000 hours theory (see Gladwell, 2009). JohnS expanded on this view stating that the concept should in reality mirror what he described as an apprenticeship for teachers with years of service aligned to pay incentives. This should be supported by structured CPD for which schools should have a statutory obligation to provide both in-house and externally sourced resources. Additionally, teachers should be obliged to visit other schools including opportunities for secondments to other institutions and involvement in research projects.

7.5.17.8 On-going professional development of teachers

AngelaS expressed the view that the school's staff appraisal system should be closely linked to a programme of professional development to build on strengths and address areas of weakness in staff expertise. However, she lamented the lack of opportunity due to a lack of financial support. JohnS stressed the need for CPD throughout a teacher's a career, although needs would change as the teacher developed expertise. He also talked

about the advantages of developing Master Teachers as developed in Singapore and Shanghai who are recognised for their expertise and their role in supporting other staff. Such a system also provides a career route for some teachers. He noted the recent formation of the Chartered College of Teaching (CTC) and Chartered Teachers, but posited the view that the incentives are unclear and that there is a lack of information about the project. JohnS held the view that teachers are more familiar with previous initiatives to raise the status of teaching, for example Advanced Skills Status (AST) and Specialist Leaders of Education (SLEs) noting that the objectives of the CCT are unclear. He questioned the incentive for SLEs to work with other teachers to improve their practice, but did support the notion of developing Master Teachers who are focused on supporting other teachers. However, he was clear that this required a profession with teachers continuously striving to improve their expertise and skills.

7. 6 Chapter 7 summary

Participants recalled how their own experiences of assessment up to the end of key stage 3 contrasted to that found in later stages of their education and more particularly with their experiences as teachers in current times. Their earlier experiences were characterised as periods of less pressure on pupils and teachers and where references to pressure were made during their time studying for examinations, respondents spoke of their self-imposed pressure to do well. Three of the six participants reported how their own experiences of assessment as students influenced their current practice as teachers. Participants presented an unequivocal view that the amount and importance of assessment, particularly as a measure of accountability, had changed significantly over their time in education resulting in a growth of pressure on pupils, teachers and schools. However, all six participants recalled few strong memories relating to assessment theory and practice in their ITT.

Although participants did not present a universal definition of educational assessment, all six participants were of the view that the use of assessment should be for the benefit of pupils. Five of the six participants stated that assessment in reality is more driven by accountability measures and the desire to produce performance data rather than teaching and learning. Despite acceptance of the need for schools to be held accountable, there were mixed views on the inclusion of teacher assessment in high-stakes tests and examinations with all six participants raising concern over the lack of trust in teacher based assessments, in the main driven by measures relating to accountability. However, all participants spoke positively about tests, but raised concerns over the uses to which the

outcomes were applied. On the whole, participants expressed more trust in their immediate colleagues than those outside of their own institutions.

Chapter 8: Interviews with Key Influencers: data analysis

8.1 Introduction

This chapter presents the data gathered from interviews with Key Influencers working within or directly having influence on the education system in England. (See Section 3.11 for an explanation of the approach and selection of Key Influencers).

8.2 Participants

The pseudonyms, occupation and roles of the Key Influencer participants are provided in Table 8.1 below. More details can be found in Chapter 3, Table 3.4.

Table 8.1: List and roles of Key Influencer interview participants

Name and Reference	Occupation & Relevant roles	Name and Reference	Occupation & Relevant roles
Ali **AliA**	Policy advice and evaluation. Academic	Ash **AshA**	Policy advice and evaluation. Academic.
Pat **PatA**	Policy advice and evaluation. Academic.	Terry **TerryA**	Policy advice and evaluation. Academic.
Sam **SamCS**	Civil Servant. Standards and testing.	Hilary **HilaryLA**	Local Education Authority assessment lead adviser
Jo **JoPol**	Politician. Government education policy.	Logan **LoganPol**	Politician. Government education policy.
Lou **LouPol**	Politician. Government education policy.	Mel **MelPol**	Politician. Opposition party education policy.
Amar **AmarPA**	Professional association executive. Academic.	Hayden **HaydenPA**	Policy advice professional association executive. Academic.
Chris **ChrisR**	Regulatory Executive. School inspection.	Jaden **JadenR**	Regulatory Executive. Examination and test regulation.

8.3 Descriptions of educational assessment: process, purpose and practice

There was no universal definition in the participants' descriptions of educational assessment. However, a common attribute of the various descriptions was that assessment involves gathering information from which inferences can be drawn. Such information or evidence was generally related to inferences around what an individual has

attained, achieved, knows or can do. Further, participants' descriptions of educational assessment fell into two broad categories: as a tool to support teaching and learning or as a means of measuring performance. As a tool to support teaching and learning, the terms assessment for learning and formative assessment were used interchangeably. All participants referred to both general categories, to support teaching and learning or as a means of measuring performance (see Section 8.3.1) in their descriptions of educational assessment. However the degree to which they emphasised each of these two broad categories tended to reflect their professional roles in relation to the education system in England. For example the key purpose of educational assessment for ChrisR, a school inspector, HaydenPA, a professional association senior executive and HilaryLA a local authority assessment adviser, centred on informing teaching and learning.

Academics tended to take a broader view in describing assessment as either summative or formative thus relating assessment not only to a process but also by the purposes ascribed to the outcomes. For example PatA noted:

> To follow Lee Cronbach, (assessment) is just a procedure for drawing inferences. We give people stuff to do, they do it, we look at what they do with the task and we draw conclusions. So following that, an educational assessment is a procedure that allows us to draw conclusions about educational processes. The question is, how justified are you in drawing the conclusions you want to draw, given the evidence you've got?

Politicians' initial descriptions of educational assessment tended to lean towards high-volume tests and examinations although this was not without reference to more localised assessments made by teachers within a school setting. In describing educational assessment as a tool to be used for tracking progress either as an on-going process or at a given end point, JoPol expressed the view that assessment is:

> …probably one of the best things we've learnt over the last twenty years. So we didn't have the narrative, that story of each individual child and what they were doing. So I think teachers' ability to monitor the progress of students as they go on, I think is excellent.

However, JoPol balanced this view by stating that: "Equally it's the worst thing they've ever done over the last twenty to thirty years. Because it fulfils too many purposes". The idea that educational assessment was more generally viewed as a process of testing and

examinations was noted by most of the participants. This view was exemplified by TerryA who stated that:

> Assessment typically in England is associated with assessment which occurs at the end of a period of learning. And we focus particularly on the idea of the provision of certification, the idea of a large formal external assessment that occurs at the end of a long period of learning, which enables people to understand and recognise what a particular person has achieved.

JadenR reinforced this viewpoint suggesting that teachers are heavily focused on qualification outcomes noting that: "Qualifications over-dominate education. The level of hype around it all, the extent to which they determine what is taught and when; the extent to which they determine assessment practice". However, JadenR explained a possible duality of purpose amongst teacher's intentions; on the one hand to open up future prospects for their students whilst on the other hand motivated by performance measures focused on their respective school.

8.3.1 Measuring performance across the education system

The use of educational assessment as a means of generating performance measures going beyond that of individual pupils was a feature discussed by all participants, with particular reference to accountability measures. For example, TerryA noted that the strong relationship between assessment and the accountability of schools influenced the behaviour of schools in particular to optimise school performance more often through the provision of a curriculum; "…that is collapsed into qualifications".

MelPol noted the influence of government policy on the structure of ITT as a cause of assessment being de-valued and has also undermined teachers' agency in the process as the system has moved towards greater accountability:

> …the government…they've been critical of our teaching universities and therefore assessment has gone from being something that was intrinsically valued by a teacher, (who) had agency to undertake it…to a complete accountability session now. So all the money's gone from supporting schools, improving curriculum and assessment to just accounting, to being pen pushers, accountants in terms of how we assess children, and teachers.

Also speaking as a politician, JoPol put this down to the multiple purposes to which educational assessment now serves, purposes for which it was not originally intended and that now:

> …rather than assessment being seen as a tool to help children, I think it's seen as a fairly heavy handed check on teacher and school performance, and I think that's a problem.

AmarPA referred to Newton's (2007) twenty-four purposes of assessment of public assessments stating that these; "…24 different purposes of public assessment are very complex – some really useful, some less so". TerryA summarised the aspects to which educational assessment is associated in the English education system into three key areas; the development of formative assessment along with; "…the ineluctable association of assessment with accountability. And then the over determination of curriculum by formal examinations". AshA described the focus of assessment on tests and examinations as a feature of English culture and in a way that was detrimental to a broader understanding of how educational assessment works.

MelPol referred to assessment as; "being central to pedagogy", mirrored by JadenR who described assessment as a key feature of teaching and as part of the human condition:

> Because from the day one that they (teachers) walk into a classroom, what are they doing? They're assessing…we're all assessing all the time. That's what we do as human beings.

AliA was the only participant to challenge the concept of educational assessment stating that:

> Nowadays I tend to like to call it measurement because I think there's something scientific and objective about what we do, at least that's the aspiration.

8.4 Teachers' understanding of educational assessment

There was a level of scepticism over the quality of teachers' understanding of educational assessment amongst all the Key Influencer participants. PatA summarised the state of assessment in schools in England as being "pretty poor" suggesting that teacher's assessments may fall short of the scientific or objective aspirations required of AliA's more

focused description of measurement. HaydenPA posited the view that despite the importance of assessment to teaching and learning, teachers' knowledge and understanding of assessments may not be widespread and expanded this view to include those who judge schools on the basis of assessment data:

> So assessment's important, knowledge of assessment is important, and at the minute it's pretty scarce; and the people that come in and judge the schools as well have pretty scarce knowledge.

There was a consensus amongst the Key Influencer participants that despite the centrality of formative assessment or assessment for learning in the teaching and learning process, the approach is at best not well understood by teachers. TerryA was less direct than PatA and HaydenPA in criticising the extent and quality of teachers' assessment, but stated that the phrase 'assessment for learning' was interpreted in different ways leading to a level of ambiguity around the term. And TerryA concluded further that the concept has moved from its original sense, often having less impact on teaching and learning:

> The general phrase which has become established as assessment for learning…doesn't mean any one thing and it doesn't always relate straight back to the work which was done – the brilliant work which was done by Black and Wiliam…because it's mutated into many, many different things. And of course many people claim they're doing something which is allied to Wiliams and Wynne Harlen and so on, but are doing something very different, sometimes something which is defined as formative assessment but is much more impoverished in terms of its impact on learning.

Other academics in particular supported this view with AliA and PatA lamenting that assessment in English schools has become largely focussed on monitoring or tracking learning rather than on formative assessment strategies. Some of the underlying reasons for this were related to initial teacher preparation and opportunities for continuous professional development and these are picked up in later sections of this chapter. However, other reasons were proffered. For example, the lack of teachers' ability to fully deploy a formative assessment approach was raised by LoganPol citing the impact of National Curriculum levels of attainment and the pre-occupation amongst teachers of measuring progress towards SATs and GCSEs as a contributory factor in the de-skilling of teachers in the use of assessment as a teaching and learning tool. However, the

possible reasons for the focus on tests and examinations were summarised by AmarPA as the increased involvement of politicians and the rise of accountability measures:

> When I first started teachers determined the curriculum. They determined the assessment system, they determined how it was to be assessed and they determined what the outputs were, what purposes the outputs were put to. From 1976 when Jim Callaghan made the Secret Garden speech politicians have become more and more involved. Eventually, what happened was that a National Curriculum was developed in the mid 80s. The exam system was changed in the mid 80s and there was more involvement from politicians from the mid 80s onwards. And the accountability measures have grown and grown and grown.

This viewpoint echoed JadenR's analysis of the over-domination of qualifications in determining what is taught, and when and how public tests and examinations determine assessment practice. Whilst qualifications were cited as over-dominating the secondary phase of education, the role of statutory testing was noted as having a similar effect for the primary phase of education. The tension between the use of assessment to improve teaching and learning and the over-focus on National Curriculum tests was a common feature raised by participants. Speaking from the perspective of a civil servant with responsibility for guiding Government policy, SamCS stated:

> …when I think more widely about it…the powerful kind of role that assessment can play is actually informing classroom practice and in driving teaching and learning and improving pedagogy. And for me that's actually where the argument should sit, however, I also recognise that because of the way in which the system is set out… it's got a tectonic nature hasn't it statutory assessment?

The 'over-domination' of qualifications and tests was constantly related to the influence of the accountability system in England throughout all of the interviews with Key Influencers; this is discussed in the next section.

8.5 Educational assessment and accountability: the general state of educational assessment in English schools at the current time

Participants consistently referred to the influence of the accountability system on the way in which educational assessment is perceived and practiced in schools. The key accountability measures most frequently referenced in this context were set around the use of tests and examinations to compare the relative performance of schools and the use of test and examination data, particularly related to school inspections. Indeed AliA's analysis concluded that all discussions related to assessment design and what's happening in schools are confined within the context of performance tables and accountability. JadenR also stated that the pressure brought about by the nature of accountability has resulted in qualifications being specifically designed to meet the demands of the accountability system stating:

> …because of the nature of accountability, we have to make decisions around the design of qualifications that means they are designed in a particular way to take into account those pressures.

Even though JadenR proffered that assessment design has been curtailed to reflect the requirements of the accountability system, AliA suggested that in essence the GCSE and GCE A level qualifications have changed very little over time and broadly serve the same purpose for which they were introduced; and that National Curriculum tests have changed very little over the last decade. However. AliA concluded that planned changes to vocational qualifications are likely to lead to fundamental change:

> But if you look at the vocational world that's a whole different kettle of fish because there are some very, very, very big reforms going on which are actually trying to change the shape of assessment in really big structural ways. And they were in relation to the Wolf report and the Whitehead report and the Richard report. They were all basically saying there was something broken about vocational assessment that needs a radical approach to solving it.

However, all participants noted that accountability is a necessary and desirable facet of a state education system, for example JoPol stated:

> Oh, it's a good thing, accountability, absolutely. Yeah. I don't think it's easy. And I think if you look at schools now, they are so shaped by the accountability framework.

LoganPol supported the idea of accountability both as a means of holding schools and government to account through a process of national testing of all children and achievement in general qualifications. LoganPol enforced the need to assess all children in all state schools rather than say using a system of sampling in order to provide more granular data at school level:

> But it (sampling) doesn't help the government to find out where the weaknesses are in the system. We want to know which schools are letting down the children, that's what you need to know.

JadenR spoke of the need for accountability measures but questioned the over-reliance by politicians on qualifications data and drawing on background experience as a researcher stated:

> I have quantitative leanings. So I'm very aware of the limitations of using outcomes and qualifications for school accountability places because of the noise in the data. And it's actually really not helpful for parents to be using that for their own personal decision-making, so I don't buy a lot of that. Having said all of that, I understand why politicians want measures of successful schools and how we need to be able to focus on where schools are doing well and where they're failing.

A similar view was aired by SamCS who explained a growing personal appreciation of the positive impact of statutory tests as accountability measures, especially in identifying areas of failure, noting that dealing with any resistance to national testing as an accountability measure is "a price worth paying". But despite supporting a system of accountability, participants with roles outside of government policy raised questions over the efficacy of the use of test and examination assessment outcomes alone in determining the quality of school performance and their impact on the day-to-day practice in schools. This was often made manifest in what could be described as the unintended consequences of the current system of accountability. For example HaydenPA stated that:

> I'm not against assessing per se, I am against anything that ranks people against others, and I think the, that's the kind of consequence of where we are. I mean the

idea that we look at our children and value their worth in terms of their potential performance later on in tests, in certain narrow aspects of learning, and your worth is determined by that, it is horrific. But it leads you down that path very easily, particularly if the school is then going to be held accountable for the performance of the children.

Participants with a political role also acknowledged the unintended consequences of the accountability system. For example JoPol noted:

I think the problem we've got is the consequence of the assessment and accountability where we've got hung up.   So if the threat is of sacking, of closing, losing sponsorship of the academy, or bad headlines, until we take that out of the accountability framework, I think we're stuck.

SamCS was of the view that using any assessment outcome as an element of an accountability measure will inevitably have an impact on how that assessment is used, stressing that accountability is not the only factor that distorts school practice. However, there was a general consensus that tests and examinations as key measures in the school accountability system had a disproportionate impact on the day-to-day running of schools.

8.6 Narrowing of the curriculum

There was wide recognition that the design of the school accountability system generates a focus on national tests and qualifications often to the detriment of broader educational goals. Of these concerns, narrowing of the curriculum was most cited. For example JoPol noted how what is measured influences perceptions of value, pointing to the 'relegation' of the arts and physical education as an example in terms of curriculum time and weighting in accountability measures. Academic participants also commonly stated that the narrowing of the curriculum was a consequence of accountability measures, for example AliA who noted that using the outcomes of tests and examinations; "…is a very narrow assessment of quality. And when it's misinterpreted then that's problematic".

Whilst acknowledging that schools undoubtedly react to accountability measures TerryA stressed that there is a range of possible reactions open to schools so that narrowing of the curriculum is not inevitable. According to TerryA, the accountability system is built on a supposition that schools will react in a rational way to a set of measures but pointed out variance in what might be described as rational. TerryA gave the example of some schools

using 'almost panic' multiple candidate entries for the English Baccalaureate (EB - an accountability measure across a predetermined group of GCSE subjects) as a way of maximising their attainment in this measure describing this as what some schools think is a rational response. However, according to TerryA, this was not true of all schools:

> And then other schools, which I support far more, are adopting a very rational approach to entry for the EB measure. And taking very deliberate steps to ensure that they still deliver a very broad and balanced curriculum, at the same time optimising what they achieve as it's viewed through the EB measure.

TerryA provided a further example of the unintended consequences of an accountability measure for schools that focused on the number of pupils achieving five GCSEs within the A* to D range of grades. In this example, TerryA articulated a level of frustration that such utilisation of assessment outcomes as a measure of accountability was detrimental to both high and low achieving pupils in that high achieving pupils were not 'stretched' in their learning, whilst pupils likely to fail the measure were largely ignored. TerryA referred to this as having; "…very educationally undesirable consequences" and expressed the view that the Government now has a more analytic and critical perspective on how accountability measures can lead to undesirable practice, noting that:

> If some schools can devise a really, really good curriculum response to these measures and the measures driven by assessment are nonetheless a good measure of the performance in schools, then what seems to be a failing is not the emphasis on assessment itself but on a failure to disseminate models of good practise at school level.

JoPol raised concern about how schools modify their curriculum offer in order to maximise performance in the Progress 8 accountability measure[58] through starting GCSE courses earlier which has resulted in some pupils dropping some subjects after only two years of study at secondary level. ChrisR explained the importance of tests and examinations as a

---

[58] Progress 8. Progress 8 (replaced the 5 A*-C accountability measure and) was introduced in 2016 as the headline indicator of school performance determining the floor standard. It aims to capture the progress that pupils in a school makes from the end of primary school to the end of key stage 4 (DfE, March 2019).

means of assessing learning but agreed with the view that it presents a narrowing effect on the curriculum, but differentiated the impact on key stage 2 to that of key stage 4:

> I do think though, the use of it (assessment) at certain points; for example at the end of Key Stage 2, because we are narrowing the domain so specifically to two things or to three things, reading, writing and maths – and then narrowing it further still within that because any assessment naturally will narrow the domain that you're looking at, that that drives certain behaviours. It isn't as stark in Key Stage 4 because the curriculum that is being tested is naturally wider.

8.7 A balanced approach to school accountability

TerryA spoke of the need to balance the possible approaches to accountability stating that policy makers had neglected to fully recognise the possibilities of a combined approach using assessment outcomes and school inspections by Ofsted noting a level of naivety as a result of the fact that; "…we're one of the countries doing it for the first time". The view that there should be a broader framework for holding schools to account other than a narrow focus on assessment outcomes was also expressed by AshA. However, AshA spoke of the seductive nature provided by the simplicity of using test results:

> …because it's simple, it comes up with numbers, you compare numbers and the like so it's very seductive testing that way as a measure of how a school is doing.

TerryA was also of the view that in general, the Government has made improvements to how it determines and uses accountability measures, a view echoed by JadenR, who referred to the broader measure of Progress 8 and using the first examination entry as a means of reducing multiple-entries as a means of gaming the system.

AshA and JadenR also expressed a level of optimism in Her Majesty's Chief Inspector of schools (HMCI) recently stated direction to Ofsted inspectors to reduce their focus on assessment data and place more emphasis on how schools provide a broad and balanced curriculum with AshA adding:

> And she's (HMCI) hammering exam factories that produce exam results. So they're widening their scope so hopefully there'll be a bit of backwash on that.

However, MelPol held the view that Ofsted is not fit for purpose. Despite airing similar concerns over the narrowing effect of accountability measures on the curriculum offered by schools, AliA balanced this view against the prerogative of policy makers to use accountability measures to focus on what they perceive as subjects deserving more attention:

> We've had ministers openly saying that they want the focus to be on those core subjects. So, if that's a policy intention then that's kind of their prerogative in a way. They were elected to make those kinds of decisions and if they want the curriculum to be focused around those particular subjects then accountability tools are quite good ways for achieving that.

8.8 The use of assessment data within schools

Despite participants' concerns over the use of test and examination data as a key accountability measure, none were in favour of taking either National Curriculum tests or qualifications out of the equation. The concerns were more generally around the narrowness of the measure and the impact on the behaviour of schools. There was also concern that the obsession with data derived from national tests and examinations influences wider practice in schools driving them to generate internal data often of a spurious nature.

HaydenPA lamented the obsession with data and raised concern that teachers do not have the knowledge to interpret and use it effectively:

> So it's a complete industry, built on nothing …very often teachers don't have any knowledge whatsoever of how to interpret data, how to interpret a graph, how to be misled by information.

However, SamCS was more forgiving of the misinterpretation of data noting that it is often used for purposes for which it was never designed:

> So I think that the extent of which it's quite easy to misinterpret data because you are simply using data that wasn't created for the purpose that you are using it.

The quality and purpose of assessment data generated by schools was also questioned by LoganPol who put this down to the impact of the introduction of floor standards[59], another now defunct accountability measure, described by LoganPol as:

> …a kind of tougher set of consequences that would occur to you as a head or a teacher if you didn't deliver, that were quite dire. Then I think schools started taking huge steps to make sure, using vast amounts of data, that those kids were on their way to getting those GCSE results. I'm not sure how effective it was anyway, because a lot of that data was bogus data anyway, just designed sometimes to protect a particular teacher in a particular year.

Speaking as a school inspector, ChrisR was clear that it required a "sceptical eye" regarding the usefulness of test and qualifications data in evaluating a school's overall quality, particularly as any test or examination only samples a particular subject or domain. MelPol also supported the use of tests but noted the outcomes should not be used; "…as a rod to beat teachers in schools with". But JadenR was clear that despite shortcomings in what data on assessment outcomes provide, it was a necessary addition to compensate for shortcomings in the methodology used in school inspections. Given a choice, JadenR advocated the use of qualification outcomes as a more reliable source than inspection interviews or observations, but suggested that a mixed methods approach would be the ideal model. LoganPol expressed a similar viewpoint noting that inspections alone would be insufficient for accountability purposes and equally that inspections rely on data generated by externally set examinations. However, the argument presented by LoganPol (and by PatA below) that educational standards declined in Wales as a result of the withdrawal of National Curriculum testing and their use in performance tables was contested by AliA who stated:

> I would ditch the performance tables tomorrow. I don't buy the argument anymore that they're important. I also don't buy some of the more recent arguments in defence of not taking away the performance tables, like looking at what happened in Wales when they were taken away. Sure things change, but there's no clear

---

[59] The school performance measures published in performance tables are used to identify schools that are failing to meet minimum performance expectations or 'floor standards'. In 2016, a school will be above the floor standard if:

- at least 65% of pupils meet the expected standard in reading, writing and mathematics (ie. achieve that standard in all three subjects) or
- the school achieves sufficient progress scores in each of reading and writing and mathematics. DfE 2016

evidence that they changed in a particularly damaging way, that's a much more complicated story. I'd have a different kind of accountability, a different kind of Ofsted. And I would be inviting the teaching profession to think more seriously and differently about assessment.

However, LoganPol was adamant that the introduction of National Curriculum assessments (SATs) and performance tables were the drivers of improved standards of educational performance. The point raised by AliA about engaging the teaching profession "more seriously and differently about assessment" was also mentioned by HaydenTA and in reference to the NAHT Commission on Accountability, spoke of the belief that:

> If the profession has a greater opportunity to reflect on big ideas, in relation to things like assessment and what are the kind of things that will make a difference to learning, and if our leaders are more confident in that space…then suddenly we start to have more confidence talking about the things which we are judged against.

However, SamCS pointed out that as long as data are generated, someone will compile it and that the system of using test and examination data as a measure of performance has become the norm, making any move away from such an approach an "un-realistic" proposition. AmarPA again accepted the use of assessment outcomes as legitimate measures of system performance, but in similar fashion to JadenR questioned the validity of using the data for assessing schools noting that:

> Basing accountability entirely on public examinations is perhaps less than 100% valid. I think it is significant, but I think it is merely one of a number of elements – a very important one, but there are others which are not used. It's high-stakes not just for the pupils because that gives them passage to other things once they reach a certain level. But it's also high-stakes for schools and it's the high-stakes for schools which is perhaps less than valid.

8.9 Opportunities to use detailed national data to inform practice

SamCS raised the issue of the missed opportunity to use the more detailed data generated by national tests that provides information on the performance of individual test questions as a means of informing teaching and learning. Since the introduction of scanning test papers and on-screen marking to allow the marking of individual questions in isolation from the whole test paper, the system has provided the opportunity for detailed analysis of the

performance of questions that could be used to assist teachers in identifying strengths and weaknesses of their pupils, or of particular teaching groups. The same system is also used to mark the vast majority of GCSE and GCE A level papers. However, SamCS noted that despite the possibility, the DfE has not found a way of presenting the data it holds on National Curriculum tests to schools in a manageable way and lamented:

> So we have got this vast amount of data, yet all we publish is a line graph. It is a lot of work for a line graph.

Similar data are produced by the examination boards in England, but in direct reference to the data generated by the DfE, SamCS spoke of possibilities for analysis that may not be to everyone's liking stating:

> You can see can't you that it would pretty much shock, be horrifying to some people but you could see how you could actually look at certain postcodes and you could see where algebra is being badly taught here or brilliantly taught here and so on and use these details as exemplars. But we have never really managed to get it into a format where that becomes possible.

SamCS expressed the view that the granularity of these data would make it problematic for individual schools to trawl through the detail and draw any conclusions, but ChrisR spoke of the opportunity for schools to reflect on the information provided by these data to identify good practice to the benefit of teaching and learning within a particular school. However, due to the pressure of; "…the headlong push to the next thing" and a general lack of cohesive planning in schools, ChrisR was not convinced that such an opportunity would be seized by schools to utilise and reflect on these assessment data.

8.10 Assessment data and performance management

Eight of the 14 participants questioned the validity of using assessment data as a tool for the performance management of schools and teachers. AmarPA, ChrisR, AshA, AliA and JadenR spoke of how the use of data generated by internal or external assessments for the purposes of performance management can lead to the perversion or gaming of the system. For example ChrisR stated:

> I find it hard to conceptualise any assessment system that uses teacher assessment other than for internal use that would ever be anything other than

gameable…and we know that even teacher assessment used internally, if schools then set targets around it for performance management, that's entirely internal, it still gets – you know, gamed. So there's not a great history on this I'm afraid.

AshA expressed the view that teachers' assessment should ideally be included as part of an "eased" accountability system, but raised concern that the high-stakes nature raises the risk of distortion:

I'd welcome teacher assessment but you can't do it in such a high-stakes accountability system because it's bound to get distorted. Hence, we would ease up on the accountability system, not remove teacher assessment.

HilaryLA, HaydenPA and JoPol spoke of the threat and fear caused by the use of assessment data as a facet of the performance management of teachers, school leaders and those working for other educational bodies. JoPol spoke in favour of the use of targets, but described differences in the way schools use targets as a means of aspiration, compared to Government's approach as a threat or 'punishment' for school managers:

So I'd take the punitive bit out of targets. Schools use targets internally as an aspiration, as an encouragement. Government use it as a punishment.

And HilaryLA, speaking from a local education authority position explained being on the receiving end:

The numbers at the end are really important, I have been responsible for an LA, there was a great deal of pressure on us improving our outcomes.

8.11 The pressure and stress of accountability – across all players in the system

Other participants referred to the issue of pressure noted in the previous section. The impact of pressure caused by the high-stakes use of assessment outcomes was frequently related to the pressure on pupils to perform well in tests and examinations. Some of this pressure was referenced to the high-stakes nature of qualifications for pupils as the means for progressing to the next stage of their education described by AmirA as the; "…passage to other things once they reach a certain level. They are then allowed to progress into

different career paths". And whilst accepting; "…that in an ideal world you wouldn't want to put children through multiple tests", SamCS expressed the view that for National Curriculum assessments:

> I don't actually believe the burden is that high, I particularly don't believe at Key Stage 1, the best schools I have seen the kids don't know they are doing the tests. Key Stage 2, I think that there is this whole industry that's developed around kind of Key Stage 2 mania: the tests shouldn't have an impact on individual children's outcomes, I do think there is a question about whether they are being used inappropriately in secondary (schools) to stream children into classes.

However, SamCS acknowledged the increase in pupils putting pressure on themselves as a result of living in a social media environment and parental demands and spoke of a "disconnect" from Government stating that the tests do not or should not have a direct bearing on pupils and the perceptions held more generally by teachers, schools and parents of the high-stakes nature of tests stating that the Government needs to be more proactive in dispelling fears around the possible consequences of poor performance in tests in isolation from other factors.

As in the case of HilaryLA who explained how pressure was felt at local authority level, two of the politicians spoke of the feeling of pressure as education ministers. For example LoganPol recalled:

> I worry every year when the Key Stage 2 results come out…and every time the Key Stage 2 results come out, I really worry about two things: has the proportion reaching the expected standard gone up, and are we closing the attainment gap for disadvantaged pupils? So those worry me.

But LoganPol noted there was less of a worry around the publication of GCSE results; "…because we claim to have eliminated inflation of GCSEs". The reasons behind the pressure on ministers was explained by LoganPol:

> Because we're spending forty billion pounds of taxpayers' money. And we've got to account for that, and we've got to be able to demonstrate to Parliament, to public, to taxpayers, that that money is being effectively spent.

JoPol recalled a conversation with a senior government adviser about a world cup qualifying game for England the night before a National Curriculum test and the possible consequences for government ministers of a nation of tired children sitting the test the following day:

> I said to (adviser) it's going to be a disaster, they're all going to stay up late, the kids, and they're not going to perform as well. What? Where are we? Where are we that I should have thought that that was a problem? If they were tired and didn't do well, that roll on effect wasn't just for the kid and for the school, it did absolutely go to the top of the ministerial chain. Now that was silly, and it's a good example, because I did have that very serious conversation.

However AliA posted a reminder that qualifications have to serve real world functions that will reasonably attract a level of pressure from public scrutiny, and equally, that National Curriculum tests have evolved into tools of the accountability system, therefore they too will be the focus of public interest and therefore perceived pressure.

8.12 The perceived quality of tests and examinations

Despite identifying negative aspects of tests and examinations, for example narrowing the curriculum and the use of data for multiple purposes, most participants made positive references to the quality of both the national tests and examinations used in the English education system. For example PatA noted:

> The fact is, that GCSEs in particular are not bad assessments, they are reasonably good tests worth teaching to. And as Wales have learnt when you get rid of assessments the whole system loses focus and achievement goes down.

HilaryLA favoured the reinstatement of statutory testing at key stage 3 as a means of improving standards:

> So if you have accountability at the end of Key Stage 3, you would have to have better teaching at Key Stage 3…I don't think Key Stage 3 teaching is as good as Key Stage 4 teaching…that would then mean that the accountability is spread across more year groups.

However, HilaryLA was clear that testing should not be the main motivation for "better teaching". And whilst celebrating the quality of National Curriculum assessments, SamCS pointed out that such recognition is often lost or ignored in more common discourse about testing:

> I think that the effort (in producing statutory tests) is probably not well understood or recognised…but I think that we could perhaps do a bit more to recognise that actually the tests we produce for primary are broadly an international exemplar and that isn't what you would pick up if you read the press… or the TES message board.

TerryA also talked of a desired increase in the amount of testing, but the emphasis was on low stakes testing to support day-to-day teaching and learning rather than at national or statutory level for the purposes of accountability. However, LouPol was of the opinion that the assessment system in England has focused on one form and for one purpose; "…which is all about one sort of assessment, can you pass the hurdle?"

8.13 Trust in teachers' assessments

There was a broadly held view across participants that teachers' assessments in the context of a high-stakes accountability system are not trusted in the English education system. Evidence of the lack of trust was presented from a range of circumstances, for example where examination outcomes are used as a measure of school performance, the performance management of teachers and comparison between schools. In terms of teacher's assessments being an element of examinations used as measures of accountability, the concern centred on the pressure to manipulate the outcomes. For example AliA noted:

> I think there is a general lack of trust in teacher assessment. Rightly so, because I wouldn't trust it given that examination results are used for accountability purposes and in some instances to pay teachers' wages and to ensure their promotions and so on and their career paths. It would be easy if there were teacher assessments to fiddle the figures.

MelPol was however forthright in the view that teachers should be trusted but that messages from politicians in particular undermined such trust:

That's part of the problem – the thing that sickened me as a politician, as a teacher, was when a politician would get up who had no pedagogical knowledge at all and say, the way to solve every social ill in our society is to put it on the curriculum, as if we weren't busy enough in the first place. So yeah, I do trust teachers, and I think that…a minister saying, I trust teachers, will be quite a powerful, will be quite prophetic in that sense.

AshA expressed the view that ideally teachers' assessments should contribute to high-stakes assessments but recalled an example of a former prime minister airing the view that this provides the opportunity for teachers to self-promote in an accountability system reliant on examination and test results:

But I think it's very difficult when…you're back to the accountability range, you know, John Major's thing of: "This is teachers rewarding themselves". Like they're marking it and their school depends on good scores – lots of pressure there. So I can see why coursework has been kind of abused…so it's a problem when the results are so important.

A particular concern raised by AmarPA related to the low level of trust between teachers:

My whole history of working with teachers has been one that has signalled that lack of trust. And it's not just the lack of trust from the public about teacher assessment, it's the lack of trust between teachers of assessments, even down to within the same school - that the teacher in year 4 doesn't trust the assessment of the teacher in year 3.

A similar view was expressed by ChrisR, JadenR and AmarPA but here the focus was on instances where pupils transfer from infant to junior schools. For example ChrisR noted concerns about the veracity of some assessments:

…inherently teacher assessment is unreliable. It's been proven and there's great research to show that that is the case, especially when you're using it for high-stakes accountability purposes. I mean it's understandable – therefore you shouldn't set up something that's going to force people into a position to do it. That's why you see the anomalies between the progress between Key Stage 1 and Key Stage 2. It's all well documented, the evidence around infant schools and their

outcomes compared with primary schools and then the progress scored for junior schools for example.

Despite raising these concerns, both ChrisR and AmarPA made clear that in their view cheating was not widespread with AmarPA stating: "I don't think teachers generally set out to cheat – that's not my experience". And for ChrisR, part of the concern was more about the subversive practice of re-taking GCSEs in search of higher grades. However, HaydenTA noted that because of a discourse around educational assessment that centres on concerns over gaming and "playing the system", teachers' assessments are seen as being problematic rather than as a key pillar on which the status of the teaching profession is based:

> At the minute it's almost like, teachers assessment doesn't count, everybody's worried that it's about gaming, so we can't have the teachers assess anything. We could take away assessment from teachers, but you take away one of the three pillars of what it means to be a teacher, it's pedagogy, it's curriculum, it's assessment.

In reference to teachers' assessments, JadenR made a distinction between what was described as "unconscious bias" on the one hand and intentional malpractice on the other in relation to teacher's involvement in high-stakes assessments. Further, JadenR questioned whether the use of the word 'trust' was actually applicable in discussing concerns over teachers' assessments stating that teachers are no more or no less fallible in their assessment practice as the rest of the population. And given that teacher's assessments involve activities that make judgments about the quality of their students' work that ultimately result in judgements about their performance as teachers through the accountability system, there is inevitably a perceived or real conflict of interest. For TerryA, this conflict prompts the need for a discussion about the purpose, need and fairness of teacher assessment. TerryA concluded that because of the internal and external pressures placed on teachers, the demands are ultimately unreasonable:

> Because what we were asking of teachers was unfair, of them especially. They were having to maximise their scores for those schools that had headteachers breathing down their neck to maximise their scores for their schools. But at the same time they were being professionally asked to be the independent objective assessors operating as an agent of an exam board. That's too conflicted, you know, we constructed the role badly.

8.13.1 The role of awarding bodies

AliA agreed with other participants that there is a general lack of trust in teachers' assessments and cited this as the reason for the need for the establishment of awarding bodies as guarantors of qualifications. In practice, awarding bodies work as a higher authority and provide system credibility and quality assurance. AliA noted that awarding bodies have the responsibility for assuring the standards of qualifications and where such organisations devolve responsibility for example through the assessment of coursework, then they are accountable for developing controls to ensure the integrity of their qualifications. AliA reflected further on the need for controls used by awarding bodies to quality assure assessments made by teachers on coursework in GCSE examination noting that a lack of assessment understanding amongst teachers further undermines trust in the process.

8.13.2 Trust in the qualifications system

The controls put in place by English awarding bodies and regulators are the response to what AshA described as a low trust system, especially in comparison to some other jurisdictions:

> I'd say England is very representative of a low trust system. We have low trust in teachers, so we do everything to check against the teacher rather than incorporating the teacher's judgement in this.

According to AshA however, the systems put in place by the awarding bodies still attract a high level of appeals and requests for the re-marking of examination scripts by schools who hold a similar level of distrust of the system: "They don't trust the system entirely". This view was challenged by AliA who cited a series of papers and lectures by Baroness Onora O'Niell on the discourse around the lack of trust in authoritative bodies in which O'Niell concluded that in reality there is still a high level of trust in our society. AliA related this conclusion to the English awarding bodies noting that discourse around the improvement of assessment is not necessarily generated by a lack of trust:

> And what she (O'Niell) actually said was, "In fact we do have a lot of trust for our authorities," even though there's a lot of discourse that would suggest the other way. But I don't think anyone's saying: "We're gonna get rid of awarding

organisations per se, because they're so bad that we just don't trust them". I think the reality is, we actually do trust them.

AliA supported this view by noting that despite concerns about the quality of marking for example, if there was no general plausibility in the results of examinations and tests, the system would collapse. However, AliA noted that a general understanding of the inevitable imperfections of any assessment is not widely appreciated and that some critiques of the assessment system are the result of political motivation. However, AliA noted that the assessment community has become more transparent through, for example, being more open about the possibility of error, the procedures for re-marks and appeals against results before certification, and the publication of mark schemes concluding that such openness; "…helps to improve understanding of the system and the understanding of the fact there might be error there. It also helps to empower people to put it right".

8.14 Teachers' understanding of the assessment system

Other participants raised the general level of assessment understanding amongst the teaching profession as an issue. For example, AmarPA suggested that what some might describe as unethical assessment practice may in fact be more to do with a lack of understanding. And LouPol stated that improved trust is reliant on good training, but as HilaryLA noted, despite the reliance on trust and the goodwill of teachers, access to professional training and development is very limited and more often "squeezed" by other priorities. MelPol noted that changes to the curriculum over time have created pressure on teachers to keep up-to-date with more demanding content and associated assessment expertise – a situation that has caused some teachers to quit the profession. MelPol stated that; "…teachers should be good at assessments, it's an integral part of the job" but if it is identified as a weakness, it should be rectified through continuous professional development. However, MelPol noted that access to such support has become problematic as; "…we know that continuous professional development's been burnt, slashed and burnt". AmarPA, and JoPol supported increased opportunities for professional development, from which the level of trust may increase. For example JoPol stated:

> …maybe if we had the Chartered Assessors, if we did more professional development… Because do you at this moment in time trust teachers to do all the assessments, I would say no. But it's a desire, I wish we could. But it won't happen until we support them in being able to do that.

8.15 Is there such a thing as teacher assessment?

For TerryA, the concept of teacher assessment in high-stakes assessments is itself open to question. This challenge is based on the view that although teachers are more than capable of making distinctions in the quality of their own students' work to place them into a defensible rank order, they cannot make reference to the wider population through a set of independent metrics. In other words, teachers can only base their judgements within a particular setting driven by localised incentives and drivers. However, SamCS somewhat countered TerryA's view stating a warming towards the use of teachers' assessments through a greater appreciation of its limitations and purpose. However, as with TerryA, SamCS noted the difficulty of achieving a standardised outcome in assessments carried out by teachers leads to more reliance on standardised tests.

8.16 Changes to assessment practice over time; the curriculum and accountability

Some of the Key Influencer participants have been part of the education system for some time and recalled how assessment practice has changed over their period of involvement. The growth and impact of the accountability system was the most cited key change in the application and use of educational assessment, and for AmarPA, the change has been driven; "…much more for the purposes of accountability than for the benefit of individual children". AshA suggested that the current population of teachers have in reality not needed to fully understand assessment, as the arrangements are determined nationally without their involvement. This point was also picked up by LoganPol who noted that the overall effect on teachers was that; "…they've kind of drove (Sic) out of the skill of testing children". However, AshA noted that previous practice was not necessarily "all good":

> ….you know you can carbon date some of us by chuntering on after two pints about Mode 3 CSEs. In that sense there was a time when teachers designed their assessments and their curriculum. But those days are long gone and they weren't all good, some of it was very lazy stuff.

Whilst AshA and AliPA were critical of what was described by AshA as interference in the details by; "….whimsical ministers and secretaries of state", both supported the introduction of a National Curriculum and the general move to more aspirational standards of performance for all students. However, the idea that a 'National Curriculum' that is no longer statutory for a large proportion of schools questioned its status. However, AshA

noted the irony of a system through which the demands of the examination system still; "…forces you to do some of the curriculum".

The quality of tests and examinations, the more detailed syllabuses and public access to the mark schemes used by awarding bodies were also recognised as significant and beneficial changes to the assessment system by JadenR who noted that in the past:

> …our knowledge of assessment was weaker. Our operational systems were weaker. If you look at what we've now got in terms of quality of marking, in terms of on line, live quality assurance, of course it's better.

JadenR noted that access to mark schemes has become the norm since the late 1990s and is no longer the right of a privileged group of examiners, often in the past coming from independent or selective schools, but pointed to the "down side" that the pressure to succeed in examinations has increased to the point where:

> …it seems to me we're in a bit of a mad world where through the publication of mark schemes we've now got kids learning mark schemes. That's not good, is it? That's a bit bonkers, isn't it? But how can we not publish mark schemes?

HilaryLA spoke of the impact of the more recent change of the removal of National Curriculum levels that was trailed by the Government as a means of freeing up schools to devise their own assessment systems. HilaryLA noted that the system has become more prone to change leading to teachers being more cynical about the motives of politicians. This view was illustrated by reference to a conversation between headteachers who had decided there was little point in working out their own assessment systems because the government would likely issue further demands in due course. However, not all participants expressed the view that there has been substantive change in the assessment system. AliA was singularly of the view that despite a sense of evolution, the fundamentals of the qualifications system had remained:

> It's gradually evolved over time, but actually it's not a million miles away from how it was for a certain group of the cohort who were examined back 150 years ago.

8.17 Teacher's knowledge of educational assessment

This section of the interviews with Key Influencers produced the most unified set of responses with the most common view expressed being that teachers' knowledge about educational assessment was in general weak. Participants suggested that much of this stems from the lack of attention to educational assessment in initial teacher education. Reasons for this included time pressures in ITT courses and the differing routes into teaching being a cause of variability in approach, duration and course content. Other reasons included the paucity of continuous professional development in this facet of education and the lack of high-level specialism and interest in universities. This was despite the view often expressed by participants that assessment is central to the teaching and learning process and a central element in the school accountability system in England. For example, HaydenPA expressed the view that the use of assessment outcomes in the English education system has become an "industry", yet despite this, teacher's knowledge is insufficient:

> So it's a complete industry, built on nothing…we don't have an opportunity to learn about assessment, very often teachers don't have any knowledge whatsoever of how to interpret data, how to interpret a graph, how to be misled by information. So assessment is important, knowledge of assessment is important, and at the minute it's pretty scarce.

This viewpoint was shared by JadenR who referred to persons in schools and colleges charged with the responsibility for managing the data emanating from examinations and tests, but with little real understanding of their limitations. SamCS raised the same issue relating to teachers' general understanding of data stating that this is an outcome of the lack of focus on assessment in initial teacher education and professional development. AmarPA was more complimentary about the broad understanding of assessment issues in schools, but based on some recent work with primary phase teachers, differentiated between the primary and secondary phases of schooling noting that although the ability to define key concepts in educational assessment may be lacking, the general awareness of assessment issues is stronger in the primary phase.

AshA was of the view that teachers in general draw on assessment processes in their day-to-day interactions to make modifications to their teaching, but lacked understanding of some of the key assessment concepts as also noted by ChrisR:

I think that the lack of assessment knowledge with teachers is around things like the reliability and validity of it. I think the idea that you do – you set some tests and that if a kid gets 5 out of 10 on one and then later gets 7 out of 10, they think there's improvement. I think that sort of thing is not well understood.

A key reason given by AmarPA for the deficiency in assessment understanding, especially in respect of key stage 3, was that heads of departments had not focussed well enough on this phase of education concentrating more on the "end game" of GCSE and A level qualifications as the key determinants of what drives the secondary phase of the education system. JadenR expressed the view, with a level of self-confessed cynicism, that teachers do have an understanding of the limitations of the reliability of assessment gained through their own experiences, but they applied different standards to the marking carried out by examination boards compared to that of their own marking suggesting this a ploy in the pressured system of accountability.

8.18 The lack of assessment expertise in England

Two of the Key Influencer participants spoke of what they described as the lack of a community of assessment experts in England. AliA noted that summative assessment in particular lacks traction in the higher education community and that compared with ten years ago the interest amongst academics in formative assessment has declined:

…there's isn't a big community in this country, in higher education… who know about assessment, but certainly not summative assessment. But there isn't a big community looking at formative assessment either in the way that there was ten years ago. So, that assessment reform group fizzled out in 2010 and nothing's taken its place.

This viewpoint was echoed by SamCS who spoke of the difficulty amongst national agencies in recruiting assessment experts to develop national rests and examinations that result in a dependency on in-house training. SamCS added that a more engaged and informed community of post-graduate assessment experts would in the long term prove beneficial to classroom practice, as knowledge of how to develop good assessment instruments would become more widespread and ultimately be of benefit to teachers.

LoganPol added a more general criticism of the ITT sector citing a lack of engagement by the sector in debates around education stating: "I'm critical of them generally…because they're quite impenetrable these colleges".

8.19 Assessment instruction in Initial Teacher Education provision

The majority of Key Influencer participants shared the view that ITT is deficient in preparing teachers in terms of educational assessment. However, participants expressed a range of underlying reasons for this. At one extreme, AshA suggested that in reality there is little need for developing the understanding of assessment amongst teachers because:

> …they haven't needed to understand it…if you were a young teacher you've been through nothing except a structured curriculum, you have tests in a certain way. They've not been enabled.

AliA presented a similar view noting that because of the reduction in coursework and controlled assessments in high-stakes assessment, there appears little point in preparing teachers in respect of summative approaches. Other participants shared the view expressed by AliA that key elements of educational assessment are seen as being in the gift of external agencies. For example PatA expressed the view that coverage of assessment theory and practice in initial teacher education is:

> …dire and, you know, it's a disaster…it's serious, in England the situation is serious but not hopeless. Or you might actually say it is hopeless but not serious, because teachers know nothing about assessment but it does not matter because the only assessment that is done is done by external agencies.

This view of the externality of assessment was re-stated by AmarPA who recalled a series of conversations with teachers at the Education Show in Birmingham in which:

> …over 90% of those teachers said, "Assessment's not part of what I do, it is done to me." And they viewed assessment as something that was external and done to them in a bureaucratic fashion. I found that very sad.

LouPol also referred to the context in which teachers work and was of the view that frequent changes by governments to high-stakes assessments throughout the system are a central cause of the lack in understanding. The influence of government policy on summative educational assessment was also recognised by AliA as a factor in reducing its place in ITT, and even though formative assessment has retained some broader interest, it too is subject to an element of fashion. PatA was also critical of the influence of government stating they; "…never felt the need to take assessment seriously".

8.20 Variation in routes into teaching

Three participants raised the increased number of routes into teaching as an issue. LouPol suggested this was a reasonable response to questions about the fitness for purpose of the more traditional routes into the profession, whilst HilaryLA raised concern that despite all routes into teaching being subject to some form of accountability, there was some doubt over the level of comparability of standards. For PatA, the key to a successful course of teacher preparation should not be related to the route, but to the quality of output. However, some of the current provision was questioned:

> What I think is completely unacceptable at the moment is that there are lots of people getting teaching certificates from SCITTs who really should not be teachers.

Despite raising concern over the potential differences in content and quality in the differing routes into teaching, LouPol was of the view that this was not necessarily the "fault of institutions" but rather as a consequence of competition between trainers to attract candidates onto the various routes. A further issue regarding the current diversity of routes into teaching was raised by HaydenPA who questioned the level of expertise available in some ITT provider institutions. AshA suggested that the lack of knowledge amongst those who train teachers is reflected internationally but may be the result of a focus on other priorities. However AmarPA was of the view that expertise in educational assessment theory and practice does reside in university based initial teacher education but concurred that short time allocations prevented the development of any depth in student's understanding: "To do that requires a lot of time and they don't have it".

8.21 Insufficient time given to educational assessment in ITT

Although the competing demands on ITT course content was recognised, there was general agreement between participants that despite the attention summative assessment

receives in practice, teacher educators devoted little time to educational assessment theory and practice. For example JadenR noted:

> …so as part of teacher training how much time gets spent on assessment? And yet they spend an inordinate amount of time in the classroom doing bloomin' mocks every five minutes. Some of them spend an inordinate amount of time on preparation for summative assessment. And yet it doesn't really feature in teacher training. Isn't that nuts?

HaydenPA concurred noting that initial teacher education has; "…been paired back and paired back and paired back" to the point that it places unreasonable pressure on providers to cover a range of elements with; "probably…even less (time) to think about issues around assessment". The competing demands made of teacher educators was also raised as being problematic by JoPol, as they laboured to deliver other elements in response to the ever changing demands of Government. The shortage of time to cover what might be deemed as desirable content in an ITT programme was also raised by AshA and ChrisR who stated that thoughts of increasing content relating to assessment theory and practice may be 'fanciful':

> So the idea that you're going to get loads of time on PGCE courses is probably fanciful. But you can I think plant the seeds…because what they will see in schools will inevitably start to shape their practice. And if we're putting them into schools that are using all those spreadsheet methods, they will just see that's how it is. And so they need to have the seeds of doubt put into their minds that that isn't the way it is.

However, TerryA spoke of the need for more content on assessment theory and practice in initial teacher education courses, a suggestion that had attracted some criticism:

> I got into terrible trouble here. Because I suggested that there should be much more assessment in initial teacher training, PGCE and so on. And it was claimed that there was already quite a lot. But it's not significant in terms of what really needs to be understood.

Even here LouPol questioned the potential difference between the stated curriculum and the enacted curriculum of teacher preparation courses:

Sometimes you look at the curriculum of teachers in training and you do wonder, you know, how much of the syllabus they actually cover, and what's in it; and whether the preparation they get is good enough for when they're in the classroom.

A similar concern prompted LoganPol to conclude that there should be a 'national curriculum' for initial teacher education courses. However AliA recognised that in any curriculum, priorities need to be established noting that summative assessment, because of its perceived externality, could be viewed as less relevant in initial teacher education and that a deeper focus on assessment for learning may be more beneficial given the competing demands of the ITT curriculum content. However, SamCS was less dismissive about the relegation of summative assessment stating that knowledge of summative assessment is of equal relevance to the classroom.

Despite AliA's view that formative assessment was; "…just seen as part of good teaching", other participants raised concern that knowledge and understanding of the concept was at best variable as expressed by TerryA:

And the variable interpretation of AfL actually indicates very strongly the unevenness of understanding of the principles of assessment. So we wouldn't accept that variation in medicine in a particular clinical technique for example. So I think there is absolutely a national need to improve assessment literacy.

AliA, HilaryLA, and HaydenPA presented the same analysis of teachers' understanding of assessment for learning pointing to wide variation in the way it is interpreted and applied. For example, AliA stated that the concept is "woolly" and open to interpretation, but when done well, is of great benefit.

8.22 The impact of schools in determining acceptable assessment practice

The role of schools in shaping the practice of newly qualified teachers was repeatedly raised even where initial preparation had laid the foundations. For example AmarPA stated that: "Unfortunately, they then go into schools where perhaps assessment practices are not fit for purpose". Relating this aspect to the variability in the content and quality of routes into teaching, TerryA noted that this results in differences in understanding of

educational assessment that may well be compounded by the variable levels of understanding of assessment practice in schools. AshA presented a similar argument:

> I think one of the drawbacks of the current model of teacher training is that because the schools are responsible you get put into a school and that becomes the model. At least in the old PGCE course there was room for a bit of questioning on, "Why are we doing this?"

HilaryLA and AliA also raised the point that in reality it would be difficult for newly qualified teachers to challenge the assessment approach used by a school.

8.23 Initial teacher education and continuous professional development

Participants made a number of links between initial teacher education and the need for subsequent professional development in the field of educational assessment with JoPol stating that professional development should be a compulsory facet of the teaching profession. However, despite what TerryA described as; "…a latent demand, both in terms of need and in terms of desire" for CPD amongst teachers, the lack of access to quality training post qualification was a concern. For example ChrisR, who recognised a growing interest in assessment theory and practice, stated there is:

> …virtually no time spent on this in training and then only certain people getting chosen to go on CPD, and that CPD sometimes not being very good because it's quite often provided by people who have just come through the system and think of assessment as putting those numbers on those spread sheets. Whereas actually, what we want is CPD that explains…reliability, validity, how do you work out whether a test is reliable? And all those kind of things.

AshA raised similar concerns about continuous professional development courses noting that the focus was more often on:

> …how to get better grades, what we need to do and how to prepare for an Ofsted inspection, you know, that kind of thing. So it's all kind of technician stuff rather than, "You're a professional with freedom to move on this".

HaydenPA and LouPol suggested that schools should have teachers with designated responsibility and training in assessment, for example Chartered Educational Assessors, but that all too often teachers are given responsibility without adequate training and support. HaydenPA explained this was often as a result of poor leaders; "…who haven't themselves engaged in any further professional studies".

8.24 A new model of initial teacher education

Six of the participants called for consideration of a new model for initial and post teacher education all of which suggested the need for a longer period of instruction and support. LouPol, JoPol, HilaryLA and HaydenPA suggested the establishment of a period of time in post should be introduced following initial training. LouPol described this as; "…almost a graduate apprenticeship". PatA drew on international approaches such as in Singapore, Germany and Finland calling for a system that would pay teachers more, but in return for a three-year internship before qualification, a system akin to that used in the legal profession. And within the internship, additional rigorous tests of knowledge would need to be passed, one of which would be on assessment literacy.

Sound selection procedures and tenure in post were also seen as a key determinant of producing successful teachers and whilst PatA was not fundamentally against:

> …so-called employment based roots…recruiting people who stay for a couple of years, then that is a bad system. Not because the training is worse, but because all teachers are worse in their first two years.

8.25 What teachers need to know about assessment on entering the profession

Although there was general agreement that there was insufficient coverage of educational assessment in initial teacher education, most participants spoke of the difficulty of introducing more content into the preparatory stage, even where courses ran for twelve months or more. AliA added that courses on assessment offering any real depth of coverage take time and, in a system, where teachers now play a small role in statutory tests and qualifications, consideration of including summative assessment processes and techniques may well be redundant. AshA and AmarPA suggested that newly qualified teachers have a broad knowledge of assessment issues because of their personal experience as pupils and students in a system heavily reliant on tests and examinations;

even so, AmarPA questioned their depth of understanding. AshA suggested that providing newly qualified teachers with the facility and time to reflect on the purposes and fitness of purpose of assessment would be beneficial. Other participants spoke of the need for a set of assessment basics such as an understanding of validity, reliability and bias but as JadenR suggested, these are complex concepts. AmarPA was of the view that the basics of educational assessment theory and practice are covered in some initial teacher education courses including attention to areas such as asking effective questions, for example based on Bloom's taxonomy to support teachers in probing their pupil's level of understanding. However, JadenR questioned if the expectations made on newly qualified teachers are realistic:

> I just wonder whether we're expecting too much of teachers, expecting them to engage with the finer points of assessments. Maybe we need to start small and our ambition is to grow, because they're just so overworked and feel so beleaguered.

8.26 On-going teacher development of assessment knowledge

The view that there should be a line of progression in developing teacher's understanding of assessment over time was an approach popular amongst participants. For example AshA was of the view that:

> There may be a good case for saying that things like understanding assessment is probably better done two years into practice… that's when I would think as a professional you need to broaden your view of assessment.

However, AshA noted that the mechanism for deciding on the appropriate content was unclear and raised concern that training was in danger of focusing on how to use data to increase performance in tests and qualifications rather than on improving the teaching and learning process. There were also suggestions that continuous professional development should be a professional obligation as in the field of medicine, with JoPol suggesting that this should be set in a time managed framework. But AmarPA raised concern that the level of required funding is not in place for schools and that any individual teacher following a self-funded course such as a Masters, would not necessarily recoup the expense. JadenR was equally supportive of access to continuous professional development, but made clear that the motivation should come from the profession: further, the frequency of changes to the system added a potential barrier.

8.27 Examiners and examination boards

Three participants suggested teachers should experience the role of an examination marker or examiner as a means of gaining or extending assessment expertise. JadenR referred to the training offered by examination boards to schools but noted that this tends to focus; "…purely on qualifications rather than actually on assessment" along with open access videos on the processes and procedures undertaken as part of the administration of qualifications. However, JadenR noted: "But I guess if you're a really, really busy teacher, are you going to go and seek that out? And maybe it all just seems a little bit distant?" Nevertheless, JadenR and AmarPA were of the view that taking part as an examiner or marker, besides being financially rewarding, does provide beneficial knowledge. ChrisR agreed with this view but was of the opinion that the experience was relatively narrow and in line with HaydenPA's call for a Chartered Educational Assessor in every school, suggested there was a need to have:

> …somebody in the school who is really red hot on this who can then disseminate and train within the school, otherwise it's just too piecemeal.

The narrowness of approaches to assessment was also questioned by LouPol who spoke of the need to encompass wider methodologies to cater for subjects that don't fit conveniently into externally set tests such as those found in vocational or creative areas. AliA concurred noting that in terms of vocational qualifications:

> …maybe in those particular contexts there is more of an argument for a lot more professional development in summative assessment. Whether that argument holds true to the same extent in other contexts I'm not sure.

8.28 The future of assessment in the English education system

The Key Influencer participants presented a range of scenarios and challenges for the future of educational assessment in England. SamCS suggested that some of the changes in assessment practice over the last ten years have not only been necessary but have taken the system from a period of evolution into one that has seen more willingness to engage in conversation. LouPol spoke of the development of technological aids such as automated marking but noted the limitations of traditional tests and the need for more nuanced assessments especially in assessing creative or vocational skills that rely on

interpersonal relationships in what was described as a 'hands on' approach by teachers. LouPol stated that there was a growing need and desire to develop new approaches:

> So I think the tectonic plates are really changing, and the training of teachers has got to adapt, all the main players, the examination boards, the regulators, have all got to realise that's happening.

AshA raised a similar issue in relation to the assessment of vocational skills stating that new developments such as the T Levels qualification require different approaches to those used for similar qualifications in the past: however, the prognosis was not optimistic because of the complexity of the qualification.

8.29 Long-term planning and politics

SamCS supported the idea of using more technology in areas such as marking and suggested that government should take more of a lead because it is not driven by commercial influence. Further, schools need support from those with assessment expertise in order to select appropriate technologies from what was described as a proliferation of alternative products. SamCS spoke further about the lack of any cross-party consensus on a range of key areas of policy including that of education, noting that "political instability" has created a tendency to focus on immediate priorities at the expense of a long-term vision. However, SamCS noted that access to what are limited resources acts as a barrier. According to SamCS, such a situation could be aided through taking some of the; "…heat and the fire away from statutory (assessments)" to facilitate more constructive thinking as teachers in essence are interested in how well their pupils are progressing and in comparing their ideas and judgments with others. This viewpoint was shared by HilaryLA who noted that too much effort is spent in reactionary rather in proactive mode but noted that most teachers wanted change to be swift rather than extended over a longer timescale. PatA spoke of the possibility of establishing one national awarding body for general qualifications noting the trade-off between flexibility and competition with that of maintaining standards of performance. However, one foreseen challenge would be that of keeping such a body free of any political interference, but that, PatA suggested, could be overcome by having a chief executive who reports to the Crown as in the case of Her Majesty's Chief Inspector of schools. AshA predicted change on the basis that the current system cannot continue forever citing the pressures of accountability and concerns about the narrowing of the curriculum as examples. However, AshA was

uncertain about from where any radical change might emerge. For HilaryLA, there is need for more stability in the system:

> I get a bit cross about the accountability that is thrust down everybody's throats where you change the system so often that you don't know whether that's a reflection of the system or whether it's a reflection of the teaching – there are so many variables, how are you saying what's made the difference?

8.30 The use of teachers' assessments

SamCS spoke of a personal realignment of views about teachers' assessment noting a shift away from being anti-teacher assessment, to one that understands more fully the limitations in applying a national standard, and appreciative of the advantages, such as; "…where what you are trying to get is an overall sense of pupil achievement over a period of time". However, SamCS concluded that teachers' assessments could not be used as part of a statutory performance measure; "…because I think that the weight of accountability is far harder to counteract". The underlying reason presented for this was not to question what are perfectly sensible judgments made by individual teachers, but that there is no certainty that standards are being applied fairly and accurately across the country.

SamCS also spoke of the recent decision to stop collecting teachers' assessments of maths, reading and science in key stage two on the grounds that the data are not used by the DfE or Government in the school accountability performance measures. However, teachers will continue to assess writing, a measure introduced following the Bew Review on the grounds that it was difficult to assess through a test. This point was raised by JadenR as being contradictory given that tests are used to assess writing in general qualifications: "So why on earth have we ruled out a test?" However JadenR noted that the DfE is in the process of considering other approaches to assessing writing such as using a devolved system of comparative judgment for the assessment of National Curriculum assessments. MelPol also supported more devolution of assessment to increase the agency of teachers but cautioned against any impact on workload.

More recent changes to remove or minimise the use of coursework and controlled assessments in general qualifications were raised by PatA noting that coursework had lost its original intention of being work produced throughout the course to being an additional

component produced at the end of a course. Further, controlled conditions are difficult to implement so:

> …you have got parents doing their work for their kids, you got older brothers and sisters helping them out, you have got no control over standards. So I think coursework as implemented in GCSE was pretty much a disaster. You need something like controlled conditions but even so, it is not adding very much.

Nevertheless, having outlined issues created by coursework and controlled assessments, PatA favoured the involvement of teachers in deciding the qualification grades awarded to their students. PatA spoke of the "purity" of the English qualifications system in that the role of the teacher is to help students to get the highest grade possible against externally set standards. However, using exams alone produces limited information on which important judgments are made, so ways need to be found to involve teachers' judgments as part of the overall assessment:

> What I am saying is, we need to bring more evidence into this system …the ideal system has to find a way of bringing in teacher judgements, teacher knowledge about their own students but it also has to control the problem that teachers find it very hard to align their judgements with those of teachers in other classrooms in their school. And certainly very hard to do it with other classrooms in other counties in the country.

JoPol suggested a system whereby schools work more closely together to moderate standards supported by Chartered Educational Assessors who would provide the quality assurance. However, LoganPol was clearly against the inclusion of coursework and controlled assessments in high-stakes assessments on the grounds that they are a distraction from teaching and unreliable; "…so I think examinations are the best way of ensuring children, students, remember and learn what they've been taught". HilaryLA presented a similar view that tests enhance teaching and learning. AliA was of the view that a move to a system totally made up of high-stakes external examinations should not necessarily be the goal suggesting the need for a more balanced approach and spoke positively about the development of a community of assessment experts - but noted that the idea had never gained traction and that any such movement would have to be valued by the profession.

8.31 An informed profession

Four participants spoke of the need for more investment in continuous professional development and a better-informed workforce with JoPol stating that a move to develop Chartered Educational Assessors would improve the levels of trust in teachers' assessments. But LouPol made clear that a level of suspicion should be deployed when new responsibilities are given to schools and teachers without proper training and the provision of time. SamCS preferred a whole school approach to professional development rather than delegation of assessment responsibilities to one individual.

8.32 The influence of accountability measures

The centrality of assessment outcomes to the school accountability system was frequently raised as an area in need of wider discussion and debate. Noting that the two have become totally intertwined, JoPol called for a national debate about the impact of using assessment in this context and the possible alternatives. AshA would welcome more teacher assessments but warned that if it was used in a high-stakes accountability environment, it would inevitably be distorted: as an alternative, AshA would prefer an easing of the relationship between tests and the accountability framework. However, TerryA was clear that variation in the performance of schools has to be tackled and that it's only through assessment; "…that we know what's actually going on in terms of attainment" and that more systematic use of formative assessment needs to be encouraged. This approach should, according to TerryA, be supported by the availability of rigorously produced examples of good practice. Despite reservations about the impact of the accountability system, JadenR questioned if it was now possible to:

> …become non-performative? So even if school accountability measures changed, can we move back to a day when the stakes around qualifications are significantly less? I think there's always going to be a strong role for examination and testing. Do I think that needs to drive out all teacher assessment? No, of course not, and maybe we'll end up with a bit more of it at some point, almost certainly.

MelPol favoured the establishment of an independent assessment and curriculum authority with the responsibility for improving the performance of schools moved from ministers to regional politicians supported by; "…a world-class teaching profession. If we're going to have that, it's got to be world-class at planning curricular, delivering curricular and assessing curricular".

8.33 Changes to the format of examinations and tests

Other participants spoke of the need to reform the format of tests and examinations. For example PatA suggested the use of a system of less predictable examinations by issuing papers that covered different sections of the syllabus administered in one examination session. AmarPA suggested a system of teacher assessments moderated by the use of an externally set test that would expose any anomalies in and across schools. Such an approach would provide greater insight into the performance of pupils than that provided by a relatively short-timed test. AshA was concerned about the efficacy of vocational qualifications such as T Levels in that approaches to the assessment of vocational qualifications are too heavily influenced by the preoccupation with generating levels of equivalence with academic qualifications in order to satisfy performance tables which in the past has driven qualifications such as the GNVQ towards externally set tests rather than the use of more authentic assessment methods designed to measure for example, practical skills. According to AliA, proposals for vocational and technical qualifications are influenced by the assessment design used in the NVQ model and competency based assessments which was:

> …quite a bold approach when it came in in the 1980s/1990s. And you know, it's clear that there have been some quite serious problems with that. And teacher assessment is part of the problem.

8.34 The role of professional bodies

HaydenPA raised concern about the erosion of the professional identity of teachers with HilaryLA adding that teachers need to be more independent and more generally viewed as professionals stating that much of their activity is now centred on compliance and procedure rather than through professional choice. HaydenPA referred to the recently formed Chartered College of Teaching as having the potential to change perceptions of the role and status of the teaching profession. SamCS made the point that:

> There is a sense where I don't understand why people aren't more proud of the education system in this country and proud of the fact that their children are doing fantastically well. And I think there is something about we don't really trumpet that, we are not really proud of it and there is a sense that everything is always going horribly wrong, whereas actually a lot of is going really very right.

8.35 The impact of international tests

JoPol acknowledged the increasing interest by ministers in international tests and surveys such as TIMMS and PIRLS but admitted that:

> I'd never heard of it (TIMMS & PIRLS) until I became a minister. I think some teachers still haven't. I've never met a teacher whose class have done the TIMMS or PIRLS test…I've never seen a test. I mean, I swear by the results, I look at the results and as a minister show pleasure in us doing them well. But personally I've never met a – because it's a sample. There is an example of a test that doesn't put pressure on people.

JadenR agreed that there is increased interest in these tests amongst politicians stating:

> So I think that whole thing is a lovely money making scheme, isn't it? Now, whether it ever was intended to start off like that, but isn't it now? There's no doubt that politicians are hugely driven by these league tables and can have dramatic effects on the education systems in some countries. And I think that's – it's quite scary. And it's just over-use of a particular form of data, isn't it? That's not good. And politicians will use these things to give themselves the cover for doing whatever politicians want to do, won't they?

8.36 Chapter 8 summary

As with teacher interview participants, the Key Influencers did not share a universal definition of educational assessment although recognition of the importance of inferences drawn from assessment was strong. Participants noted the dominance of qualifications in determining views and understanding of educational assessment along with the growth of assessment outcomes as measures used in accountability systems at a national level and more locally within schools. However, the centrality of assessment to the process of teaching and learning was broadly recognised. Key Influencers were broadly of the view that assessment knowledge and understanding was low across the education system and although supporting the need for accountability, there was a general view that the accountability measures have distorted the focus of education in schools, for example through narrowing of the curriculum. Such practices were also cited as drivers to use assessment within schools as devices to track progress against externally set accountability targets. Despite identifying some of the negative effects of tests and

examinations on practice, most participants were of the view that tests and examinations as used in England are of good quality.

It was a broadly held view that trust in teachers' assessments in high-stakes assessments is low but that much of this is driven by the impact of the centrality of assessment to the system of accountability. However, it was also noted that what might be described as unethical practice may be more an outcome of the lack of training in assessment theory and practice than malice. The majority of Key Influencer participants also shared the view that ITT is deficient in preparing teachers in terms of educational assessment, though it was acknowledged that the pressures on the range of content of initial education courses to be covered in limited time was influential in restricting the range and depth of topics covered.

Chapter 9: Discussion

9.1 Chapter over-view

This chapter presents a discussion of the main findings of this research project set against the key research questions that underpin this inquiry. Following a short introduction the chapter is presented in three main sections each taking one of the three research questions in turn. The first of these sections, Section 9.3, focuses on research question 1 and discusses the findings presented primarily in Chapters 2 and 4 on how the role of teachers in what have become high-stakes assessments has changed since the Norwood Report was published in 1943. Section 9.4 relates to research question 2 and discusses the perceptions of educational assessment held by the teachers and Key Influencers who took part in this research project. Section 9.5 addresses research question 3 and discusses the findings of the teacher and ITT provider surveys on the curriculum content relating to educational assessment found in initial teacher training programmes. The chapter concludes with a summary of the discussion.

9.2 Introduction

As presented in Chapter 1 of this study, the centrality of assessment to the process of teaching and learning is generally uncontested in the academic literature on education (see for example Newton, 2007; Shaffner, 2008; QCDA 2011; Wiliam, 2013; CCEA, 2014; and Wong & Kaur, 2015). This notion of centrality is also reflected in the views of teachers and Key Influencers who contributed to this research project as illustrated in the data presented in Chapter 7 and Chapter 8. However as this study shows, the uses to which educational assessment outcomes are put in the English education system have increased significantly over time and now cover in excess of twenty such uses (see Newton, 2007).

9.2.1 The growth in general qualifications

In the early 1950s, formal qualifications at the end point of schooling were accessed by a minority of pupils remaining in full time education up to the age of 16 as stepping-stones to the GCE A-level used for selection into higher education. However, the number of pupils entering and achieving passes in general qualifications has increased significantly over the period of interest from 10.7% of the relevant age group passing five or more General Certificate of Education Ordinary levels (GCE O levels) at schools in England and Wales

in 1953 to 79.6% achieving 5 or more passes grades A* to C in the General Certificate of Secondary Education (GCSE) in 2011 (Bolton, 2012).

9.2.2 Norwood's vision unfulfilled

The growth in externally set examinations was not the expectation of the Norwood Review (1943) that envisaged a post-war system of primary and secondary phases of education with differentiated provision in the latter but with relatively few qualifications and the centrality of teachers in the assessment process rather than that of awarding bodies whose purview would be restricted to that of university selection. Norwood's vision was never realised and by 1980, the curriculum and examinations were externally controlled, a situation largely tolerated by teachers (Lawton, 1980). Moreover the number of pupils accessing general qualifications grew, so did the uses to which the assessment outcomes were put, so over time assessment has increasingly become a tool of surveillance (see for example Foucault, 1977a; Broadfoot, 1966; Allen, 2012; Mason, 2019) within a culture of performativity and accountability (see Ball, 2003 & 2013). Such uses have held direct consequences for the role played by teachers in what have become high-stakes assessments. Although Norwood's vision of a teacher led assessment system did not come to fruition, teachers did play a significant role in the assessment of qualifications between the 1960s and early 1990s through for example coursework assessment and Mode 2 and Mode 3 examinations (see Chapter 4). However, over time such influence became more restricted as concerns were raised over the validity and reliability of such approaches to assessment (see for example QCA, 2005; Colwill, 2007) in a system where examination outcomes formed a central element of the accountability system. In 2010, tighter controls were adopted through the introduction of controlled assessments in GCE and GCSE examinations, but more recently, the role played by teachers in high-stakes assessments has diminished even further as controlled assessments were deemed as failing to resolve the issues identified in coursework over validity and reliability (See Ofqual, 2011 & 2013; Opposs, 2016). Few GCE and GCSE examinations now include any form of teacher involvement in the assessment process. Further, in 2017, the government removed the statutory requirement for schools to report teacher assessment judgements at the end of KS2 where test data alone is used for accountability purposes (DfE, 2017a) and from 2020, key stage 1 assessment will become non-statutory, being replaced by a new reception baseline assessment designed to form the datum for a new progress measure. These are amongst the latest in a long line of changes to the role played by teachers in what have become high-stakes assessments through their use as key accountability measures. Over time, the justification for change has often related to issues

of validity and reliability and in more recent times, consideration of the impact of assessment on teacher workload. However, this study contends that changes to the role of teachers in high-stakes assessments over time have been contingent on socio-political influences rather than being in pursuit of an ideal model of educational assessment. It is therefore somewhat ironic that in light of the decision to cancel examinations for 16 and 18 year olds in England in 2020 due to the Covid-19 pandemic, qualification outcomes will be based on grades and rank order of pupils as determined by teachers.

9.3 Research question1. How has the role of teachers in high-stakes educational assessment changed since the Norwood Review (1943) to the present day?

Chapter 4 used the work of Paul-Michel Foucault as a framework of inquiry or 'extra dimension' in the analysis of the intersection of power and knowledge (Kendall & Wickham, 2003) to examine the role of teachers in high-stakes assessment since the mid-twentieth century. In Chapter 4, I identified three epochs highlighting significant change in the role of teachers in the process of educational assessment in schools in England: such epochs are described by Foucault as epistemes, that is "…periods of history… that were organised around their own specific world-views" (Danaher et al, 2002. p. 15). The three epistemes help in identifying key shifts in thinking around the principles and purpose of educational assessment and the part played therein by teachers. These three epistemes, summarised in Table 9.1, are now discussed in turn.

9.3.1 Episteme 1. The war years and the post-war ambitions for education (1943 to the mid 1950s)

The first of these epochs stemming from the war years and post war goals of education would appear to offer raised status for teachers as enhanced professionals (McCulloch, 1993) through being the key operators of educational assessment in a post-war period driven by the optimism of a tri-partite system of secondary schools designed to meet the needs of individual students and the nation with parity between the different types of schools as proposed by the Norwood Report. And as noted by Lawton some three decades later, the control of examinations "…is of crucial political significance" (1980, p. 83). However, Norwood's vision was not universally supported even within his own committee as illustrated by Brereton's concern that examinations were becoming to be regarded by many people "…merely as a necessary evil" (1944, p. vi). Further, Norwood

(1943) noted that not all teachers favoured such an approach and that the public may not be fully prepared for change (pp. 45-46).

Even so, the Norwood Review concluded that as long as teachers accepted external control of; "…what they shall teach and external assessment of the way in which they have taught, they can never rise above the rank of journeymen" (Norwood, 1942b., In McCulloch, 1993, p. 134). But teachers did not embrace the opportunity offered by Norwood with teachers' groups; "…at best lukewarm and often actually hostile in their reactions…raising issues about their own perceptions as to the proper role of teachers" (McCulloch, 1993, p. 136).

Table 9.1: The three epistemes identified in this study: key characteristics and approximate period

1. The war years and post war goals of education (1943 to the mid 1950s)
• discourse on a more egalitarian post-war education system • free access to education for all • creation of a secondary phase of education divided into three strands of provision • access determined by a process of selection as children reach the age of eleven • high-stakes assessments leading to qualifications are thought to be unnecessary for the majority of the population • teachers viewed as the central actors in the administration of qualifications
2. The expansion of high-stakes assessment (mid 1950s to late 1980s)
• teachers do not achieve the role of running a school-based examination system envisaged in the previous period • numbers taking qualifications gradually increase • alternative qualifications are made available to a wider group of the school population • the practical implications of expanding the provision of schools serving different sections of the population make slow progress • growing opposition to the tripartite system of secondary education • new qualifications see a more central role for teachers • growing debate around the control of the curriculum
3. The age of public accountability and performativity (1988 – to present)
• a move to centralised control through the establishment of a statutory national curriculum and associated assessments • the establishment of an open system of school accountability based on test and examination outcomes • institutions and the individual become highly visible and objectified through the apparatus of surveillance such as test and examination data, school performance data and inspection reports • the role of teachers in high-stakes assessments diminishes

9.3.2. Episteme 2. The expansion of high-stakes assessment (mid 1950s to late 1980s)

The second episteme witnessed an expansion of increasingly high-stakes assessments and the gradual increase in the school population through free access for all to an extended period of schooling up to the age of 15. This was mirrored by a gradual increase in entries to public examinations and for a period of time, the extensive use of tests for the sole purpose of selecting pupils around the age of eleven for the now varying routes through secondary education offered by secondary modern, technical and grammar schools. As demonstrated in earlier chapters, the selection test for children aged around eleven years old, commonly known as the 11+, was not statutory nor was it derived from one centrally designated source. Norwood and the 1944 Education Act had set the foundations for a tripartite system of secondary school provision, but local authorities were at liberty to propose their own means of meeting the requirements of the Act with some favouring multilateral schools rather than the tripartite system (see for example Dent, 1944) whilst others held out for a system of comprehensive education which became a major development in the 1970s. Indeed the format of the 11+ was never clearly articulated and universally agreed and was envisioned by some, including Norwood, as an unspecified combination of tests and teacher assessments formed over a period of time: others became more reliant on tests, some of which were commercially produced (see for example Section 4.5.1.3 on the National Foundation for Educational Research (NFER) and Moray House in Chapter 4). There were also differences in how pupils passing the 11+ were allocated to secondary school places. The education system was slow to respond to the ambitions of the tri-partite system with the building of appropriate schools curtailed by post war austerity. This meant that despite the broad public usage of the terms 'passing' or 'failing' the 11+, the numbers 'passing' were more commonly determined by the availability of grammar school places.

However, entry to the newly created GCE O level was restricted by the fact that candidates had to have reached the age of 16 before they could enter for the examination at a time when the majority of pupils left school aged 15. This resulted in the proliferation of school leaving certificates designed and administered by local authorities as a means of validation. Such certification attracted no national currency reinforcing the view that students holding these certificates were viewed, in comparison to the relatively few students holding GCE qualifications, as educational failures (Tattersall, 2007). Lobbying by teachers was instrumental in this restriction being relaxed with increasing numbers of pupils in secondary modern schools being entered for the examination and following the

Beloe Report (1960), the introduction of the Certificate of Secondary Education administered by regional examination boards provided access to qualifications to a wider group of students and a system of external validation. This was despite the Education Minister's objection to the widespread use of external examinations (see Beloe, 1960; and Aldridge et al., 1991). The lobbying for change demonstrated a seminal act of power by teachers in the secondary phase of education as they sought wider access to examinations driven by the teaching profession, an opportunity missed in the early 1950s. As a result, teachers became key players on the regional examination boards even to the point of designing their own examinations (Mode 2 and 3) which were validated by the boards: this period has been described as the golden age of teachers' control of examinations (Le Grand, 1977, cited in Whitty, 2001, p. 161).

However, there were clearly rumblings of discontent amongst the public and the political classes over what was described as the 'secret garden' of a system of education hidden from public view "…nor that profane hands should be allowed to touch it" (Callaghan, 1976. See also Chitty, 1991). It is clear from a review of the literature and the findings in Chapter 4 that teachers had in fact gradually seized considerable influence and control over the general examination system and by default the school curriculum. Further, the system was more loosely regulated when compared to the current arrangements under Ofqual. But equally clear, teachers did not receive specific training in assessment theory and practice - even those designing and marking public examinations. Lead examiners for CSE and GCSE qualifications, those responsible for producing test papers, were generally recruited directly from the teaching profession, and selected on their classroom experience, subject knowledge and experience as markers. GCE A level followed a similar model, although some lead examiners were drawn from university departments. This is not to say that either group was necessarily deficient in expertise or that awarding bodies did provide technical support to some degree. However, examiners and markers were (and still are) all employed on an ancillary basis with no recognized qualification in assessment in a system described as a "cottage industry" by QCA's Chief Executive Ken Boston (Sutherland, 2008).

9.3.3 Episteme 3. The age of public accountability and performativity (1988 – to the present)

The outcome of the 'great debate' initiated by then Prime Minister James Callaghan's speech (1976) could be described as the foundation of the third episteme, the age of public accountability and performativity. Callaghan raised questions around a range of issues

including new teaching methods, concerns raised by industry about new recruits from schools being deficient in basic skills and why a high proportion of girls 'abandoned" science before leaving school; he cited these issues in his call for a public debate on the educations system. However, it took until 1988 before key reforms such as the statutory National Curriculum took hold followed by the first publication of school performance tables in 1992. Here, the educational assessment outcomes of tests and examinations became central measures in the system of school accountability. In the early 1990s, there was clear concern with regards to teachers' involvement in assessing general qualifications used to gauge school performance with suspicions over their motivation, for example as exemplified by the then Prime Minister John Major's assertion that teachers were in effect marking their own homework (Major, 1991); this intervention resulted in reduced weightings of coursework in GCE and GCSE syllabuses. At the same time newly developed externally set statutory National Curriculum assessment tasks were being introduced into schools following recommendations by the Task Group on Assessment and Testing (TGAT) in 1987: the assessments related directly to the content of the newly formed statutory National Curriculum. The standardised tasks were to be marked and moderated by teachers and administered over time in 'normal learning contexts': the approach was built on the recognition of teacher's professional skills and mutual support; "…giving them both key responsibilities and communal safeguards against idiosyncrasy" (TGAT, 1987, para.16). The reduction of teachers' involvement in qualifications on the one hand, but a key role in statutory tests on the other appears contradictory. However teachers did not make a play for this and in fact shunned their potential role in statutory tests on the grounds of increased workload through a series of teacher union boycotts and court actions in the early 1990s. This was interpreted by some as symbolic of a deeper rejection of the "marketisation" of education (Coles, 1994). The then Secretary of State for Education John Patten referred to the objection to tests as "Neanderthal", and following the Dearing Review (1994) a revised system of externally set and marked tests was introduced with test outcomes joining general qualifications in the 1996 performance tables published by the Department for Education. This was predicated on a view that the publication of performance data from an administrative authority has a positive effect on student achievement (see Gill and Benton, Chapter 1, Section 1.4.1) and despite challenges from teacher associations, performance tables based on test and examination results have remained as a key feature of the English education accountability system. This is in spite of the teacher boycotts and court actions of the 1990s that although reflecting Foucault's notion of 'power is everywhere', in that teacher actions resulted in modifications to the administration of the tests, the final result of these particular

skirmishes ultimately took teachers further away from the central role, and control, envisaged by TGAT (Whetton, 2009) without any real impact on the accountability regime.

9.3.3.1 The impact of accountability measures

Over the period I have described as the age of public accountability and performativity, there has been a clear reduction in teaches' involvement in statutory tests and examinations accompanied by increased emphasis on measures of accountability based on assessment outcomes. The teacher interviewees in this study were united by the view that accountability was the key driver of the education system in England and the cause of some frustration. Even so, the need for accountability was recognised by all teacher interviewees even to the point of JohnS supporting league tables. The Key Influencers were equally supportive of a system of accountability, for example JoPol who stated: "Oh, it's a good thing, accountability, absolutely" (JoPol p2/3) and LoganPol who recognised the need to hold schools and government to account. However, teacher and academic Key Influencer participants were more critical of the use of tests as a focal point of accountability than the majority of political or civil servant participants. Even here, the detrimental impact of using tests and examinations as a key measure was recognised by two of the politicians as exemplified by MelPol and JoPol who noted that assessment as a tool to support teaching and learning has been devalued and has undermined teachers' agency. Much of this criticism centred on their use to measure teacher and school performance noting the unintended consequences of narrowing of the curriculum and gaming the system - the latter noted throughout this study as a key driver of mistrust not only *of* the profession but *within* the teaching profession (NAHT, 2014). Trust in teachers' assessments is covered in more detail in Section 9.6.7 below. Further, there is evidence that the growth in the qualifications market and use of tests has driven a culture of intense competition as schools look to improve their position in performance tables and even amongst teenagers looking to secure work or university places (see Wolf, 2008). The increased competition amongst schools has fuelled suspicion over the reliability of teachers' assessment. The literature review cited studies reporting concern with the reliability of teacher's marking and over-assistance by teachers in their pupil's coursework (for example Johnson, 2011; HMCI, 2011; Ofqual, 2013; Opposs, 2016; Rimfeld et al., 2019) which has fuelled a system driven away from validity to one of increased reliability and manageability (Whetton, 2009). And despite studies finding concern with the quality of marking in tightly governed regimes of marking as used by awarding bodies and government agencies administering National Curriculum assessments (for example

Meadows and Billington, 2005; QCA, 2009; NFER, 2013), teacher associations critical of these regimes have as yet to present a tested and viable alternative system.

9.3.3.2 The lack of a unified teacher voice

As illustrated in Chapter 4 Section 4,5,3,11, the Assessment Charter developed by schools in Hampshire demonstrates the opportunity and freedom for schools to mitigate at least some of the unintended consequences identified above of a system built around assessment outcomes (see for example Pilgrim's Cross C of E Primary School, Hampshire). This has been further reinforced by changes in Ofsted's School Inspection Framework introduced in September 2019 with an increased focus on the quality of education and a decrease in what had been an over-emphasis on test and examination performance data. However, teachers lack a coherent voice or unified language with regards to educational assessment with several teacher unions and associations representing their views; a situation exploited by Kenneth Baker in 1988 (Cloake, 2017). The recently formed Chartered College of Teaching does however present a forum to develop professional discourse around educational assessment, but this organisation is as yet in the early stages of development.

Chapter 5 illustrated how power relations played a part in the way educational assessment has changed in the English system over time and how the teaching profession has both influenced and responded to national policy. Although the Teacher and Key Influencer Interviewees made little or no reference to changes in educational assessment beyond the last 30 years, there was a clear sense of change within their own periods of engagement with the education system. Much of this reflected Foucault's position that change does not assume progress (Foucault, 1994; Smart, 2002; Kendall & Wickham, 2003) and that changes in moments of history are contingent, that is, they are dependent on the events and conditions of everyday life rather than steps towards an ultimate vision of truth (see Turkel, 1990). Teacher interviewees did however make frequent references to changes from the time they were pupils noting significant increases in the use and impact of assessment as a means of accountability and a growing sense of pressure driven by the system of accountability typified by SaraS who stated: "…this is just what I remember but I might have forgotten, but it seemed a lot simpler than now". In reality, this level of change mirrors the intentions of the educational reforms of the late 1980s and early 1990s introduced to create a step change from the education system of the 1970s and 1980s. However, it also demonstrates how change can be introduced and extended within an identifiable epoch rather than being the outcome of a one off high profile event.

In short, epistemes are not clearly bounded. The 1988 Education Reform Act is arguably a 'one off' and highly significant event, but the intended reforms stuttered in the early years and issues of contention in the early 1990s still resonate today, for example in regard to the content and statutory nature of the National Curriculum, statutory assessments and performance tables.

As noted above, teacher interviewees supported the idea of schools being held to account and testing and examinations attracted relatively little technical criticism with some favouring more testing – albeit of a low stakes nature, for example JohnS and HelenP. This reflected the finding in Chapter 5 that schools purchased optional tests from QCA in large numbers (see Chapter 4 Table 4.5). However, there is need for some caution here in the teachers' ability to question the technical quality of test instruments in the light of their lack of training. For example, SaraS noted: "I don't think people (teachers) are very aware of how to put an actual assessment together": GrahamP recalled his 'shock' at the lack of assessment knowledge amongst his school colleagues and AngelaS made the assertion that; "It's clear from my own faculty that some people know diddlysquat about assessment and how it goes". Key influencers coming from an academic background however, were more likely to cite the technical limitations of assessment, for example around the concept of assessment error, unconscious bias or the limitations of assessment data - what JadenR referred to as 'noisy data'. But other than in the case of some politicians and civil servants, there was a point of unification around the limits to which test outcomes can be used as tools of accountability and the negative impact this can have on general perceptions of educational assessment: this was captured, perhaps ironically, by a politician Key Influencer who described the general perception of assessment: " …rather than assessment being seen as a tool to help children, I think it's seen as a fairly heavy handed check on teacher and school performance, and I think that's a problem" (JoPol).

9.3.4 Research question 1 summary of findings

Drawing on the work of Foucault as a framework of analysis (Chapter 4), this study has identified three educational assessment epistemes between the 1940s to the present day. And although changes from one episteme to another are blurred, each transition has seen the emergence of a power/knowledge dimension as educational assessment has become a central tool of school and teacher accountability, a technology, and an instrument of power that has governed the role played by teachers in its development, administration and use. This study has shown that such power is not top down but rather diffuse (see Foucault, 1976) as teachers have at times rejected opportunities to take more central roles

and pushed back against government policy, at local and national levels as in the early 1990s for example. Much of this rejection has been justified on the basis of concerns around teacher workload, but as the role played by teachers in high-stakes assessments has reduced, so has their level of control. This has been noted by Lawton (1980) in Chapter 4 which has led to a decline in their professional status and agency as reflected by HaydenPA and MelPol in the empirical data presented in Chapter 8. This study has shown how the role of teachers in high-stakes educational assessment has changed from an initial vision as presented by the Norwood Report that saw teachers as the key players in educational assessment to one where their input and influence has been marginalised. The move to a system of high-stakes accountability, for schools and teachers in particular, built around the results of tests and examinations has raised concern over the reliability of teachers' assessment judgements as exemplified by JadenR's view that, teachers are aware that the judgments they make on their students such as predicted grades are used to evaluate their own performance; this alone creates an atmosphere of mistrust. As a consequence, the assessment system has adopted procedures that favour a focus on reliability over the broader concept of validity. And whilst some would argue that more teacher assessment would increase validity (Wiliam, 2001; Harlan, 2007; and Johnson, 2013), government policy has moved in the opposite direction as shown for example by recent changes to the collection of teacher assessments at the end of Key Stage 2, and the removal or reduction of coursework and controlled assessments (Ofqual, 2011 & 2013; Opposs, 2016) with justification given via concerns over the reliability of teachers' assessments. The teachers who took part in the interviews for this research frequently spoke of assessment as something almost external to their day-to-day reality, an imposition of accountability or as GrahamP put it, something that "…is done unto you". Such views lead to present assessment as something external to teaching and learning rather than integral to the process. And yet a system with over thirty years' experience of an externally set curriculum and assessment regime arguably negates the need for teachers to have a deep understanding of assessment theory and practice. This set of circumstances may to some extent condition and explain the low focus on assessment theory and practice in initial teacher training discussed in Section 9.7 below. However, the under-utilisation of professional expertise and an atmosphere of mistrust between teachers, schools and policy makers undermine the notion of progress.

9.4 Research question 2: How do teachers and Key Influencers perceive educational assessment and the role played by teachers in the English education system?

To answer this question, this section draws on the data derived from interviews with six teachers and fifteen Key Influencers presented in Chapters 7 and 8 respectively.

9.4.1 Defining educational assessment

Despite constant references, there was no consistency across interviewees as to what actually constitutes educational assessment. The teacher interview participants more commonly described the purpose of assessment as a support for their teaching or as a means of certification rather than attempting to define it. The focus on assessments as being tests and examinations was cited by one academic (AshA) as a feature of English culture in a way that is detrimental to a broader understanding of assessment, with one teacher participant suggesting that most teachers see assessment as something that is "…done unto them" (GrahamP) rather than empowering them as professionals. It was common amongst teacher participants to find initial descriptions of assessment couched in terms of external tests or examinations rather than within the classroom context. However, in describing the prime purpose of assessment, all teacher participants expressed the view that assessment should be for the benefit of pupils. In speaking about definitions or descriptions of educational assessment, other than academic or regulatory participants, there was no consistency in the use of terminology associated with educational assessment. This was reflected in the literature review (Chapter 2), for example Newton's (2007) call for greater precision in the language of assessment. One other academic (AmarPA) expressed the view that although teachers in the primary phase are aware of the issues around assessment, they do not possess the terminology associated with assessment: "I think they would find it difficult to define validity, to define reliability, to define equity" (AmarPA). Referring to secondary phase teachers, AmarPA noted there was less involvement in the testing regime and a lack of effective assessment for learning, particularly in years 7, 8 and 9 where the influence of external tests and examinations was at its least. This view that assessment for learning was less than effective or misunderstood was a consensus view amongst Key Influencers and supported by half of the teacher participants. This raises questions about the quality of provision in ITT as 100% of respondents to the ITT Provider Survey reported the inclusion of formative (and summative) assessment as key principles of assessment being included in their

courses with 95% being moderately or very confident in their level of provision. However, fewer respondents to the Teachers' Survey reported the inclusion of these key concepts (73.9%) with almost a third feeling 'not very' or 'not at all confident' in these aspects of assessment. The teacher interview participants' views would seem to support the findings of the Teacher Survey with frequent references to the lack of understanding around the purposes of both formative and summative found in newly qualified teachers. For example HelenP recalled the lack of knowledge shown by recent interviewees for a post in her schools, and whilst AngelaS recalled that formative and summative assessment was part of her training, she noted that coverage was superficial. The underlying cause could also be one of a lack of clear terminology or language as identified by Newton (2007) in the literature review and suggested in the teacher interviews by AngelaS who noted that various elements of the content of her course were not explicitly linked to assessment.

The Key Influencers were equally critical of the level of teachers' general understanding around key assessment concepts with AliA suggesting that the lack of focus on assessment in ITT could be a reflection of trends in education. TerryA was equally critical about the level of understanding of assessment principles and referred to the wide variation of interpretation in aspects such as assessment for learning: Other reasons for the lack of understanding around key concepts in educational assessment may well be driven by the formation of a National Curriculum and the provision of an assessment regime, such as levels of attainment, that have negated the need for teachers to gain the basic skills and understanding and culminated in the loss of focus in initial training and an erosion of teachers' expertise. Such a view was expressed by LoganPol who proffered that the use of National Curriculum levels of attainment and the focus on formative assessment over a twenty-year period had driven out teachers' skills of testing.

9.4.2 Accuracy in assessment terminology

The lack of accuracy in the use of terminology was also apparent and reflects findings in the literature review (Stiggins, 1991 & 2014; Newton, 2007; Gardner, 2007; Klenowski, 2012; Looney et al., 2017): for example in the way assessment terms like 'moderation' and 'standardisation' were used interchangeably by teacher participants and reference to 'SATS' (which were discontinued in the mid 1990s, but the term is still generally used to describe statutory tests) was universal. As illustrated in the literature review, similar references are common in the mass media (see for example Busby, 2016). Some of this may be attributed to constant and sometimes subtle changes in the education system leading to mixed messages, misunderstanding and confusion, for example the changes to

the grading system in the GCSE (see Ofqual, 2015a), the move from coursework to controlled assessments and the removal of levels of attainment in 2014. The lack of accuracy amongst teachers with regards to key concepts in educational assessment is likely to be the result of the paucity of training in these facets as shown in the literature review and survey data: for example the Carter Review finding that assessment theory and practice is poorly represented in ITT, a view supported by Teacher Interview participants and the Teachers' Survey. This aspect is covered in more detail in Section 9.5 below.

9.4.3 Educational assessment in the English education system and the role of teachers

All interview participants spoke of significant changes in educational assessment practice since the 1990s, particularly around the increased use of assessment either within schools or across the system with the consensus being one of general support for educational assessment either as a tool to aid teaching and learning or as a means of recognising achievement. However, the key driver of change was universally noted as being the use of test and examination outcomes as measures of school accountability. Whilst the need for school accountability was recognised, the current regime was viewed as having limitations and the source of ever increasing pressure on schools, teachers and pupils to meet performance targets: this finding reflects much of educational literature presented in earlier chapters. For the Key Influencers with academic backgrounds and the teacher interview participants, concern was expressed that the purposes to which assessment outcomes are put have increased, driven and distorted by school accountability measures and the obsession with student, teacher, school and system performance data: a viewpoint commonly found in the literature (see for example Foley and Goldstein, 2012:Ball, 2013; Mason 2019).[60] It was also noted that assessments are frequently used for purposes for which they were not designed, a viewpoint found in the academic literature examined in Chapter 3 (see for example Newton, 2007). For teacher interview participants many of these changes have resulted in increased pressure and perceptions of decreased professional standing as noted in Section 9.5.1 above for example (see also O'Neil, 2002).

---

[60] Other Key Influencers were less inclined to note the changes in purpose or any resultant distortions.

9.4.4 Perceptions of pressure caused by assessment

The teachers' perceptions of pressure contrasted with their recollections of assessment as pupils, although their awareness clearly grew as they progressed through the system, in particular in relation to their studies for GC(S)E and GCE A level qualifications. As the teacher participants recalled their own experiences as pupils, there were few recollections of any pressure on their teachers or themselves during the primary phase of education. Indeed it was generally viewed as a relatively 'idyllic period' by HelenP and GrahamP in particular. However, there was clearly a growing awareness of assessment through their own experiences of coursework and public examinations. Whilst this was reported in terms such as being 'coached towards the exam', or in 'playing the game' to find extra marks in coursework, all interviewees put this down to self-imposed pressure for the purpose of progression to the next stage of education rather than pressure on or from teachers. This was more often contrasted with current practice, for example HelenP who noted that in her role as a senior leader in her school, she felt considerable pressure from the publication of school performance tables; this was in contrast to her recollections as a pupil. However, care should be taken in interpreting these results as young children may not perceive such pressure, and teachers may equally form a layer of protection for their pupils. References to pressure were also common amongst Key Influencers. Whilst academics put this in the context of pressure on teachers and pupils, politicians, particularly those who had held ministerial positions, spoke about their own feelings of pressure around the proportion of pupils meeting expected standards or closing the attainment gap for disadvantaged pupils as set by government targets: these concerns were expressed in the context of their own accountability to the electorate as policy makers and holders of the public purse (see LoganPol and JoPol in Chapter 8). References were also made amongst academic Key Influencers and regulators to the pressure on test and qualification designers to produce assessment instruments driven by the need to meet accountability rather than educational purposes. For example JadenR expressed the view that assessments might be designed in a different way were it not for their purpose as measures of accountability.

This suggests a system of accountability in the form of a recurring cycle of events, each impacting on the different constituents of the education system each one becoming dependent on the other. The interview data suggest that although participants held concerns over the current system of school accountability, there was no rejection of the need for accountability with some Key Influencer participants expressing views such as enduring teachers' resistance to national testing is a 'price worth paying' in respect of improving educational performance and identifying areas of failure; this resonates with the

motivation expressed by then Secretary of State John Patten (1993) identified in Chapter 4. Concern was expressed however by those participants with roles outside of government policy on the over-reliance on test and examination outcomes and more broadly across all interview participants on the resultant unintended consequences, for example narrowing of the curriculum or gaming the system: this reflected findings in the literature review (Spens Committee, 1938; Norwood, 1943; Gove, 2013; House of Commons, 2017; Ofsted, 2019). There were some expressions of optimism however in respect of policy changes such as more focus on curriculum breadth created by the Progress 8 measure and the re-alignment of Ofsted inspections from one heavily focused on performance data to one that focuses on the quality of curriculum provision and limitations on repeated entries for GCSE examinations.

Across all of the interviews, there was particular support for the use of tests at national and local level; this is consistent with findings in the literature review, and for example through the amount of non-statutory tests purchased by schools (p. 57). However, as reflected in the findings of research question 1, there were discrepancies over the purposes to which the outcomes are put and the over-reliance on test and examination data within the accountability framework. Politicians and policy advisers were more supportive of the current regime whilst in contrast practitioners and academics were more critical of the validity of much of the data used in the English system with references to the need for more scepticism of the veracity of data and improvement in the level of understanding amongst teachers of its limitations and implications. This supports the findings of the Teacher Survey with 49% of respondents stating that the effective use of assessment data was not included in their ITT course. Support for tests and examinations used in accountability measures amongst teacher interview participants frequently included preference for some level of teacher assessment in the process. Politicians and policy Key Influencers however tended to make a clear distinction between tests and other means of assessments used as measures of accountability or as qualifications and assessments undertaken by teachers; the general consensus here being that teachers cannot play a significant role in any high-stakes assessments. This was in contrast with the views of the primary phase teacher participants who favoured an approach using a combination of teacher and external test outcomes as currently used in Key Stage 1 statutory assessments (see HelenP, TimP and GrahamP in Chapter 7). However, although assessments carried out by teachers were viewed as important by Key Influencers as a means of producing information to be used within schools to inform teaching or to pass information between teachers and schools, there was no suggestion of their use in accountability measures.

9.4.5 The influence of personal experience on teacher interviewees' assessment practice

One aspect emerging from this research in regard to the teacher participants' recollections of their own experiences as pupils was the impact it has had on their own teaching practice. For example, GrahamP spoke of the influence of his mathematics teacher and the elation of passing A level maths, a feeling he wants to pass on to the children he now teaches. JohnS and AngelaS both spoke of their early fascination with the way examination papers were constructed and marked, both reporting that they sought examination syllabuses and past papers, an action that has impacted on the way they now construct their own internal assessments. Further references were made by teacher participants about their efforts to reduce or minimise pressure on their students, especially in light of their own experience as pupils and in comparison with current concerns over student wellbeing (see for example: Busby, 2016; Education Select Committee, 2017). This supports the findings of Looney et al (2017) that teachers understanding of assessment goes beyond knowledge and skills to one that is influenced by a structure of emotions and beliefs formed through their own experiences of being assessed.

9.4.6 The impact of accountability measures on assessment practice

Teacher and Key Influencers referred frequently to the impact of accountability measures on the day-to-day practice of teachers in schools, a constant theme being that of 'teaching to the test': this is consistent with finding in the literature review and other documentation such as Ofsted reports and commentaries by Her Majesty's Chief Inspector of Schools. One Key Influencer (AshA) expressed the view that all discussions related to assessment are set in the context of accountability. Much of this centred on the generation of data based on assessments for purposes of performance management within schools and comparison with other schools through what were commonly described as 'league tables' by interview participants. Assessments conducted within schools were frequently used to predict the ultimate performance of pupils in external tests and examinations using data 'flight paths' to monitor and predict the progress of pupils. However, there was concern expressed by teacher participants that there was too much emphasis on summative rather than formative uses of assessment outcomes, although references were made to the possible uses of item analysis data from summative assessments as a means of evaluating the effectiveness of teaching. Where internal data were used more judicially to inform teaching and learning or to help students prepare for tests and examinations or to manage teacher workload, there was support for the approach from teacher participants,

but concerns were again raised about the validity of such data and limitations in teacher's understanding of these data.

There was strong support for the use of externally produced standardised tests within schools which reflects findings in Chapter 4 regarding the numbers of tests purchased by schools (see Table 4.5). Teacher participants frequently suggested greater use of testing, but of a low stakes nature and purely for use in guiding teaching and learning. There was also some support for a regime of external moderation between schools - but with added kudos for those taking an active role as moderators. This approach was in line with the system proposed by TGAT in 1988, but rejected by teacher associations as it was viewed as an increase in workload. Some of this enthusiasm was tempered by caveats over the need for training and better provision of supporting materials such as well-designed marking schemes and exemplification of standards of performance. Such a system was suggested by teacher participants as being potentially 'powerful' (for example AngelaS) in establishing professional trust and as a means of supporting improvements to the teaching and learning process. However AngelaS countered this view by supporting the removal of teacher involvement in coursework, for example over or suspicions of cheating and gaming, caused largely by the pressure of accountability rather than a consequence: "…of the pressure on teachers to produce data rather than a real desire to be devious" (AngelaS). JohnS and SaraS also raised concerns about coursework including over-aiding by teachers and gaming. These viewpoints align with findings in the literature discussed in Chapters 2 and 4. Three teacher participants described teachers' practice as being driven by competition with other schools, a situation described by GrahamP as one that works against collaboration: all three noted an increase in competition between schools over their time as teachers. This viewpoint would seem to challenge the earlier suggested proposal for a system of inter-school moderation.

9.4.7 Trust in teachers' assessments

All teacher and Key Influencer interview participants raised concerns that teacher assessment is not trusted for example by politicians or awarding organisations, an outcome reflected in the literature review and surveys. Discussions around trust in teachers' assessment produced a tangible level of intensity from both teacher and Key Influencer participants. For example GrahamP who stated his goal as to make sure that teachers' assessments are trusted and empowered to make summative high-stakes assessments. But this was not universal amongst the teacher participants as JohnS noted that any form of teacher assessment must be confined to low stakes because of concerns

over their trustworthiness. As for the Key Influencer participants, AliA proffered that the lack of trust in teachers' assessment is the reason why awarding bodies were formed in what AshA described as 'a low trust system' of education in England.

However, this study has shown more nuanced perceptions of trust. Whilst the evidence broadly supports the findings of NAHT Commission on Assessment (2014), this research project has shown that levels of trust are differentiated across constituent groups (see Chapter 6 Teacher Survey Data analysis). The majority of teacher interview participants reported no issues with trust within their own institutions, and that with appropriate training and changes to the accountability system, trust could be restored across the teaching profession. This reflects the finding from the Teacher Survey in which 93.8% of teachers stated that they felt trusted by their immediate colleagues. However, two secondary phase interview participants, JohnS and AngelaS did raise concerns over trust within schools in relation to the assessments made by teachers, for example the possibility of inflating results to meet performance management targets. Issues of trust were more commonly related to concerns about other schools, for example accusations of cheating in public examinations and the lack of trust between infant and junior schools as exemplified by TimP who noted that children are re-tested on entry into his junior school to form a baseline assessment. Similar actions were also highlighted as children enter the secondary phase as a sign that teacher assessments and National Curriculum external Key Stage 2 test outcomes are not trusted – as further illustrated by SaraS who noted that key stage 2 assessment; "doesn't correlate" with pupil's actual performance in the first stage of secondary education. A similar picture is presented by the responses to the Teacher Survey in which 38.3% of respondent stated that their assessment judgments were not trusted by colleagues in other schools. Primary school interview participants suggested this could be caused by the gap between the results of National Curriculum tests held in May and what HelenP described as; "quite an easy run in from May to July" impacting pupil's performance the following September as judged by secondary phase teachers. Teacher participants from the primary phase expressed the view that they would 'like to think' that their assessments were trusted by secondary phase teachers, but the fact that secondary schools carry out baseline assessments in year seven undermined such trust. SaraS held the view that supervision of GCSE examinations was more robust than the invigilation of Key Stage 2 tests, although JohnS disagreed stating his concern that secondary teachers 'cheat' or game the system, particularly in coursework. This view is echoed in the literature review, for example Ofqual (2012).

However, all issues concerning the lack of trust raised by interviewees were referenced to the adverse impact of the school accountability system of performance tables, school inspections or performance management within schools; this view was broadly held amongst Key Influencers. Teacher interview participants reported that malpractice is often driven by suspicions that 'other schools do it' and feeling mistrusted by external agencies, particularly the Standards and Testing Agency (STA) and Ofsted. This was reflected in the survey of teachers where 56.5% of respondents reported that they felt mistrusted by the STA and 50.7% mistrusted by Ofsted. It is of some note that 85.6% of respondents in the Teacher Survey felt that politicians did not trust their assessment judgments, although one of the politicians in the Key Influencer group suggested this was fuelled by messages from politicians that undermined any level of trust and may well seep into the consciousness of teachers. Indeed, AngelaS, one of the teacher participants, raised doubts over trusting her own assessments, though this was referenced to a lack of assessment knowledge and training. However feelings of not being trusted went beyond politicians with 52.5% of Teacher Survey respondents feeling mistrusted by awarding bodies in terms of their coursework assessments: issues around the reliability of teachers' assessment may account for some of these reactions (see Opposs and He, eds., 2012: Ofqual, 2013: Rimfeld el al., 2019). But the issue of trust within and external to the teaching profession raised as a concern by the NAHT Assessment Commission (2014) is clearly reflected in the views of participants in this study as illustrated more broadly in the literature review (see ATL et al., no date; Brookhart, 2011; Gardner, 2007; Harlen, 2005; Johnson, 2013; Looney et al., 2017; Stobart, 2008; Wolf, 2008 & 2014; and Green, 2007). It also underpins recent moves in National Curriculum assessments and general qualifications to reduce or remove teachers from the assessment process. This has gains in terms of the reliability of assessments but has costs in terms of reduced validity as the focus of assessments moves to that which can be assessed through a written test. The political decision to remove external testing of writing in Key Stage 2 provides a clear example here: the test was replaced by teacher assessment, but the outcomes are not used in school accountability performance measures. Further references were made to the government's decision to ease teacher workload by ceasing the collection of teacher assessments, currently a statutory requirement, even though the data are not used for accountability purposes. The broader impact of these moves was viewed by some interview participants as reducing the standing of teachers as professionals.

9.4.8 Trust versus recognition of fallibility

The interviews also raised some concern over the use of the term 'trust' with JadenR suggesting that there needs to be greater recognition of the fallibility of human judgments, especially where such judgments impact on the students and institutions in which teachers work: this, it was suggested, should generate the formation of a system that protects the integrity of such judgments rather withdraw them from the assessment process. In essence, teachers are conflicted by a system that expects objectivity as they act as agents of examination boards or other external agencies on the one hand and being measured and held to account by the outcomes of their own assessments on the other; this was considered as unfair pressure by some Key Influencer interview participants, for example JadenR and TerryA. However there were widely differing views on how to deal with these conflicts ranging from increased use of teachers' assessments, popular with most teacher participants, supported by better training and systems of moderating teachers' assessments, through to the total removal of any teacher involvement in tests or examinations used in any form of measures of accountability, generally more popular amongst Key Influencers, though supported by JohnS, one of the teacher participants. Reaching any form of consensus appears distant.

9.4.9 The expected future of educational assessment in England

Teacher interview participants were generally resigned to the current state of assessment in schools in England and their own roles in the system. Most teacher interview participants expressed the view that they had little if any influence on national policy and that in reality, they adopt the procedures as found or dictated by their own schools. There was also more demand for stability rather than fundamental change, though some concern was raised around what one teacher participant, JohnS, described as the "haemorrhaging (of) expertise" as more experienced teachers leave the profession. This latter point was linked to challenges in finding assessment expertise, for example in local authorities, and in the growing number of multi-academy trusts with a focus on competition that work against the sharing of knowledge. This reflected a general feeling of despondency amongst teachers regarding any thoughts of fundamental change perhaps typified by SaraS who spoke of the energy of senior leaders in schools focusing on how to meet expectations rather than how the system could differ. Noting that she could not recall any discussions amongst teachers in regard to possible alternatives, SaraS questioned the level to which teachers feel equipped to express any opinions on assessment. However, there were mixed views amongst teacher interview participants regarding perceptions of expertise in assessment,

particularly within their own schools. Two participants were very clear that assessment knowledge in their own schools was good compared with other schools they had visited or knew; a view that reflected the findings of the Teacher Survey. Conversely, two other participants raised doubts over the levels of expertise found in their own schools with a further participant, AngelaS, raising doubt in her own knowledge of assessment theory and practice.

9.4.10 Access to continuous professional development

There was a more general feeling that assessment knowledge amongst teachers was often superficial and in need of further professional development and support, such as exemplification materials produced by national bodies: this viewpoint was common across teacher and Key Influencer participants. However, access to high quality professional development was clearly an issue for the majority of teacher interview participants, a viewpoint supported by four of the Key Influencers, for example ChrisR who raised concern that access to CPD is very limited and frequently provided by "…people who have just come through the system and think of assessment as putting those numbers on those spread sheets". TimP could not recall having access to any external CPD on educational assessment suggesting that the lack of financial resources and reduced provision in support from Local Authorities were key factors. The lack of general availability has resulted in reliance on in-house support, which is clearly dependent on the level of expertise available within an institution, or in the case of AngelaS, a teacher interview participant, a self-funded Masters course. JohnS was of the view that schools do not have sufficient expertise to run in-house training and doubted that there was any real incentive so to do. The Teacher Survey supports JohnS's assertion that levels of in-house expertise in assessment are low with 80.8% of all respondents agreeing or strongly agreeing that there are significant gaps in the capacity of schools in theoretical and technical aspects of assessment: 66.6% of the ITT Provider Survey were also in agreement. This also supports the findings of the Carter and McIntosh Reviews. This raises important questions about the quality of in-house professional development in assessment at a time when access to high quality external support is in decline. Concerns raised by Key Influencers over the lack of a community of assessment experts in England further compound this issue.

Recent moves by government agencies to improve teacher training and development through the development of the ITT Core Content Framework and Early Career Framework are a clear and much needed response to the criticisms levelled by the Carter Review. However, based on the evidence in this research, the reliance on mentoring and

support from expert colleagues with sufficient expertise in educational assessment will prove a challenge. The interviews with Key Influencers and teachers and the Survey of ITT providers all suggest a lack of expertise in educational assessment that the system of mentoring will need. Of further concern is the finding in the Teacher Survey that similar expertise amongst ITT providers is also perceived as lacking. So while the need to improve the level of training and development around educational assessment has been triggered, the extent to which the aims can be met is dependent on access to expertise that will be required to turn the frameworks into structured courses of instruction: at the present time such levels of expertise are in short supply.

9.4.11 Research question 2 summary of findings

To answer research question 2, data from surveys with teachers and ITT providers was combined with face-to-face interviews with teachers and Key Influencers in the English education system. The data has shown that teacher interviewees and Key Influencers showed little consistency in what the term educational assessment constitutes with teachers predominantly associating assessment with external forces driven by a system of high accountability. The lack of an agreed and universal language to discuss educational assessment as identified in the literature review (Chapter 2) and the paucity of provision in regard to assessment theory and practice as noted by the Carter Review was borne out in the teachers' survey and interviews (Chapters 6 & 7). Although the interviews and Teachers' Survey supported Carter's view that there was a lack of assessment knowledge amongst ITT providers, the Survey of ITT providers (Chapter 5) presented an opposite viewpoint: the reasons for this require further research. However, the data did show broad agreement that assessment expertise is lacking in schools.

All six teacher interview participants spoke of significant change to their roles in high-stakes assessments over their time as teachers noting increased pressure with the accountability system as the key driver of their diminished role: this was noted by both teachers and several Key Influencers as the cause of teachers' reduced agency and perceived professionalism. For the teacher participants, reflections on their own time as pupils played a significant part in their evaluations of current practices, a feature not broadly recognised in the literature. Despite any misgivings about the use of assessment outcomes in the accountability system, there was widespread support for the use of assessment as a tool for teaching and learning and the use of tests in schools, though most preferred such use to be in a low stakes environment. This aspect gave light to the high numbers of tests purchased by schools as identified in earlier chapters and would

support the vision of the TGAT Report that was rejected by teacher unions in the early 1990s.

The lack of perceived trust as identified by the NAHT Commission on Assessment was a consistent feature of the data shown in Chapters 5 and 6. However, the data in this study has shown the levels of trust manifest in the teaching profession are complex with relatively high trust amongst immediate colleagues, but perceptions of low trust shown by politicians for example (see Figures 6.10 & 6.11). As with other issues identified in this section, accountability was frequently cited as the source of mistrust and there was a level of resignation amongst the majority of interviewees that despite their varying views on high-stakes assessment, any change was distant.

There was a consensus amongst all interviewees that teachers' current assessment knowledge is superficial and that access to good quality training and development is lacking and dependent on internal provision. Given the current levels of expertise, frequent concern was raised about the ability of schools to provide good quality training and development. It could also be argued that given the limited role played by teachers in high-stakes assessment, the need for initial training may be reduced. However, this overlooks the critical importance of assessment in the teaching and learning process. The next section of this chapter addresses research question 3 by focusing on the ITT curriculum and the level to which teachers feel prepared in the field of educational assessment for their early years of teaching.

9.5 Research question 3: What does the ITT curriculum in England offer in terms of assessment theory and practice?

The literature review in Chapter 2 demonstrates a lack of research and knowledge of the curriculum offered by ITT providers in England (Sheffield Institute of Education, 2019; HMCI, 2019) and the materials submitted to the Carter Review (2015) by 145 ITT providers showed a paucity of detail in relation to educational assessment theory and practice (DfE, 2015c). Research Question 3 aims to gain such a detailed insight into the content provided in ITT courses and the views of teachers, teacher educators and Key Influencers on the extent to which it presents an adequate grounding for teachers entering the teaching profession in the maintained sector of the English education system. Question 3 considers the perceived quality of instruction in educational assessment in initial teacher preparation through perceptions of confidence in the level of instruction, an aspect not covered

explicitly by the Carter Review, at a time of increased diversification of routes into teaching and the way educational assessment is used in the English education system.

9.5.1 Aspects of educational assessment included in ITT courses

The analysis of responses in the ITT providers survey suggest that the content referred to in the questionnaire is included in most initial training courses, with some aspects included in all – for example the fundamental principles of assessment and the effective use of assessment data to ensure that teaching is both supportive and sufficiently challenging. However, the responses received in the teacher survey did not reflect the level of provision suggested by ITT providers. There were also areas covering key concepts in educational assessment such as bias, validity, and reliability where around one–third of providers stated these aspects were not included in the courses: again, figures from the teacher survey reported even lower returns on the inclusion of these aspects: see Table 9.2 below: (see also Chapters 5 and 6 data analysis).

Table 9.2: A summary of the percentage of respondents who answered that each aspect of educational assessment had been included in their ITT course by survey: teachers versus ITT providers' views
Note: All of these differences in percentage responses between groups are statistically significant at the 5% level (using a chi-square test of [61]association). See Appendix 11(a) for details.

Aspect of educational assessment	ITT providers survey response	Teacher survey responses
The fundamental principles of assessment	100% (n=47)	73.9% (n=199)
The effective use of assessment data to ensure that teaching is both supportive and challenging.	100% (n=49)	51% (n=202)
Bias	68.1% (n=47)	17.8% (n=197)
Validity	72.3% (n=47)	24.2% (n=198)
Reliability	70.2% (n=47)	26.8% (n=198)
Setting standards of educational performance	79.2% (n=48)	26.3% (n=198)

---

[61] Statistical Test using Medcalc. MedCalc uses the "N-1" Chi-squared test as recommended by Campbell (2007) and Richardson (2011).
The confidence interval is calculated according to the recommended method given by Altman et al. (2000)
https://www.medcalc.org/calc/comparison_of_proportions.php

These figures suggest a series of contradictions in the responses to the surveys in this study. For example, whilst providers' responses suggest 100% inclusion of the fundamental principles of assessment, key principles like validity and reliability appear to be missing in courses from almost a third of provider responses. And although providers show relatively high levels of inclusion of aspects of assessment theory and practice in their provision, this is not reflected by the responses in the teachers' survey as illustrated in Table 9.2.

There is also some disconnect between how confident ITT providers were that they had fully prepared trainee teachers and how confident teachers felt on completion of their courses (see Chapters 5 and 6 data analysis). For example, while ITT providers and teachers were most confident regarding 'the differences between formative and summative uses of assessment', 75% of ITT providers felt 'very confident' they had fully prepared trainee teachers in this aspect compared to 28.2% of teachers (the highest 'very confident' ratings on the teacher survey). The surveys showed that ITT providers were generally more confident that their course prepared their trainees in key aspects of educational assessment than the teachers. While over 39% of respondents to the teacher survey were 'not at all confident' their ITT course had fully prepared them for their first years in teaching in terms of key assessment principles, ITT providers reported a different picture as shown in Table 9.3 below.

There are other examples of divergence in terms of levels of confidence. For example, whilst 72.9% of ITT providers were moderately of very confident in the use and misuse of assessment outcomes and the ethics of assessment, this compared with only 28% of teachers. And in the case of using relevant data to monitor progress, 97.8% of ITT providers were moderately of very confident compared to 51% of teachers.

Care needs to be taken in interpreting these results. For example, to some extent this disconnect may be due to the different periods in which respondents to the teacher survey had trained. This difference could also be influenced by ITT providers reflecting on current provision, rather than reflecting on course content over time in their responses (with 33.2% of respondents to the teacher survey having trained prior to 2000). Additional explanations could also be a likely gap between what one presumes has been learned because it has been taught; retention of assessment-related principles over time; defining and understanding assessment-related principles as separate/more theoretical compared to the everyday practice of teaching and the differences in the language around assessment – as highlighted by Newton (2007).

Table 9.3: A summary of the varying levels of confidence in the preparedness of newly qualified teachers: teachers versus ITT providers' views

Note: All of these differences in percentage responses between groups are statistically significant at the 5% level (using a chi-square test of [62]association). See Appendix 11(b) for details.

Key assessment aspects	% of teachers' levels of confidence that their ITT had fully prepared them for entering the profession.		% of ITT providers levels of confidence that their courses prepare newly qualified teachers.	
	Not at all or not very confident	Moderately or Very confident	Not at all or Not very confident	Moderately or Very confident
Criterion- and norm-referencing	70.5	27.5 (n=120)	42.4	57.6 (n=33)
Utility	76.6	23.3 (n=120)	41.9	58.3 (n=36)
The design of standardised tests	72.8	27.2 (n=125)	41.7	58.3 (n=36)
Validity	68.9	31.1 (n=122)	36.6	63.4 (n=41)
Bias	74.2	25.8 (n=120)	37.5	62.5 (n=40)
Reliability	67.5	32.6 (n=123)	35.0	65.0 (n=40)

However much of this reflects other findings in the literature review relating to the lack of knowledge and confidence found in the teaching profession at present, for example Johnson, (2012), McIntosh (2015) and Ofsted (2020). This may also be confounded by the lack of agreement around the detail of what constitutes an acceptable or desirable curriculum of assessment theory and practice (Ingvarson and Rowley, 2017), or the "…significant variability in ITT courses" as identified by the Carter Review (2014, p. 5). Of course more recent DfE instigated initiatives to clarify and improve understanding of the desired content of ITT and support for newly qualified teachers in respect of educational assessment may address these issues in the longer term, but the expertise on which such a system depends is scarce as found in the interviews with teachers and the literature review (for example Stiggins & Conklin, 1992; Green, 2007; Stiggins, 2008; Isaacs, et al., 2013; NAHT, 2014; McIntosh, 2015). Further, the different routes into teaching bring to bear differing pressures on what can actually be achieved in terms of curriculum content within the initial stage of teacher education with ITT providers facing the pressure of what

---

[62] Statistical Test using Medcalc. MedCalc uses the "N-1" Chi-squared test as recommended by Campbell (2007) and Richardson (2011).
The confidence interval is calculated according to the recommended method given by Altman et al. (2000) https://www.medcalc.org/calc/comparison_of_proportions.php

can reasonably be expected in the time available for instruction: this aspect was frequently identified as a constraint in the ITT Providers' Survey and recognised by Key Influencers with experience of teacher education. Schools will also find challenges in designing the level and intensity of support for newly qualified teachers graduating from differing ITT routes each presenting differing levels of engagement with assessment theory and practice.

The teacher interviews showed some ambivalence amongst participants in their views on their preparedness for teaching in regard to educational assessment. In several cases this was a reflection of their route into teaching. For example TimP trained as a primary teacher through a Graduate Teaching Programme (GTP), so although he stated that on reflection he was not prepared in terms of educational assessment, he was at least familiar with the processes used by the school in which he trained and he felt confident in asking other staff for support. Two other participants stated that they were not at all prepared but left to develop their own understanding over time with one other noting the positive influence of her mentor who by chance had an interest in assessment whilst another noted his training coincided with the introduction of major new government policy and therefore more emphasis during his PGCE course. For one of the secondary phase teachers, the support she received from her mentor during training was deficient, but a later move to another school where assessment practice was better established provided real opportunity to develop in this aspect. However, one senior secondary phase teacher participant whilst recognising the importance of the role of mentors, noted that in light of other pressures experienced in schools, the role of mentor "…goes down to the lowest rung" (JohnS).

9.5.2 The Carter Review – ITT content and levels of expertise in educational assessment

There was broad support in both surveys for Carter's assertion that in all areas of ITT content, the most significant improvements are needed for training in assessment with 59.2% of providers in agreement, and 83% of teachers. There was also broad agreement that there are significant gaps in the capacity of schools regarding assessment, with almost 67% of providers and 81% of teachers in agreement. However, there was a clear divergence in views around the capacity of ITT providers with 73.5% of ITT providers disagreeing with Carter compared to 71.6% of teachers who did agree with Carter. Further divergence was found in responses to questions regarding the levels of confidence that ITT institutions had the right level of expertise to deliver all aspects of educational assessment theory and practice to the same depth and quality as all other aspects of their

ITT course with, over 95% of ITT provides reporting they were confident in their level of ability compared with just under 56% of the teacher survey respondents (see Sections 5.3 & 6.3).

The weakness of knowledge around educational assessment amongst the teaching profession was a unifying issue in the interviews with Key Influencers with some pointing to the irony that despite the centrality of assessment to teaching and learning and the accountability system, knowledge remains scarce. This was often exemplified by the reference to key aspects of assessment such as validity, reliability and the use of assessment data – aspects also found wanting in the Teachers' Survey. The low number of assessment experts coming out of or working in the higher education community was also raised as an issue, an issue that was recognised as a challenge for national agencies in recruiting personnel with the right levels of expertise. However, there were no explicit references from Key Influencers to the lack of capability amongst teacher educators in terms of their knowledge of educational assessment theory and practice.

The majority of Key Influencers were nevertheless of the view that initial teacher training is deficient in preparing teachers in terms of educational assessment: this reflects findings in the literature review (see Stiggins and Conklin, 1992; NAHT, 2014; McIntosh, 2015). However, a range of opinion was expressed as to why this is. Some were of the view that such knowledge is of little need in a system where assessment is driven by external forces such as the preparation of tests and examinations by external agencies and the reduced participation of teachers in high-stakes assessments such as in coursework. One leading academic expressed the coverage of educational assessment theory in ITT as:

> "…dire. Or you might actually say it is hopeless but not serious, because teachers know nothing about assessment but it does not matter because the only assessment that is done is done by external agencies" (PatA).

There were other references from Key Influencers to teachers viewing assessment as being external or 'done' to them, a sentiment expressed by participants in the teacher interviews and found in the literature (Stiggins, 2014). Frequent change to government policy was also cited as an inhibiting factor, although there was recognition that despite the general lack of instruction around summative assessment, more attention has been given to formative assessment - though even here it was thought that transient government support for this aspect was somewhat dependent on whether it was as one teacher Interviewee AngelaS put it, "flavour of the month" or not. The ITT Providers'

Survey however showed that 100% of respondents included coverage of formative assessment in their courses, although a quarter of respondents in the Teachers' Survey stated that they did not experience such with around one-third expressing the view that they were 'not very' or 'not all confident' in this aspect of assessment. The teacher interviews also raised some concern about teachers' depth of understanding in formative assessment with interviewees speaking of teachers having only superficial knowledge, a concern noted in the literature review (see Section 2.13). One teacher interviewee also raised the possible impact of teachers' subject knowledge on the way assessment was used or understood in the primary phase of education. This was based on the demands of teaching a wide range of subjects with limited expertise in at least some of the subjects leading to a lack of understanding in what or how to assess in these subjects or in terms of what standards of performance should be expected from pupils. These circumstances put considerable pressure on teachers to design and implement assessment strategies in subjects in which they are un-familiar: this has the potential to undermine their confidence and lead to poor assessment practice. One secondary phase interviewee raised a similar link between subject knowledge and assessment practice, an issue identified in the Carter Review (2015). In this example, the interviewee expressed the view that some teachers were reluctant to use deeper questioning techniques used in formative assessment practice as this may expose the lack of deeper subject knowledge held by the teacher.

9.5.3 Further development of teachers' understanding in assessment

The teacher interview participants all spoke of the lack of access to professional development in educational assessment in the early and indeed subsequent years of their teaching. However, as with recollections of initial teacher training, assessment content may well have been 'hidden' behind or not recognized as being within training sessions. For example one secondary teacher recalled attending a school training session on running mini-plenaries within lessons that on reflection; "…probably got people to start thinking about formative assessment" (AngelaS). Two other teacher interviewees spoke of work as a GCSE marker or Key Stage 2 moderator as useful personal development, but each was of the view that this was reliant on their own self-reflection. Both viewed this more generally as a missed opportunity as awarding organisations and schools could be more pro-active in turning such experiences into professional development. It was clear from speaking to all participants that most is if not all continuous professional development is now more heavily reliant on internal provision and therefore dependent on sharing existing expertise with staff in their own schools. And for one interviewee, 'professional development' often dissolves into more general issues faced by the school rather than a

focus on developing assessment expertise. It was clear however, that all participants had a 'real thirst' for professional development but a combination of competing priorities, the cost of CPD and the availability of quality courses (internally or externally sourced) were real barriers. These views reflect the findings of the LKMCo/Pearson review (2017) that a fifth of teachers didn't know where to look for information on assessment. Of these factors, the shortage of finance to access CPD was universal amongst teacher interview participants. This was also a feature raised by Key Influencer participants, for example MelPol, who referred to the 'slash and burn' approach of government support for local authorities or as identified by SamCS, that professional development is not 'particularly focused' on assessment, despite some obvious potential advantages. Other Key Influencers spoke of the lack of quality in CPD around educational assessment with the focus being more often on the collection of assessment data in preparation for Ofsted inspections rather on the fundamental principles such as reliability or validity. This was viewed as more often lead by those who as products of the current system understood only the mechanics of data collection rather than the fundamentals of assessment or issues relating to the validity of data. However, as with the teacher interview participants, there was broad agreement that CPD should be a constant feature of the teaching profession with some supporting the view that it should be a more obligatory requirement, but one that is driven by the profession.

9.5.4 The varying demands on ITT course provision

The ITT Providers' Survey illustrated a level of frustration in terms of what can be reasonably expected in terms of course content related to educational assessment theory and practice. The desirability of what could or should be included in ITT is clearly regulated by a number of constraints. The amount of available time is a central issue as educational assessment 'competes' with other elements of instruction such as behaviour management, subject knowledge and broader aspects of pedagogy. This issue was frequently cited by ITT providers but equally recognized by teacher interview participants and Key Influencers. For most, this underpinned recognition and support for further professional development particularly in the early years of teaching.

However, this also raised challenges such as the capacity of schools and ITT providers to offer the required level of expertise and the varying approaches to assessment found across schools – the latter intensified by frequent changes to educational policy. The pressure of time on ITT providers, especially those offering courses with duration of one year or less was viewed as being particularly problematic. But the notion of preparing

students to accommodate the demands of varying systems used by schools is equally of concern. It is clear from the literature review that the removal of National Curriculum levels in England has resulted in a proliferation of assessment systems (McIntosh, 2015: Boylan, 2016; DfE, 2016d) used by schools and this has an impact on preparing students to enter the profession. For example, those undertaking initial training through a university based route may well experience two or three teaching practice placements in schools that have adopted various systems to generate and record assessments. However, those undertaking a school based route or one within an academy chain, may well experience different schools, but all using the same system. This issue was raised by TimP, one of the teacher interview participants, who noted little difficulty in understanding the assessment approach used by the school in which he trained through a Graduate Teaching Programme (GTP) and then took up full-time employment. However, on reflection, he concluded that what he actually understood was the operation of the system of assessment rather than the fundamentals of assessment. This raises an important issue: it would appear that training is more likely to focus on the adoption of prevalent 'assessment systems' that are in reality administrative rather than academically sound approaches to educational assessment: as result, teachers are in effect acting as 'task mangers': see Section 2.13 (see also Twiselton, 2000, 2004 and 2007).

9.5.5 Research question 3 summary of findings

This section has looked in detail at perceptions of the content relating to educational assessment offered in ITT courses in England. The data produced in Chapters 5 & 6 makes a contribution to filling the gap in knowledge of ITT curriculum content identified by the Sheffield Institute of Education (op cit) and adds to the limited evidence submitted to the Carter Review. The surveys and interviews used in this study were designed to gain an understanding of the range of educational assessment aspects covered in ITT along with the levels of confidence in the extent to which newly qualified teachers were prepared for the classroom in these aspects. The aspects included in the surveys were designed to cover a wide range of key concepts in assessment theory and practice.

The data has shown that despite 100% of ITT provider survey respondents stating that their courses covered the 'fundamental principles of assessment', almost a third of respondents did not cover key aspects such as bias, validity and reliability: this appears to be contradictory. According to the teacher survey, these aspects were reported as not being covered in their training courses by over two-thirds of respondents. There were also similar discrepancies between providers and teachers in the levels of confidence in the

extent to which ITT graduates were prepared for teaching in relation to aspects of educational assessment. Whilst some care needs to be taken in interpreting these data, as for example the teacher respondents trained at different times over a thirty-year period and the ITT providers could be reflecting on current practice, the lack of knowledge and confidence expressed by teachers does reflect the findings in the literature review.

More recent initiatives taken by the DfE to strengthen the Teachers' Standards (DfE, 2011c) and the ITT Core Content and Early Careers Frameworks (DfE, 2019b) should help to address some of the shortcomings identified in Chapter 1 of this research project. However, much of this will be dependent on the knowledge of current practitioners and as this study has shown, knowledge in educational assessment is generally lacking. Other pressures on the system were also identified such as the amount of time allotted to ITT that limits the extent and amount of coverage on a range of aspects required in preparing teachers for the classroom. This is further complicated by the varying routes into teaching some of which are highly dependent on the levels of expertise in schools offering courses or placements. The variation in provision in terms of course content and the quality of mentoring support in different routes into teaching were highlighted by the teacher participant interviews.

The data from this study broadly support the findings of the Carter Review that the most significant improvements needed in ITT provision are in aspects of educational assessment with almost two-thirds of ITT providers and over 80% of teachers in agreement. There was also broad agreement that there are significant gaps in the capacity of schools around educational assessment. However, there was a clear discrepancy in the capacity of ITT providers in terms of their expertise with almost three-quarters disagreeing with Carter compared with almost three-quarters of teachers who did agree with Carter: the interviews with Key Influencers showed strong alignment with teachers in this aspect. Questions were raised by some interview participants as to how much assessment knowledge teachers actually need in a system driven by external assessment whilst others pointed to the centrality of assessment in the teaching and learning process and therefore the need for better understanding. However, access to good quality CPD was widely noted as scarce or of variable quality.

9.6 Research implications

This study has addressed three key questions relating to changes in the role played by teachers in high stakes assessments, perceptions of educational assessment held by teachers and Key Influencers, and the curriculum content relating to assessment theory and practice in initial teacher education. This section presents a brief discussion of the implications of the key research findings of this study for educational policy, practice and further research. However, within the field of education, Badley (2003) presents a picture of research in a state of crisis underpinned by four false assumptions (see Section 4.3.4) that have served to undermine the reputation of research (see also Section 3.2; Hargreaves, 1996; Tooley and Darby,1998; Wilson, 2014). In response to the concerns outlined by these authors, the methodological approach and the intended outcomes of this project have been driven by an underlying philosophy of pragmatic optimism (Foucault, 1976a; see also Fabbrichesi, 2016; in Section 4.3.4). The intention of using this approach is to produce useful, practical outcomes (Cohen et al., 2018) that can be achieved through pragmatic strategies to influence and support reconstruction (Koopman, 2011; Badley, 2003) in the field of educational assessment in the English education system: given the impact of the Covid-19 pandemic on the current and future provision of examinations and tests (see Section 1.8), this may be particularly pertinent. The discussion is presented under the headings of the implications for educational policy, educational practice and further research. However, in practice, these areas are interdependent and need to be seen as a whole rather than separate entities. Recommendations for further research and suggested participants based on this discussion are presented in Chapter 10.

9.6.1 Implications for educational policy

Policy makers have been accused of drawing on research to; "…legitimise policy rather than to inform it" (Clegg, 2005, p. 418, cited in Wellington, 2015. See also Wilson, 2014). And as this research project has shown, change over time in the role of teachers in what have become high-stakes assessments has been largely contingent driven by prevailing political ideology lacking in inherent logic, and constituted by discursive and non-discursive practices, often by way of urgent response to particular problematisations (Bailey, 2016; see also Badley, 2003). However, Badley (2003) as with Koopman (2011), offers a way forward based on a more pragmatic, pluralistic model of inquiry that echoing Foucault's approach, would provide a reconstructive service. As with Koopman's notion of; "…building bridges across familiar philosophical divides" (2011, p. 3), the findings of this research project presented in Chapters 4 and 6 show a wide chasm in trust between policy

makers and the teaching profession suggesting a need for reconstructive dialogue between both parties mediated by reference to robust research evidence. More recent moves resulting from the Covid-19 pandemic have widened levels of trust further as illustrated by the government's stated preference for external examinations over teacher-based assessments (Williamson, 2020a) - despite recourse to centre assessed grades following wide public concern with the "mutant algorithm" (Johnson, 2020) proposed by Ofqual for awarding qualification grades in summer 2020. Further concerns about awarding high-stakes qualifications in 2021 are already widely established amongst teachers, teacher associations, university chancellors and students on the grounds of fairness and equity (see for example Rethinking Assessment, 2020; The Times, 2020) and there is every likelihood of further changes or modifications to the currently stated government policy to run the 2021 series as planned albeit with a slightly delayed timetable, 'more generous grading' and the release of advance information on examination topics (Williamson, 2020b). And even if examinations do take place, the government's stated use of 'generous grades' to mirror standards set in 2020 will have an impact on the comparability of performance standards set in 2019, 2022 and beyond which will pose significant challenges for Ofqual in applying its statutory remit (see Ofqual, 2020c). However, the over-reliance on external paper-based assessments (Dunford, 2020; HelenP, Section 7.5.4) has been questioned (see also Harlen, 2005b; Wiliam, 2001; Johnson, 2013) and there have been renewed calls to scrap National Curriculum tests and the GCSE (see McCann, 2016; Henshaw, 2016; Rethinking assessment, 2020). Much of the opposition to external assessment is generated by concerns over its dominance in the English regime of accountability measures and their adverse impact on teacher and pupil wellbeing raising questions over the ethics of administering such tests as shown in the literature review in Chapter 2, Chapter 4 and the teacher and Key Influencer data presented in Chapters 7 and 8. Yet there was universal acceptance of the need for a fair system of accountability amongst the interview participants. But as this study shows, the lack of expertise amongst the teaching profession with regards to educational assessment (see for example Carter, 2014) and the lack of research into teacher reliability (Harlen, 2004; Wilmut, 2005; Stanley et al., 2009; Ofqual, 2013 and 2019) undermines teacher or school-based alternatives to the external regime of examinations. If approaches to educational assessment are to change, there needs to be a more fundamental review of policy around the aims and purposes of education, the validity of measures of accountability and the potential for teacher-based assessment supported by improved teacher education. Such a system would also require robust approaches to standardisation and moderation. However, the aims and purposes of education should drive the agenda rather than the means and methods of assessment (see Badley, 2002,

p. 307): in short, assessment should serve the purpose of education and not dictate its focus. Such a review should be informed by evidenced based research involving all stakeholders in the education system in a state described by Foucault as *parrhesia* (speaking freely – even to power; see Section 4.3.4) and not fettered by political ideology. However, this has further implications for educational assessment practice and research as presented in the next two sections.

9.6.2 Implications for educational practice

As noted in the previous section, teachers' knowledge and understanding of educational assessment has been questioned (Carter, 2014; McIntosh, 2014), and the evidence presented in this study supports such concerns (see Chapters 6 and 8 in particular). Further concerns have been raised in the academic literature about issues such as teacher bias (Gardner, 2007; Lamprianou and Christie, 2009; Kirkup and Twist, 2015) and the ability of teachers to predict reliable GCSE and GCE grades (see Gill and Benton, 2015a and 2015b). Yet as the data in Chapter 6 shows, concepts such as bias, performance standards and reliability are not perceived as being adequately covered in initial teacher education. It should be noted that although ITT providers presented a different view to teachers in their confidence that these aspects were adequately covered in their courses (Chapter 5), broader concerns over teachers' understanding of assessment theory and practice were expressed in the teacher and Key Influencer data presented in Chapters 7 and 8 respectively. It was also suggested by some Key Influencer and teacher interview participants that where teachers applied techniques such as assessment for learning or questioning of pupils, it was undertaken with no real depth of understanding and was more a form of 'going through the mechanics' (AngelaS, Section 7.4.1), an approach in line with Twiselton's notion of 'task managers' (see Section 2.13). The Teachers' Standards (DfE, 2011c) and the ITT Core Content and Early Careers Frameworks (DfE, 2019b) do note assessment as a required area of study in ITT and in the early years of teaching, and require instruction in educational assessment, but the Standards operate at a high level with little detail as to what actually needs to be covered and at what stage in the early careers of teachers. Further, given that the successful implementation of the Standards and Framework depends on the knowledge of ITT providers and school mentors, there is a clear need for greater detail and support materials for those charged with ensuring the requirements are met and to an acceptable standard. A clear implication here is that further work needs to be undertaken to establish a detailed curriculum for ITT and for the on-going development of teachers, and for extended training for those holding teaching positions in ITT and in mentoring newly qualified teachers in schools and colleges.

From the evidence presented in Chapter 7 in particular, there is a clear appetite in the teaching profession for continuous professional development, but this in turn requires quality provision, the support and facilitation of school leaders, local authorities or multi-academy trust managers, and resources from central government. It should also be noted that such provision should not be viewed as solely to support high-stakes assessment, but to the benefit of the teaching and learning process (Wiliam, 2013; see also Section 8.17). Such a move needs to be co-ordinated as more recent policy to offer a range of routes into teaching, with school-centred training favoured by current government ideology rather than informed by research and evaluation (Golding, 2015; Roberts and Foster, 2016), has served to limit access to higher education institutions involved in high quality research: in practice, this has been replaced this with a high dependency on extant expertise within schools – the prevalence of which is not recognised by the literature or by the findings of this study. However, organisations such as the Chartered Institute of Educational Assessors (CIEA) and the Chartered College of Teaching (CCT) are well placed to support schools, but there needs to be broader recognition of the benefits from government, the support of the teacher associations and crucially by senior leaders in schools and colleges. But despite calls for the greater involvement of teachers in high-stakes assessment (for example Bew, 2011a&b: Rimfeld et al., 2019; ATL et al., no date), it is only through the development of teacher expertise supported by robust research evidence that alternatives can be posited and the professional status of teachers in this field re-established. The implications for research are discussed in the next section.

9.6.3 Implications for further research

The findings of this research project support the view that there is a need for a greater focus on the content and quality of educational theory and practice available in initial teacher training and in the further professional development of teachers. However, there is as yet no consensus on what newly qualified teachers need to know and be able to do with regards to educational assessment (see Section 2.2; Ingvarson and Rowley, 2017) neither is there confidence in the ability of teacher trainers in this field (Gardner, 2007; Carter, 2015), a point supported by the research data presented in Chapters 6 and 8. And despite the calls to fill the gap in the paucity of research into the reliability of teacher-based assessments over many years (Harlen, 2004; Wilmut, 2005; Stanley et al., 2009; Ofqual, 2011 and 2013), there is still a need of further research in this area (Johnson, 2012; Ofqual, 2019; Gill, 2019).

However as discussed in Section 3.2, the quality of educational research has been questioned or that; "…the policy debate is often occupied by contradictory views, which can be based on personal experience, anecdote, and political instinct" (EPI, 2020. See also Hargreaves, 1996; Hillage Report, 1998; Badley, 2003; Wilson, 2014; Wellington, 2016; Badley, 2003). This study has also shown the lack of transparency in the content relating to educational theory and practice in initial teacher education courses now offered through a range of routes into teaching (Sheffield Institute of Education, 2019; Carter, 2014), a position exacerbated by the lack of consistency and understanding of the language used around educational assessment (Newton, 2007) and supported by the data presented in Chapter 8 (see for example AmarPA, Section 8.14).

But if there is to be any possibility of fundamental improvement in the use and understanding of educational assessment, this change needs to be underpinned by research evidence. In consideration of the implications for further research, Badley (op cit) offers a useful framework of analysis that reflects the Foucaltian approach covered in Chapter 3. Badley's proposal for a 'pragmatic reflective equilibrium' (p. 305) offers what he describes as 'forging a working point of view' rather than an alternative or new research paradigm. The emphasis here is on continuous reflection, whilst equilibrium is temporary, thus creating an approach that intertwines thought with action where doubt is resolved, and new doubt is generated. Key to Badley's proposition is the involvement of individuals working in a community sharing and comparing their considered judgments; "…until divergence of views is at a minimum" (p.305). The establishment of communities of professional practice would support negotiated change through the generation and impact of research as teachers become no longer viewed as technicians or consumers of educational research (op cit., 307) but as creators and co-creators of research. Such an approach does however call for greater understanding of research methodology amongst teachers, an understanding of the key issues facing teachers by researchers, and access to adequate resources provided by policy makers to support schools and researchers in these endeavours. Developing communities of practice may also be a challenge in the context of a high accountability system that in some regards places schools into a level of competition: this was an aspect referenced by teacher interviewees in Section 7.4.13. Recommendations for further research are presented in Chapter 10.

9.7 Chapter 9 summary

This chapter has presented a discussion on the three research questions of interest to this project. It has shown how changes have been made to the purpose and uses of assessment over three identifiable epistemes using a Foucault inspired analysis and the bearing this has had on the role played teachers. Over time, teacher representative bodies have lost or spurned the opportunity to control high-stakes assessments resulting in their reduced professional status with regards to the design and administration of tests and examinations. The chapter has provided insight into teacher and Key Influencer perceptions of assessment and the influence of a system of high accountability. A new and more detailed insight into the assessment related content and levels of confidence in ITT courses has also been provided. Chapter 10 concludes the thesis with what I believe to be a set of conclusions that can be drawn from this research along with recommendations for future work that stem from the implications identified in Section 9.6.

Chapter 10: Conclusion and recommendations for future research

10.1 Chapter over-view

This chapter presents the conclusions that can be drawn from this research project. The chapter begins with what I believe to be the main contributions of this research before acknowledging the limitations of the project and recommendations. The chapter concludes with a personal reflection on the experience of undertaking this research project.

10.2 The contribution of this research

Firstly, utilising the work of Foucault as a tool of inquiry, this study provides an understanding of how the role of teachers in high-stakes assessment has changed over time through the identification of three epistemes (Chapter 4). In particular, it presents a novel view of changes to the role played by teachers in educational assessment using an approach described by Foucault as a 'history of the present', that is; "…a problem expressed in the terms current today…I try to work out its genealogy. Genealogy means that I begin my analysis from a question posed in the present" (Foucault quoted by Kritzman, 1988: p. 262; cited in Garland, 2014, p. 367). This approach is not predicated or conditioned by a model of continuous improvement but one dependent on the contemporaneous social, political and technical forces of the time. It demonstrates how assessment through examinations used for qualifications (or tests as used in statutory assessments) has been extended to include use within a system of accountability as a tool of surveillance and the centrality of the examination as a controlling mechanism for the system of education.

Secondly, the study shows how the role of teachers has moved from one of centrality and high trust in the process of educational assessment to one where teachers have been marginalised from the process driven by concerns over the level of reliability and trust that can be attributed to teacher derived assessments used a as a central measure of their performance in a system of high-stakes accountability. This study suggests that this conflict in interest has driven a system that has favoured increased reliability over validity (Chapter 1, 2 & 4).

Thirdly, the study has shown that over the period of interest, teachers have been in positions of influence or power in determining their role in the assessment processes of high volume tests and examinations but used such power to remove or reduce their involvement more often than not based on arguments around workload (Chapter 4). This study argues that although these moves have successfully reduced teacher's workload, they have been accomplished at the cost of diminished professional status in high-stakes educational assessments.

Fourth, the study has highlighted contradictions between the views held by some teachers and representative bodies that high-stakes tests and examinations are damaging the wellbeing of pupils and teachers and their continued practice in administering such assessments. This raises ethical issues for the teaching profession that are as yet to be seriously explored in the academic literature. However, the study has also shown that the teaching profession lacks a collective voice as exemplified by Kenneth Baker, then Secretary of Education, in successfully establishing the 1988 Education Reform Act (Chapter 4).

Fifth, the data from the interviews with teachers and Key Influencers have shown that a shared understanding and language around the underlying key concepts of educational assessment is in short supply (Chapters 7 & 8): this supports the notion of a lack of assessment literacy within the teaching profession as found in the academic literature investigated in this study (Chapter 2).

Sixth, this study supports the findings of the Carter Review in relation to gaps in provision around educational assessment found in ITT. However, it exposes a level discrepancy between the views of ITT providers and teachers, particularly around the levels of expertise found amongst ITT providers. Further, the surveys used in this study have provided a level of detail not found elsewhere in the literature on the content of ITT courses and levels of confidence in the extent to which newly qualified teachers are prepared in key aspects of educational assessment as they engage in their early years of teaching (Chapter 5 & 6).

Seventh, the findings of the study support the NAHT Commission on Assessment's finding that there is a 'worrying lack of trust within the teaching profession' around assessment (Chapter 4). However, this study has shown that levels of trust felt by teachers are heavily nuanced and differ between groups for example between teachers within an institution

where levels of trust are relatively high to that between teachers and politicians where teachers report high levels of mistrust (Chapter 5 & 6).

Eighth, despite concerns over the reliability of teachers' assessments discussed in Chapter 3, the Covid-19 pandemic has resulted in total recourse to teachers submitting centre derived assessment grades and rank ordering of their students. The data generated by this approach to awarding qualifications in 2020 will present a unique research opportunity in gaining a deeper understanding of how teachers determine their estimated grades and their perceived accuracy. It is hoped that this study and the recommendations in Section 10.6 below will make a contribution to the post-summer 2020 qualifications results analysis.

Finally, although this study has found deficiencies in ITT provision and raised previously unidentified concerns about levels of confidence and expertise in educational assessment, it needs to be acknowledged that time is limited in ITT courses to deal with key aspects in educational assessment in detail (Chapter 8). Further, as found in the data generated by this study (chapters 4 to 8), levels of expertise in educational assessment have likely been reduced by a system in which the focus on formative assessment has reduced and its quality of practice questioned in a system of high accountability where assessment is viewed as external to teachers being. None of this questions the potential of teachers to conduct valid and reliable assessments, but this study suggests that more support will be required to develop expertise in ITT institutions, schools and mentors if the aims of the newly found government initiatives to support early career teachers are to be achieved.

10.3 The limitations of this research study

Having made claims for the contributions of this study, the limitations of the study need to be acknowledged. First and foremost, I have to declare my level of insider knowledge of the workings of national policy. Whilst this has been of great benefit in gaining access to some of the Key Influencers in education policy and research, it has to be acknowledged that the responses of some of the participants may be influenced by their background knowledge of me and my known value position that the teaching profession should be more trusted in terms of their educational assessment of students, but that this position need to be earned through increased knowledge of assessment theory and practice supported by well researched quality assurance systems. I have made every effort to avoid bias in conducting the interviews by holding back on personal views and by using

questions designed to understand the views of participants alone. I am somewhat assured however, by the strength of character exhibited by interviewees.

It should also be recognised that being a lone researcher limits the scope of inquiry. Using face-to-face interviews is time consuming and finding times to access interviewees has proved a challenge. Interviewing people with busy schedules and other demanding commitments may limit their time to think things through. Limited resources also reduce the number of interviews and given that interviewees volunteered to take part in the project, they are effectively self-selecting: this limits the extent to which the findings of this project can be generalised to the wider population. And whilst the use of surveys assists in triangulating the findings, it has to be noted that respondents to the ITT provider questionnaire were acting in an individual capacity and not necessarily expressing the views of their respective institutions. Hence, along with the teacher survey respondents, participation was entirely voluntary and may have been motivated by a range of reasons.

The approach used to address question one of this research project was based on the development of an analysis of change over time using a method derived from the influence of Michel Foucault. I have made every attempt to apply Foucault's pragmatism, but I am conscious that the combination of no other studies having taken this approach in attempting to understanding change in teachers' assessment over time combined with my own limited experience as a researcher may have impacted adversely on my mission to stay faithful to my stated methodological position.

This study has supported claims that the expertise of teachers and teacher educators is deficient. However, this is a general claim made at system level and therefore one that may well overlook pockets of exemplary practice, be it by individuals, schools or teacher training institutions. And whilst the study has identified in some detail the aspects of educational assessment in which instruction is low or lacking in confidence, it has not identified the appropriate range of content and depth of coverage of a curriculum of instruction designed for teachers as they move through the phases of their careers. This, along with other considerations is included in the next section of recommended actions emanating from this research project.

10.4 Recommendations

The following recommendations arise from the issues and findings identified in this study. The recommendations are an outcome of the discussion of the research questions as

presented in Chapter 9. The focus on four recommendations is derived from the strength of evidence for further work on key areas that will have the most impact on strengthening educational assessment practice in the English education system. The areas selected are based on gaps and suggestions found in the literature combined and reinforced by the data generated in the interviews with teachers and Key Influencers, and the questionnaire surveys.

Drawing on my own experience of working in education policy as described in Chapter 1, I have identified a range of key organisations, representative bodies, experts in the field and policy makers who are best placed to take the suggested work forward: therefore, each recommendation includes suggestions for participants in further research along with a designated lead authority.

10.4.1 Recommendation1

Detailed curriculum and associated support materials should be developed that are designed to meet the needs of teacher educators and school based mentors.

This study recognises the importance of initiatives to improve teacher's knowledge and understanding of educational assessment as they move through initial training and into the first years of practice. However, the study has shown that the current state of expertise in educational assessment is low and as such raises questions as to the extent to which any new initiative can be implemented and supported. It is therefore recommended that further work is undertaken to formulate a detailed curriculum and associated support materials that are designed to meet the needs of teacher educators and school based mentors. This will require careful consideration of what teachers really need to know about assessment and at what point in their career. This should be led by the Chartered Institute of Educational Assessors and include the involvement of the DfE, the Chartered College of Teaching, and initial teacher educators.

10.4.2 Recommendation 2

Further work should be undertaken to develop ways of supporting schools and teachers in developing their understanding and practice in educational assessment.

Recommendation 1 is targeted to support new entrants into the teaching profession. However, despite findings in the academic literature of the centrality of assessment in the

teaching and learning process, assessment literacy across the teaching profession is low. This suggests that further work should be undertaken to develop ways of supporting schools and teachers in developing their understanding and practice in educational assessment. This should be led by the DfE and include the involvement of the Chartered College of Teaching, the Chartered Institute of Educational Assessors and the academic research community.

10.4.3 Recommendation 3

Further research should be undertaken to develop workable standard setting and moderation procedures in schools and colleges.

The study of changes to the role played by teachers over time in educational assessment has shown how the impact of the school accountability system has resulted in the marginalisation of teachers from the processes of high-stakes assessments. Whilst such moves have increased the reliability of assessments, it has come at the cost of validity, particularly in subjects of a practical or vocational nature. And although teacher representative bodies have over time called for greater appreciation of teachers' assessments, they have not proposed alternative models backed by research. Further research into workable standard setting and moderation procedures is therefore recommended. This work should feature a full analysis of the outcomes of measures introduced in 2020 to deal with the Covid-19 pandemic. This should be undertaken with the full support of the Government and led by Ofqual and include the involvement of the awarding organisations, the DfE, teacher representative bodies, the Chartered Institute of Educational Assessors and the academic research community.

10.4.4 Recommendation 4

Research should be undertaken into the ethical considerations of assessment in a highly performative system.

This study has highlighted the detrimental effects of high-stakes testing on teachers' wellbeing and that of their pupils. And despite teachers' acknowledgment of these undesirable effects, schools continue to administer these and other widely available practice tests on which they are often accused of focusing their teaching. Conversely, the study has also identified a small number of schools that have developed a publicly stated charter that favours assessment only as a support for learning and prioritises the personal

growth and the wellbeing of learners over National Curriculum tests. This suggests further research into the ethical considerations of assessment in a highly performative system is needed. This should be led by the DfE and include the involvement of teacher representative bodies and the academic research community.

10.5 Personal reflections on this research study

Having spent in excess of forty-five years working in the education profession at a variety of levels, I have been forcibly struck by the number of times I have heard interview participants working at all levels of the education system express the view that taking part in this research study has given them the all too rare opportunity to reflect on their experiences and practices. This underlines what I believe has been a total privilege for me to engage in post-graduate research, an opportunity to reflect on current practice and explore new ideas, an opportunity that very clearly should be open to more people in what should be a profession underpinned by continuous professional development and engagement with the research community.

The generosity of spirit and access to world-class teachers and supervisors at the University of Leeds has made for an invigorating and humbling experience and one that will see me continue to explore further research in the field of educational assessment. Research into to the approaches and outcomes of the summer 2020 pandemic arrangements for general, technical and vocational qualifications will be of particular personal interest.

The simple goal of my research is to work in areas that will be of direct benefit to those still practising in the teaching profession – and importantly to the benefit of the young people they teach.

List of References

AERA/APA/NCME, 1999. *Standards for Educational and Psychological Testing*. The American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME).

Aldridge, P., Dean, D and Aldrich, R. 1991. Education and policy in England in the twentieth century. Publisher London: Woburn Press.

Alexander, W. P. 1946. *The Education Act. A Parents' Guide. With a Foreword by the Rt. Hon. R.A. Butler M.P.* Phoenix House Ltd. London.

Alexander, R., Rose, J. and Woodhead, C. 1992. *Curriculum Organisation and Classroom Practice in Primary Schools. A discussion paper.* Department for Education and Science.

Allen, A. 2012. *Cultivating the myopic learner: the shared project of high and low-stakes assessment.* British Journal of Sociology of Education, 33 (5). pp. 641-659. ISSN 0142-5692.

Allen, A. 2012a. Using Foucault in education research, British Educational Research Association on-line resources. [Online] [Accessed 20th November 2019]. Available from: https://www.bera.ac.uk/publication/using-foucault-in-education-research

Allen, R., Belfield, C., Greaves, C., Sharp, C. and Walker, M. 2016. *The Longer-Term Costs and Benefits of Different Initial Teacher Training Routes.* IFS report 118. Institute of Fiscal Studies.

Alvesson, M. and Sandberg, J. 2011. *Generating research questions through problematization*. Academy of Management Review, 2011, Vol. 36, No.2, 247–271. Available from:
https://www.jstor.org/stable/41318000?seq=1#metadata_info_tab_contents

AQA, 2020. Special consideration. [Online]. [Accessed 12 January 2020]. Available from: https://www.aqa.org.uk/exams-administration/special-requirements/special-consideration

Aguinis, H. and Solarino, A.M. 2019. Transparency and replicability in qualitative research: The case of interviews with elite informants. Received: 26 April 2017 Revised: 3 February 2019 Accepted: 12 February 2019. DOI: 10.1002/smj.3015. *Strategic Management Journal* published by John Wiley & Sons, Ltd. Strat. Mgmt. J. 2019;1–25. Wiley On-line library. [Accessed 18th April 2019]. Available from:
https://onlinelibrary.wiley.com/doi/pdf/10.1002/smj.3015

Arksey, H. and Knight, P. 1999. Interviewing for social Scientists. SAGE Publications.

Astle, J. 2017*. The Ideal School Exhibition*. RSA Action research Centre, London.

Attalli, Y. 2016. Effort in Low-Stakes Assessments. What Does It Take to Perform as Well as in a High-Stakes Setting? *Educational and Psychological Measurement.* Volume: 76 issue: 6, pp. 1045-1058. Article first published online: March 4, 2016. Issue published: December 1, 2016. DOI: Available from:
 https://doi.org/10.1177/0013164416634789

ATL., NAHT and NUT. No date. *Common ground on assessment and accountability in primary schools*. Published by The Strategy and Communications Department of The National Union of Teachers.

Badley, G. 2003. The Crisis in Educational Research: a pragmatic approach. *European Educational Research Journal, Volume 2, Number 2, 2003.*

Bailey, L. J. 2016. The Policy dispotif: historical formation and method, In. Ball, S., J. ed. 2016. *Michel Foucault and Education Policy Analysis.* Routledge.

Baird, J., Béguin, A., Black, P. , Pollitt, A. and Stanley, G. 2011. *The reliability Programme; Final Report of the Technical Advisory Group.* Ofqual/11/4825. The Office of Qualifications and Examinations Regulation.

Ball, S. J. 1990. *Politics and Policy Making in Education: Explorations in Policy Sociology.* Routledge. London.

Ball, S.J. 2003. The teacher's soul and the terrors of performativity. *Journal of Education Policy.* 18:2, pp. 215-228, DOI: 10.1080/0268093022000043065. Available from:
http://dx.doi.org/10.1080/0268093022000043065

Ball, S. J. 2013. Foucault, Power, and Education. Routledge. UK.

Barnard, H. C. 1966. A History of English Education from 1760. Richard Clay (The Chaucer Press), Ltd., Bungay, Suffolk.

Barber, M. 1994. *The Making of the 1944 Education Act.* [Online] Cassell Education. [Accessed 8th April 2020]. Available from:
https://books.google.co.uk/books?hl=en&lr=&id=_N3TAwAAQBAJ&oi=fnd&pg=PR5&dq=1944+education+act&ots=zMypv3lasA&sig=lLySNcDTXWTtNkqCJsVM0PkvrtE#v=onepage&q=1944%20education%20act&f=false

Bastalich, W. 2009. Reading Foucault: Genealogy and Social Science Research Methodology and Ethics. [Online] *Sociological Research Online*, Volume 14, Issue 2. doi:10.5153/sro.1905. [Accessed12th April 2019]. Available from:
 http://www.socresonline.org.uk/14/2/3.html

BBC. 1993.   On the record. John Patten interview.   Recorded from transmission: BBC-1. Date: 28.2.93. Available from:
 http://www.bbc.co.uk/otr/intext92-93/Patten28.2.93.html

Beasley, M., Gould, E., Kirkham, S. and Emery, H. 2004. *Report on Key Stage 3 English Review of Service Delivery Failure 2003–2004 to QCA Board.* Qualifications and Curriculum Authority, 2004. Available from:
https://dera.ioe.ac.uk/8228/1/10343_ks3_en_report_04.pdf

Beloe Report. 1960. *Secondary School Examinations other than the G.C.E. Report of a Committee Appointed by the Secondary School Examinations Council in July 1958.* HMSO. London. [Accessed 15th July 2017]. Available from:
http://www.educationengland.org.uk/documents/beloe/beloe.html

Benton, T. 2014. *A re-evaluation of the link between autonomy, accountability and achievement in PISA 2009.* Cambridge Assessment, Research Division.

BERA-RSA. 2014. Research and the teaching profession. Building the capacity for a self-improving education system. Final report of the BERA-RSA Inquiry into the role of research in teacher education. ISBN: 978-0-946671-37-3

Available from:
 https://www.bera.ac.uk/wp-content/uploads/2013/12/BERA-RSA-Research-Teaching-Profession-FULL-REPORT-for-web.pdf

Bertram, D. 2017. *Likert scales are the meaning of life.* [Accessed 22nd May 2017]. Available from:
https://docplayer.net/17623968-Likert-scales-are-the-meaning-of-life-dane-bertram.html

Bew, P. 2011. *Independent Review of Key Stage 2 testing, assessment and accountability. Progress Report.* DfE, England.

Bew Report. 2011. *Independent Review of Key Stage 2 testing, assessment and accountability. Final Report.* DfE, England.

Black, B. 2019. *11 Things we know about marking and two we don't…yet.* Ofqual Blog, Beth Black, Director of Research and Analysis. The Office of Qualifications and Examinations Regulation. (Ofqual). Available from:
 https://ofqual.blog.gov.uk/2019/03/05/14572/

Black, P. 1998. *Testing: Friend or Foe? Theory and Practice of Assessment and Testing.* Routledge. Print ISBN: 9780750706148, 0750706147.

Bloom, B.S., Hastings, J.T. and Madaus, G.F. 1971. *Handbook on formative and summative evaluation of student learning.* McGraw-Hill. New York.

Blower, C. 2015. *Exam factories? The impact of accountability measures on children and young people.* Christine Blower NUT General Secretary. June 2015. Available from:
https://www.teachers.org.uk/education-policies/research/examfactories

Bolton, P. 2012. *House of Commons Library. Education: Historical statistics.* Standard Note: SN/SG/4252 Last updated: 27 November 2012.
(Standard Notes are compiled for the benefit of Members of Parliament and their personal staff. Authors are available to discuss the contents of these papers with Members and their staff but cannot advise others.)

Borges, J. L. 1942. *The Analytical Language of John Wilkins.* Essay. Spanish: (El idioma analítico de John Wilkins.) Available from:
https://ccrma.stanford.edu/courses/155/assignment/ex1/Borges.pdf

BOS, 2017. The Bristol On-line Survey tool. https://www.onlinesurveys.ac.uk

Boylan, M. 2016. National curriculum sub-levels: assessment practices as assemblage. In: *Third Manchester Conference on Mathematics Education and Contemporary theory,* Manchester, 18-21 July 2016. (Unpublished). Available from:
 http://shura.shu.ac.uk/16208/3/Boylan%20National%20curriculum%20sub-levels.pdf

Bradbury, A. 2019. *Pressure, anxiety and collateral damage. The Headteachers' Verdict on SATS.* UCL Institute of Education, University College London. Available from:
https://www.morethanascore.org.uk/wp-content/uploads/2019/09/SATs-research.pdf

Brannen, J. 2005. *Mixed Methods Research: A discussion paper*. ESRC National Centre for Research Methods. NCRM Methods Review Papers. [Online]. [Accessed February 5th 2019]. Available from:
http://eprints.ncrm.ac.uk/89/1/MethodsReviewPaperNCRM-005.pdf

Brereton, J. L. 1944. The Case for Examinations, An account of their place in Education with some proposals for their Reform. Cambridge University Press.
.
Brill, F. and Twist, L. 2013. *Where Have All the Levels Gone? The Importance of a Shared Understanding of Assessment at a Time of Major Policy* Change*. NFER Thinks: What the Evidence Tells Us. Slough: NFER. Available from:
https://www.nfer.ac.uk/publications/99940/99940.pdf

Broadfoot, P. M. 1996. *Education, Assessment and Society.* Open university Press.

Brookhart, S. 2011. The Use of Teacher Judgment for Summative Assessment in the United States: Weighed in the Balance and (Often) Found Wanting**.** *Paper presented at the Invited Research Seminar Teachers' Judgments within Systems of Summative Assessment: Strategies for Enhancing Consistency.* Oxford University Centre for Educational Assessment. June 2011. Available from:

http://oucea.education.ox.ac.uk/wordpress/wp-content/uploads/2011/07/Brookhart-FINAL.pdf

Brown, M. 1996. The context for the research – the evolution of the National Curriculum for mathematics. In: Johnson, D.C. and Millett, A. ed. *Implementing the Mathematics National Curriculum: Policy, Politics and Practice.* Paul chapman publishing Ltd.

Bryman, A. 2008. Social Research Methods. Oxford University Press.

Burt, C. 1959. The examination at eleven plus. *British Journal of Educational Studies,* 7:2, 99-117, [Online] DOI: 10.1080/00071005.1959.9973017. [Accessed 12th July 2019]. Available from: https://doi.org/10.1080/00071005.1959.9973017

Busby, E. 2016. Nine in 10 teachers believe Sats (sic) preparation harms children's mental health, survey finds. *TES, June 2016.*

Callaghan, J. 1976. *'A rational debate based on the facts'.* Ruskin College Speech, Oxford. 18th October 1976. Available from:
http://www.educationengland.org.uk/documents/speeches/1976ruskin.html

Cambridge Dictionary, 2019. [Online]. [Accessed 21st April 2019]. Available from:
https://dictionary.cambridge.org/dictionary/english/developed

Cameron, R. and Molina-Azorin, J. F. 2011. The acceptance of mixed methods in business and management. *In, International Journal of Organisational Analysis: Bingley Vol.192, Iss.3. (2011).*

Carter Review. 2015. *Review of initial teacher training (ITT).* Sir Andrew Carter OBE, Chair. DfE, England.

CCEA. 2014. *Assessment Practice.* Council for the Curriculum, Examinations and Assessment. Available from:
 http://ccea.org.uk/curriculum/assess_progress/assessment_practice

CCT. 2017. *Chartered College of Teachers Aims and vision.* Available from:
https://chartered.college/vision

CFEY/Pearson. 2020. Making Waves: Building a Better Future for Assessment – final report. Centre for Education and Youth Policy in partnership with Pearson. Available from: https://cfey.org/reports/2020/12/making-waves-building-a-better-future-for-assessment-final-report/

Child, S. and Wilson, F. 2015. An investigation of A level teachers' methods when estimating student grades. Cambridge Assessment internal report. Cambridge, UK: Cambridge Assessment. In: Gill, T. 2019. *Methods used by teachers to predict final A level grades for their students. Research Matters: A Cambridge Assessment publication*. 28, 2-10. Available from: https://www.cambridgeassessment.org.uk/Images/562367-research-matters-28-autumn-2019.pdf

Children's Commissioner. 2019. Briefing: the children leaving school with nothing. [online]. [Accessed 20th September 2019]. Available from: https://www.childrenscommissioner.gov.uk/wp-content/uploads/2019/09/cco-briefing-children-leaving-school-with-nothing.pdf

Chitty, C. 1991. *The Great Debate: The Politics of the Secondary School Curriculum, 1976-88.* Thesis submitted for the degree of Doctor of Philosophy in the University of London. Institute of Education, 1991.

Clark, L. 2014. Coursework is axed as Gove toughens up GCSEs and A-Levels: Education Secretary introduces plans for harder exams including questions in foreign languages**.** [Online]. *The Daily Mail, 9th April, 2014.* [Accessed 10th April, 2018]. Available from: https://www.dailymail.co.uk/news/article-2600649/Tougher-exams-questions-foreign-languages-science-experiments-demanded-exam-shake-up. html#ixzz4MDWwIkKI

Clark, R. 2004. Lies, damned lies and education. [Online]. [Accessed June 6th 2019]. *The Spectator. November 2004*. Available from: https://www.spectator.co.uk/article/lies-damned-lies-and-education

Cloake, M. 2017. *Pride, Passion and Professionalism. The NUT and the struggle for education 1870-2017.* National Union of Teachers. 2017.

Cohen, L., Manion, L. and Morrison, K. 2007. *Research Methods in Education.* Sixth Edition. London and New York: Routledge.

Cohen, L., Manion, L. and Morrison, K. 2018. Research Methods in Education. 8th Edition. Routledge, Oxon, England.

Coldwell, M., Greany, T., Higgins, S., Brown, C., Maxwell, B., Stiell, B., Stoll, S., Willis, B. and Burns, H. 2017. Evidence-informed teaching: an evaluation of progress in England. Research report. DfE, England, July 2017. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment _data/file/625007/Evidence-informed_teaching_- _an_evaluation_of_progress_in_England.pdf

Coles, J. 1994. Enough was Enough: The Teachers' Boycott of National Curriculum Testing. [Online] Changing English, 1:2, 16-31, DOI: 10.1080/1358684940010203. [Accessed 15th July 2019]. Available from: https://www.tandfonline.com/doi/pdf/10.1080/1358684940010203?needAccess=true&

Cole, S.J. and Osterlind, S.J. 2008. Investigating Differences Between Low- and High-Stakes Test Performance on a General Education Exam. *The Journal of General Education,* 2008, Vol. 57(2), pp. 119-130. Available from: https://doi.org/10.1353/jge.0.0018

Collins Dictionary, 2020. [On-line]. [Accessed on-line 13th December 2019]. Available from: https://www.collinsdictionary.com/dictionary/english/assess

Colwill, I. 2007. *Improving GCSE: internal and controlled assessment.*
Recommendations on the nature of controls needed for internal and controlled assessment in future GCSE specifications. Report commissioned by the Qualifications and Curriculum Authority. QCA/07/3207.

Coughlan, S. 2017. School exam rules to be reviewed after cheat row. BBC News online. Available from: http://www.bbc.co.uk/news/education-41114798

Creswell, J. W. 2007. *Qualitative Inquiry and Research Design.* Second Edition. USA: Sage Publications.

Creswell, J. W. and Creswell, J. D. 2018. Research design: qualitative, quantitative, and mixed methods approaches. Firth edition, International edition. Thousand Oaks, Calif.; London: Sage Publications.

Creswell, J.W. and Plano Clarke, V.L. 2011. Designing and Conducting Mixed Methods Research. SAGE Publications Inc. Thousand Oakes, California.

Crisp, V. 2008. A review of literature regarding the validity of coursework and the rationale for its inclusion in the GCSE. [Online]. In: *Research Matters: Issue 5. January 2008.* UCLES. A Cambridge Assessment publication. [Accessed 5th September 2018]. Available from:
http://www.cambridgeassessment.org.uk/research-matters/

Cross Report. 1888. *Report of the Royal Commission on the Elementary Education Acts.*

Curtis, P. 2008. Sats for 14 year-olds are scrapped. *The Guardian. Tuesday 14th October 2008.* Available from:
https://www.theguardian.com/education/2008/oct/14/sats-scrapped

Danaher, G., Schirato, T. and Webb, J. 2000. *Understanding Foucault.* SAGE Publications Ltd.

Danaher, G., Schirato, T. and Webb, J. 2002. *Understanding Foucault.* (Second Edition). SAGE Publications Ltd.

Danielson, C. 2008, Assessment for Learning – For teachers as Well as Students. In: Dwyer, C.A. ed. *The Future of Assessment – shaping teaching and learning.* Lawrence Erlbaum Associates, New York and Oxford.

Dearing, R. 1994. *The National Curriculum and its Assessment: Final Report.* The Dearing Review. London: School Curriculum and Assessment Authority (SCAA). 1994.

Dearing, R. 1996. *Review of Qualifications for16-19 Year Olds. Summary Report.* School Curriculum and Assessment Authority (SCAA). Hayes. England.

Dempster, F. N. 1989. Spacing Effects and Their Implications for Theory and Practice. *Educational Psychology Review*, Vol. 1, No. 4 (December 1989), pp. 309-330. Springer. Available from: https://www.jstor.org/stable/23359223

Denscombe, D. 2008. Communities of practice: a research paradigm for the Mixed Methods approach. [Online]. *Journal of Mixed Methods Research, Volume 2 Number 3, July 2008, p. 270-283*. Sage Publications. [Accessed 11th January 2019]. Available from: https://journals.sagepub.com/doi/pdf/10.1177/1558689808316807

Denscombe, D. 2014. Good Research Guide: For Small-Scale Social Research Projects. McGraw-Hill Education.

Dent, H. C. 1943. *Education in Transition. A sociological study of the impact of war on English education 1939-43.* Kegan Paul, Trench, Trubner & Co., Ltd. London.

Dent, H. C. 1962. *The Education Act, 1944. Provisions, Regulations, Circulars, Later Acts.* Ninth Edition. University of London Press Ltd.

Dent, H. C.1963. *The Education System of England and Wales.* Third Edition. University of London Press. London.

Dent, H. C. 1977. *The Training of Teachers in England and Wales 1800 -1975.* Hodder and Stoughton.

DES. 1965. Circular 10/65. *The Organisation of Secondary Education.* [Online]. Department of Education and Science. London. [Accessed 21st July 2019]. Available from: http://www.educationengland.org.uk/documents/des/circular10-65.html

DES. 1970. *Circular 10/70. The Organisation of Secondary Education.* [Online]. Department of Education and Science, London. [Accessed 21st July 2019]. Available from: http://www.educationengland.org.uk/documents/des/circular10-70.html

DES. 1974. *Circular 4/74. The Organisation of Secondary Education.* Department of Education and Science. London. [Accessed 21st July 2019]. Available from: http://www.educationengland.org.uk/documents/des/circular4-74.html

DES.1988. *Task Group on Assessment and Testing: a report*. London, Department of Education and Science and the Welsh Office.

DfE. 2011. *A history of School and College performance tables.* Internal document, unpublished. For information on performance table requirements see DfE Statement of Intent, see for example;
https://www.gov.uk/government/publications/school-and-college-performance-tables-statements-of-intent

DfE. 2011a. *First report of the independent review of teachers' standards QTS and core standards Presented to The Secretary of State for Education.* Department for Education. 14 July 2011. DfE, England.

DfE. 2011b. *Second report of the independent review of teachers' standards Post-threshold, excellent teacher and advanced skills teacher standards.* Presented to The Secretary of State for Education. Department for Education. 12 December 2011. DFE-00168-2011. DfE, England.

DfE. 2011c. *Teachers' Standards.* DfE, England. Available from:
 https://www.gov.uk/government/publications/teachers-standards

DfE. 2012. *Statistical First Release: Schools, Pupils and their Characteristics: January 2012.* DfE, England.

DfE. 2012a. Extract from the summary of the contents of the first primary performance tables published by the DfE, 1996. Source: A History of School and College Performance Tables. DfE, Last revised 16 July,2012: Internal document.

DfE. 2013. *Assessing Without Levels.* Department for Education. 18th September 2013. DfE, England. Available from:
www.education.gov.uk/schools/teachingandlearning/curriculum/nationalcurriculum2014/a00225864/assessing-without-levels

DfE. 2013a. *Secondary School Accountability Consultation,* 7th February 2013. DfE, England.

DfE. 2013b. *Reforming the accountability system for secondary schools; Government response to the February to May 2013 consultation on secondary school accountability.* October 2013. DfE, England. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment _data/file/773705/Early-Career_Framework.pdf

DfE. 2013c. *Teachers' Standards. Guidance for school leaders, school staff and governing bodies.* July 2011. (Introduction updated June 2013). DfE, England.

DfE. 2014a. *Reforming assessment and accountability for primary schools Government response to consultation on primary school assessment and accountability.* March 2014. DfE, England.

DfE. 2014b. *2014 School and College Performance Tables Statement of Intent*. July 2014. DfE, England.

DfE. 2015a. *2015 School and College Performance Tables Statement of Intent.* July 2015. DfE, England.

DfE. 2015b. Workload Challenge: Analysis of teacher consultation responses Research report. February 2015. Sarah Gibson, Lizzie Oliver & Mary Dennison. DfE, England. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment _data/file/401406/RR445_-_Workload_Challenge_-_Analysis_of_teacher_consultation_responses_FINAL.pdf

DfE. 2015c. Findings from the call for evidence. The Carter Review. DfE internal document. Viewed with permission of the DfE.

DfE. 2016a. A framework of core content for initial teacher training (ITT). DfE, England. Available from: www.gov.uk/government/publications

DfE. 2016b. *Initial teacher training (ITT) census: 2016 to 2017 - Provider-level tables*. DfE, England. Available from: https://www.gov.uk/.../initial-teacher-training-trainee-number-census-2016-to-2017

DfE. 2016c. *Eliminating Unnecessary Workload around marking. Report of the Independent Teacher Workload Review Group.* March 2016. DfE, England.

DfE. 2016d. *Eliminating Unnecessary Workload associated with data management. Report of the Independent Teacher Workload Review Group.* March 2016. DfE, England.

DfE. 2016e. *Eliminating Unnecessary Workload around planning and teaching resources.* Report of the Independent Teacher Workload Review Group. March 2016. DfE, England.

DfE. 2016f. *2016 School and college performance tables. Statement of Intent.* August 2016. DfE, England.

DfE. 2017a. *Primary Assessment in England. Government Consultation Response.* September 2017. DfE, England. Available from:
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment
_data/file/644871/Primary_assessment_consultation_response.pdf

DfE. 2017b. *Get into Teaching*. DfE, England. Available from:
https://getintoteaching.education.gov.uk/

DfE. 2018a. Statistics on initial teacher training. July 2018. DfE, England. Available from:
https://www.gov.uk/government/collections/statistics-teacher-training

DfE. 2018b. Analysis of teacher supply, retention and mobility. September 2018. DfE, England. Available from:
https://dera.ioe.ac.uk/32226/1/Teachers_Analysis_Compendium_4.pdf

DfE. 2018c. *Assessment without levels: qualitative research. Research report December 2018*. Helen Poet, Caroline Sharp, Jennifer Garry, Jennie Harland, Gill Keightley, Catherine Kirkup and Kerry Martin – National Foundation for Educational Research. DfE, England. Available from:
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment
_data/file/764174/NFER_AWL_report.pdf

DfE. 2018d. *Initial teacher training (ITT) census: 2016 to 2017. Coverage; England.* DfE, England.

DfE. 2018e. *Initial teacher training (ITT) performance profiles for the academic year 2016 to 2017, England.* p. 5. Published 26 July 2018. DfE, England. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment _data/file/728820/Initial_Teacher_Training_Performance_Profiles_2016_17.pdf

DfE. 2018f. Making data work, Report of the Teacher Workload Advisory Group. Department for Education November 2018. DfE, England. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment _data/file/754349/Workload_Advisory_Group-report.pdf

DfE. 2019a. *Secondary accountability measures. Guide for maintained secondary schools, academies and free schools.* March 2019. DfE, England.

DfE. 2019b. Early Career Framework. DfE England. January 2019. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment _data/file/893150/Early-Career_Framework.pdf

DfE, 2019c. Initial teacher training (ITT): core content framework. Department for Education, 1st November 2019. DfE, England. Available from: https://www.gov.uk/government/publications/initial-teacher-training-itt-core-content-framework

DfE. 2020. *Press release. Schools, colleges and early years settings to close.* Department for Education, 18th March 2020. DfE, England. Available from: https://www.gov.uk/government/news/schools-colleges-and-early-years-settings-to-close

DfE. 2020a. DIRECTION UNDER S129(6) OF THE APPRENTICESHIPS, SKILLS, CHILDREN AND LEARNING ACT 2009. Letter form the Secretary of State for Education to Sally Collier, Chief Regulator, Ofqual, 31st March, 2020. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment _data/file/877611/Letter_from_Secretary_of_State_for_Education_to_Sally_Collier.pdf

DfE. 2020b. DIRECTION UNDER S 129(6) OF THE APPRENTICESHIPS, SKILLS, CHILDREN AND LEARNING ACT 2009. Letter form the Secretary of State for Education to Sally Collier, Chief Regulator, Ofqual, 9th April, 2020. Available from:

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/879024/Letter_from_Secretary_of_State_for_Education_to_Sally_Collier_-_VTQs_-__9_April_2020.pdf

Dornyei, Z. 2007. Research Methods in Applied Linguistics. Oxford University Press.

Downing, L. 2008. *The Cambridge Introduction to Michel Foucault.* Cambridge University Press, UK.

Draft Circular. 1945. *Extract from an Education Department draft circular on the introduction of 11+ exam, July 1945* (ED 147/134). National Archives. [Online]. [Accessed 20th July 2019]. Available from:
https://www.nationalarchives.gov.uk/wp-content/uploads/2014/03/ed147-1341.jpg

Education Act, 1944. [Online]. [Accessed 1st October 2019] Available from:
http://www.legislation.gov.uk/ukpga/Geo6/7-8/31/part/II/crossheading/provision-and-maintenance-of-primary-and-secondary-schools/enacted

Duberley, J., Johnson, P. and Cassell, C. 2012. Philosophies Underpinning Qualitative Research. In: Symon, G. and Cassell, C. 2012. Qualitative Organizational Research. Core Methods and Current Challenges. Sage Publications Ltd.

Dunford, J. 2020. It's exam-only assessment that has led us to this mess. TES, 11th August 2020. [Online] [Accessed 11th August 2020]. Available from: https://www.tes.com/news/its-exam-only-assessment-has-led-us-mess

Education Reform Act. 1988. VOL1/2. Chapter 40. London. Her Majesty's Stationery Office. Reprinted 1989. Available from:
http://www.legislation.gov.uk/ukpga/1988/40/pdfs/ukpga_19880040_en.pdf

Education Select Committee. 2017. *Primary Assessment Inquiry. Ordered by the House of Commons to be printed 26 April 2017.* Available from:
https://publications.parliament.uk/pa/cm201617/cmselect/cmeduc/682/682.pdf

EEF. 2016. Support from senior leaders 'crucial' to getting teachers to engage with research. Education Endowment Foundation (EEF) May, 2016.Available from:

https://educationendowmentfoundation.org.uk/news/support-from-senior-leaders-crucial-to-getting-teachers-to-engage-with-rese/

EPI. 2020. *Our mission and purpose.* Education Policy Institute. Available from: https://epi.org.uk/our-mission-and-purpose/

ESRC. 2018. *Secondary Education and Social Change in the UK Briefing paper.* Economic and Social Research Council. Available from:
https://sesc.hist.cam.ac.uk/wp-content/uploads/2018/09/Briefing-Paper-Exams.pdf

Etymonline. 2020. Online Etymology Dictionary. [Online]. [Accessed 12th February 2020]. Available from: https://www.etymonline.com/word/assess

European Commission. 2018. *Quality assurance for school development. Guiding principles for policy development on quality assurance in school education.* Produced by the ET 2020 Working Groups. Directorate-General Education, Youth, Sport and Culture Schools and multilingualism. Second edition April 2018. Available from:
 https://www.schooleducationgateway.eu/downloads/Governance/2018-wgs2-quality-assurance-school_en.pdf

Expert Group on Assessment. 2009. *Report of the Expert Group on Assessment. DCSF.* Available from: http://www.educationengland.org.uk/documents/pdfs/2009-expert-group-assessment.pdf

Fairclough, N. 2012. Critical discourse analysis. In, *International Advances in Engineering and Technology* (IAET) ISSN: 2305-825. Vol. 7 July 2012. International Scientific Researchers (ISR). Available from: http://scholarism.net/FullText/2012071.pdf

FFT/education datalab. 2019. *How is data used in schools today? A 2019 survey of current practice.* FFT/educationdatalab and teachertapp. co.uk. 3rd December 2019. Available from: https://fft.org.uk/how-schools-use-data/

Foley, B. and Goldstein, G. 2012. *Measuring success. League Tables in the public sector.* The British Academy. London. SW1Y 5AH. Published March 2012. ISBN 978-0-85672-600-2.

Forty, R, 1994. Foucault and Epistemology, In: *Michael Foucault Critical Assessments.* Smart, B. (Ed) Routledge, London and New York.

Foucault, M. 1970. *The Order of Things.* Tavistock Publications Ltd.

Foucault, M. 1972. *The Archaeology of knowledge*. Tavistock Publication Ltd.

Foucault, M. 1973. *The Birth of the Clinic: An archaeology of Medical Perception.* New York; Pantheon.

Foucault, M. 1976. *The history of sexuality volume 1: an introduction.* Penguin: London.

Foucault, M. 1976a. Dits et Écrits 1954–1988 (1976) Vol. II, 1976–1988 edited by Daniel Defert and François Ewald. Not published in English. [Extract online] [Accessed 20th July 2020] Available from: https://en.m.wikiquote.org/wiki/Michel_Foucault

Foucault, M. 1977. Nietzsche, Genealogy, History'. In: Bouchard, D. F. ed. *Michel Foucault. Language, Counter-Memory, Practice: Selected Essays and Interviews*. Ithaca: Cornell University Press. This is reprinted in Rainbow, P. ed. *The Foucault Reader*. Harmondsworth: Penguin, 1984.

Foucault, M. 1977a. *Discipline and Punish. The Birth Of The Prison.* Reprinted in Penguin Books, 1991.

Foucault, M. 1980a. Two Lectures. Power/knowledge. C. Gordon. London: Longman. In: Ball, S. J. 2013. *Foucault, Power, and Education.* Routledge. UK.

Foucault, M. 1988. The Masked Philosopher. In: Kritzman, L.D. ed. *Michel Foucault. Politics, Philosophy, Culture: Interviews and other writings of Michel Foucault, 1977-1984.*New York: Routledge.

Foucault, M. 1994. Nietzsche, Genealogy, History'. In: Faubion, J. ed. *Aesthetics, Method and Epistemology: Essential Works of Foucault 1954-1984 Vol. 2* London: Penguin Books pp. 369-392.

Foucault, M. 1997. Essential Works of Foucault 1954-1983, Volume 1: Ethics, Subjectivity, and Truth. New York: New Press.

Foucault, M. 2010. *The birth of biopolitics: lectures at the College de France 1978-79.* Basingstoke: Palgrave Macmillan.

Fullan, M. G. 1991. Professional development of educators. pp. 315-344. In: Wiliam, D. and M. Thompson. 2008. *Integrating assessment with learning: What will it take to make it work*? In: *The future of assessment: Shaping teaching and learning,* ed. C.A. Dwyer, 53–82. New York: Erlbaum.

Furlong, J. 2005. New Labour and teacher education: the end of an era. *Oxford Review of Education.* 31:1, 119-134, DOI: 10.1080/0305498042000337228. Available from: https://doi.org/10.1080/0305498042000337228

Gabriel, D. 2013. Inductive and deductive approaches to research. [Online]. [Accessed 21st February 2019]. Available from: http://deborahgabriel.com/2013/03/17/inductive-and-deductive-approaches-to-research/

Gale, T. 2001. Critical policy sociology: historiography, archaeology and genealogy as methods of policy analysis. *Journal of Education Policy.* 16(5). pp. 379-393.

Gardner, J. 2007. Is teaching a 'partial profession? *Make the Grade, Summer 2007.* Chartered Institute of Educational Assessors (CIEA).

Garland, D. 2014. What is a "history of the present"? On Foucault's genealogies and their critical preconditions. *Punishment and society*, 2014, Vol. 16(4) 365-384. SAGE.

Gibb, N. 2015. Reforming education through international evidence.
Schools Minister speech to the *Organisation for Economic Co-operation and Development Education Policy Outlook Conference*. Published by the DfE. Available from: https://www.gov.uk/government/speeches/nick-gibb-reforming-education-through-international-evidence

Gibb, N. 2017. Through testing times: a review of the education reform project and its future*. Policy Exchange Fringe Meeting – Conservative Party Conference 2017.* Available from:
 https://www.nickgibb.org.uk/news/through-testing-times-review-education-reform-project-and-its-future

Gibbons, S. 2012. *Big ideas: valuing schooling through house prices.* Centre Piece. Paper No' CEPCP374: Centre of Economic Performance, LSE, London.

Gill, T. 2019. Methods used by teachers to predict final A level grades for their students. *Research Matters: A Cambridge Assessment publication.* 28, 2-10. Available from: https://www.cambridgeassessment.org.uk/Images/562367-research-matters-28-autumn-2019.pdf

Gill, T. and Benton, T. 2013. *Investigating the relationship between aspects of countries' assessment systems and achievement on the Programme for International Student Assessment (PISA) tests.* Cambridge Assessment Research Report 30th March 2013.

Gill, T. and Benton, T. 2015a. The accuracy of forecast grades for OCR GCSEs in June 2014. Statistics Report Series No.91. Cambridge assessment. Available from: https://www.cambridgeassessment.org.uk/Images/241265-the-accuracy-of-forecast-grades-for-ocr-gcses-in-june-2014.pdf

Gill, T. and Benton, T. 2015b. *The accuracy of forecast grades for OCR A levels in June 2014. Statistics Report Series No. 90.* Cambridge, UK: Cambridge Assessment.

Gillan, D. 2003. The Northern CSE Boards. In: *Setting the Standard, A century of public examining by AQA and its parent boards.* Tattersall, K., Day, J., James, H., Gillan, D. and Spencer, A. AQA, England.

Gillard, D. 2011. *Education in England: a brief history.* Available from: www.educationengland.org.uk/history

Gillard, D. 2017. *Education England: The New Secondary Education (1947).* Ministry of Education Pamphlet No. 9. [Online]. [Accessed 20th August 2019]. London: His Majesty's Stationery Office 1947. Available from: http://www.educationengland.org.uk/documents/minofed/pamphlet-09.html

Gladwell, M. 2009. Outliers: The story of success. Penguin Books.

Glaser, B, G. and Strauss, A, L. 1967. *The discovery of Grounded Theory. Strategies for Qualitative Research.* Chicago: Aldine.

Glossary of Education Reform. 2014. High-stakes test. [Online]. [Accessed 14th August 2018]. Available from: http://edglossary.org/high-stakes-testing/

Golding, J. 2015. What has the Coalition Government done for the development of initial teacher education? *London Review of Education* Volume 13, Number 2, September 2015.

Gov.Uk. 2016. *Official Statistics. Initial teacher training: trainee number census 2016-2017. Provider level tables SFR-57-2016 to 2017.* DfE. November 2016. Available from: https://www.gov.uk/government/statistics/initial-teacher-training-trainee-number-census-2016-to-2017

Gove, M. 2013a. 'I refuse to surrender to the Marxist teachers hell-bent on destroying our schools: Education Secretary berates 'the new enemies of promise' for opposing his plans'. *Mail Online, March 23 March 2013.* Available from: https://www.dailymail.co.uk/debate/article-2298146/I-refuse-surrender-Marxist-teachers-hell-bent-destroying-schools-Education-Secretary-berates-new-enemies-promise-opposing-plans.html#ixzz4chIf0tBb

Gove, M. 2013b. Gove to overhaul GCSE league tables to stop schools 'gaming' the system. *Sunday Times. 14th October 2013.* Available from: https://www.thetimes.co.uk/article/gove-to-overhaul-gcse-league-tables-to-stop-schools-gaming-the-system-wwlxv2b6sn9

Gove, M. 2014a. Coursework is axed as Gove toughens up GCSEs and A-Levels. (Secretary of State for Education, Changes to GCE and GCSE examinations in 2014.) *The Mail Online, England.* Available from: https://www.dailymail.co.uk/news/article-2600649/Tougher-exams-questions-foreign-languages-science-experiments-demanded-exam-shake-up. html

Gove, M. 2014b. *The future of vocational education.* Speech, Michael Gove, Secretary of State for Education, speaks at McLaren about the future of vocational education. Published 3 March 2014. Department for Education, England.

Greene, J. C., Caracelli, V. J. and Graham, W. F. 1989. Toward a conceptual framework for mixed-method evaluation designs. In: *Educational Evaluation and Policy Analysis, 11 (3), pp. 255-274.*

Gregory, K. and Clarke, M. 2003. *High-Stakes Assessment in England and Singapore. Theory Into Practice*, Vol. 42, No. 1, The Impact of High-Stakes Testing (Winter, 2003), pp. 66-74. Published by Taylor & Francis, Ltd. Available from:
http://www.jstor.org/stable/1477320

Guba, E. G. and Lincoln, Y.S. 1994. Competing Paradigms in Qualitative Research. In: Denzin, N.K. and Lincoln Y.S. eds. Handbook of Qualitative Research, pp. 105–117. Thousand Oaks, CA: Sage.

Hadow Report.1926. [Online]. [Accessed 5th May 2019]. Available from:
http://www.educationengland.org.uk/documents/hadow1926/hadow1926.htm

Hall, R. F. 2013. Mixed Methods: In search of a Paradigm. In: *Conducting Research in a Changing and Challenging World,* Chapter 7: Publisher: Nova Science Publishers Inc., Editors: Thao Le and Quynh Le, pp. 71 – 78. [Online]. [Accessed 15th April 2019]. Available from:
https://www.researchgate.net/publication/259045135_Mixed_Methods_In_search_of_a_paradigm

Hall, B. and Howard, K. 2008. A synergistic approach: Conducting mixed methods research with typological and systemic design considerations. *Journal of mixed Methods Research, 2(3), pp. 248-269.*

Hanks, J. 2017. *Exploratory Practice in Language Teaching. Puzzling about Principles and Practices*. Palgrave Macmillan.

Hansard. 1943. HC Deb 29 July, 1943. vol 391 cc1825-928.

Hanson, W. E., Creswell, J. W., Clark, V. L P. , Petska, K. S. and Creswell, J. D. 2005. Mixed Methods Research Designs in Counseling Psychology in *Journal of Counseling Psychology* Issue: Volume 52(2), April 2005, p 224-235. Section eds: Haverkamp, B. E., Morrow, S. L., Ponterotto, J. Copyright: (C) 2005 by the American Psychological Association. I Publication Type: [Mixed Methods]. SSN: 0022-0167. Accession: 00001192-200504000-00012.

Harford, S. 2015. What Ofsted inspectors will look at (video). National Director, Ofsted. Available from: https://www.youtube.com/watch?v=H7whb8dOk5Q

Hargreaves, D. 1996. *Teaching as a research-based profession: possibilities and prospects. (The Teacher Training Agency Lecture 1996).* In: Hammersley, M. 2007. ed. Educational Research and Evidence-based Practice. Sage.

Harlen, W. 2004. The role of assessment by teachers in summative assessment systems: what is the evidence that teachers assessment can be used dependably for summative assessment? Paper presented at the *British Educational Research Association Annual Conference, University of Manchester*, 16-18 September 2004.

Harlen, W. 2005a. Trusting teachers' judgement: research evidence of the reliability and validity of teachers' assessment used for summative purposes, *Research Papers in Education*, 20:3, 245-270, DOI: 10.1080/02671520500193744. Available from: https://www.tandfonline.com/doi/abs/10.1080/02671520500193744

Harlen, W. 2005b. Teachers' summative practices and assessment for learning – tensions and synergies, *Curriculum Journal*, 16:2, 207-223, DOI: 10.1080/09585170500136093. Available from:
https://doi.org/10.1080/09585170500136093

Haycock, K. 2008. Improving Achievement and Closing Gaps Between Groups: The Role of Assessment in Helping Us To Move Faster. In: Dwyer, C.A. ed. *The Future of Assessment – shaping teaching and learning.* Lawrence Erlbaum Associates, New York and Oxford.

Hazel, W. and Ward, H. 2017. Exclusive: DfE declares ceasefire in war on university teacher-training departments. *TES.10th March 2017.*

Heim, A. W. 1954. *The Appraisal of Intelligence.* Methuen & Co. Ltd, London.

Hellawell, D. 2006. Inside–out: analysis of the insider–outsider concept as a heuristic device to develop reflexivity in students doing qualitative research. Teaching in Higher Education, 11:4, 483-494, Available from:
https://www.tandfonline.com/doi/pdf/10.1080/13562510600874292?needAccess=true&

Henry, J. 2009. Teachers unanimously back Sats boycott. [Online]. *The Telegraph.* [Accessed 23rd April 2019]. Available from:

https://www.telegraph.co.uk/education/educationnews/5141158/Teachers-unanimously-back-Sats-boycott.html

Henshaw, P. 2016. It is time to scrap all examination up to 16 – including GCSEs. *SecEd.* Published 18th May 2016.

Hillage, J., Pearson, R., Aanderson, A. and Tamkin, P. 1998. *Excellence in Research on Schools* (Hillage Report), Department for Education and Employment, England.

HMC. 2012. *England's 'examinations industry': deterioration and decay. A report from HMC on endemic problems with marking, awarding, re-marks and appeals at GCSE and A level, 2007-12.* The Headmasters' and Headmistresses' Conference.

HMC. 2018. *New data from Ofqual reveals the true extent of grading unreliability*. **the** Headmasters' & Headmistresses' Conference **(**HMC), press release, 1st December 2018. Available from:
https://www.hmc.org.uk/blog/new-data-ofqual-reveals-true-extent-grading-unreliability/

HMCI. 2011. *The Annual Report of Her Majesty's Chief Inspector of Education, Children's Services and Skills 2010/11*.

HMCI. 2012. The Annual Report of Her Majesty's Chief Inspector of Education, Children's Services and Skills. Ofsted. London. The Stationery Office.

HMCI, 2013. Annual Report, 2012/3. Her Majesty's Chief Inspector of Schools. Ofsted. Manchester.

Hoddle, L. 2006. *Did 'failing' the 11plus mean, for some, a lifetime of feeling secondary? A summary of some research findings.* [Online]. [Accessed18[th] August 2019]. Available from:
https://comprehensivefuture.org.uk/did-failing-the-11plus-mean-for-some-a-lifetime-of-feeling-secondary/

Hodges, D., Martimianakis, M., A., McNaughton, N. and Whitehead, C. 2014. Medical education... meet Michel Foucault. In: *Medical Education* 2014; 48: 563–571. John Wiley & Sons Ltd.

Hodgson, A. and Spours, K. eds. 1997. *Dearing and Beyond: 14-19 Qualifications, Frameworks and Systems*. Kogan Page Ltd.

House of Commons Children, Schools and Families Committee, 2008. *Testing and Assessment Third Report of Session 2007–08. Volume II. Oral and written evidence. Ordered by The House of Commons to be printed 7 May 2008*

Holmes, A. 2020. Researcher Positionality – A Consideration of Its Influence and Place in Qualitative Research – A New Researcher Guide. *Shanlax International Journal of Education*, vol.8, no. 4, 2020, pp. 1-10. Available from:
https://files.eric.ed.gov/fulltext/EJ1268044.pdf

House of Commons Education Committee. 2017. *Primary Assessment. Eleventh Report of 2016-17 session. Ordered by the house of Commons to be printed 26th April 2017.* Available from:
https://publications.parliament.uk/pa/cm201617/cmselect/cmeduc/682/682.pdf

Huddersfield County Borough. 1945. Minutes of Huddersfield County Borough Council General Purpose (Education) Sub-Committee, 12th December 1945
(Huddersfield County Borough Council Minutes, p. 259).

Huddleston, P. 2015. How should we assess vocational and practical learning? In: *The future of Assessment 2025 and beyond*. AQA. 2015.

Hyman, W. M. and Clegg, A. B.1954. *West Riding Education. Ten Years of Change.*

IBM SPSS Data Collection Data Entry 7 (Media Set). University of Leeds.

Ingvarson, L. and Rowley, G. 2017. Quality Assurance in Teacher Education and Outcomes: A Study of 17 Countries. In: *Education Researcher,* Vol. 46 No. 4, pp. 177-193. AERA May 2017.

Isaacs, T., Zara, C., Herbert, G: with Coombs, J. and Smith, C. 2013. *Key Concepts in Educational Assessment.* Sage. London.

Jackson, A. Y. 2013. Spaces of Power/Knowledge: A Foucauldian Methodology for Qualitative Enquiry. *Qualitative Inquiry,* October 2013. SAGE Journals. Available from: http://journals.sagepub.com/doi/abs/10.1177/1077800413503803

James, M. 2010. Educational Assessment: overview. In: Peterson, P. , Baker, E. and McGaw, B. eds. *International Encyclopedia of Education: Third Edition*, Volume 3: (Oxford: Elsevier): 161-171. Available from: https://www.sciencedirect.com/science/article/pii/B9780080448947017139

James Report. 1972. *Report of the Committee of Enquiry into Teacher Education and Training.* D.E.S., London.

JCQ, 2019. *Report of the Independent Commission on Examination Malpractice*. The Joint Council for Qualifications. September 2019.

Jellis, C. 2018. *In pursuit of reliability and validity.* [Online].Published by the Centre for Evaluation and Monitoring (CEM), July 12, 2018. [Accessed 22nd July 2019]. Available from: https://www.cem.org/blog/in-pursuit-of-reliability-and-validity/

Johnson, B. 2020. *A-levels and GCSEs: Boris Johnson blames 'mutant algorithm' for exam fiasco.* Reported by Sean Coughlan, BBC News family and education correspondent. BBC 26 August 2020. [Online] [accessed 26 August 2020]. Available from: https://www.bbc.co.uk/news/amp/education-53923279

Johnson B. R., Onwuegbuzie A.J. and Turner L.A. 2007. Toward a definition of mixed methods research. *Journal of Mixed Methods Research*. 2007;1:112–133. doi: 10.1177/1558689806298224.

Johnson, S. 2011. *A focus on teacher assessment reliability in GCSE and GCE. A Report for Ofqual.* Ofqual/11/4807 January 2011. First published by the Office of Qualifications and Examinations Regulation in 2011.

Johnson. S. 2012. A focus on teacher assessment reliability in GCSE and GCE. In: *Ofqual's Reliability Compendium.* Opposs, D and He, Q. eds. Ofqual, pp. 365-416.

Johnson, S. 2013. On the reliability of high-stakes teacher assessment, *Research Papers in Education*, 28:1, 91-105, DOI: 10.1080/02671522.2012.754229. Available from: https://doi.org/10.1080/02671522.2012.754229

Kendall, G. and Wickham, G. 1999. *Using Foucault's Methods*. SAGE Publications.

Kendall, G. and Wickham, G. 2003*. Using Foucault's Methods.* 2nd edition. SAGE Publications Ltd.

Khosravinik, M. 2014. Critical Discourse Analysis, Power, and New Media (Digital) Discourse, In; Kopytowskaa, M. and Kalyango, Y., eds. *Why Discourse Matters: Negotiating Identity in the Mediatized World*. Peter Lang. p. 287-306. Available from: https://www.researchgate.net/publication/304914039_Critical_Discourse_Analysis_power_and_New_media_discourse_In_M_Kopytowskaa_and_Y_Kalyango_eds_Why_Discourse_Matters_Negotiating_Identity_in_the_Mediatized_World_Peter_Lang_287-306

Kimberly, K., Hextall, I., Torrance, H. and Moon, B. 1989. National Assessment and Testing: the TGAT Report. Review Symposium. *British Journal of Sociology of Education*, Vol. 10, No,2, 1989.

Kirkup, C. and Twist, L. 2015. Performance Descriptors Risk Undermining the Benefits of Teacher Assessment. *NFER Thinks: What the Evidence Tells Us*. Slough: NFER.

Klenowski, V. 2012. Raising the stakes: the challenges for teacher assessment. [Online]. The Australian Educational Researcher, 2012, Vol.39(2), pp. 173-192. [Accessed 22nd January 2018]. Available from:
https://link.springer.com/article/10.1007/s13384-012-0057-9

Lambert, D. and Lines, D. 2001. *Understanding Assessment: Purposes, Perceptions, Practice*. Routledge Falmer, London and New York.

Lamprianou, I. and Christie, T. 2009. Why school based assessment is not a universal feature of high-stakes assessment systems? In: *Educational Assessment, Evaluation and Accountability* (2009) 21:329-345. Published on line. DOI 10.1007/s11092-009-9083-1 Springer Science + Business Media, LLC 2009.

Lawton, D. 1980. *The politics of the school curriculum.* Routledge & Kegan Paul, London, Boston and Henley.

Lazaroiu, G. 2013. Besley on Foucault's Discourse of Education. *Education Philosophy and Theory*, 45:8, 821-832, DOI: 10.1080/00131857.2013.785092 https://doi.org/10.1080/00131857.2013.785092.

Lesson, S. M.A. 1939. *The Spens Report, Religion in Education*, 6:2, pp. 63-70. [Online]. DOI: 10.1080/3908556167 [Accessed 28th September 2019]. Available from: https://www.tandfonline.com/doi/abs/10.1080/3908556167?journalCode=cbre18

Lightfoot, L. 2009. Teachers go to war over life in our classrooms. [Online]. *The Guardian. On-line: Sunday 5th April, 2009.* [Accessed 28th November 2018]. Available from*:* https://www.theguardian.com/education/2009/apr/05/teachers-union-sats-education

Lightman, B. 2015. Assessment in a self-improving system*. In; The future of Assessment 2025 and beyond.* AQA.

Long, R. and Danechi, S. 2019. English Baccalaureate. *House of Commons Briefing Paper* Number 06045, 4 September 2019. [Online]. [Viewed 15th September 2019]. Available from: www.parliament.uk/commons-library.

Looney, A., Cumming, J., van Der Kleij. and Harris, K. 2017. Reconceptualising the role of teachers as assessors: teacher assessment identity*.* Assessment in education: principles, policy & practice. Routledge. Available from: http://dx.doi.org/10.1080/0969594X.2016.1268090

MacLellan, P. 2016. Why don't teachers use education research in teaching? *Education in chemistry. Royal Society of Chemistry*. [Online] Available from: https://edu.rsc.org/analysis/why-dont-teachers-use-education-research-in-teaching/2010170.article

Maclure, J. S. 1979. *Educational Documents England and Wales 1816 to the present day.* University press, Cambridge.

Major, J. 1991. Education – *All our futures.* Café Royal Speech made to the Centre for Policy Studies at the Cafe Royal in London on Wednesday 3rd July 1991. Available from:

http://www.johnmajorarchive.org.uk/1991/mr-majors-speech-on-education-to-centre-for-policy-studies-3-july-1991/

Mansell, W. 2015. *Exam cheating: time for full disclosure.* NAHT.

Mansell, W., James, M. and The ARG Group. 2009. *Assessment in schools: Fit for purpose? Commentary by the Teaching and Learning Research Programme.* London. Economic and Social Research Council, Teaching and Learning Programme.

Marshall, J. D. 1996. *Michel Foucault: Personal Autonomy and Education.* Kluwer Academic Publishers. Dordecht/Boston/London.

Mason, J. 2011. Semistructured Interview. In: The SAGE Encyclopedia of Social Science Research Methods. Lewis-Beck, M.S., Bryman, A and Futing Liao, Y. eds. SAGE Publications Inc. Thousand Oakes. [Online]. [Accessed 25th April 2019]. Available from: https://dx.doi.org/10.4135/9781412950589

Mason, M. 2019. *Foucault and his Panopticon.*[Online]. [Accessed 21st November 2019]. Available from: http://www.moyak.com/papers/michel-foucault-power.html

Maykut, P. and Morehouse, R. 1994. Beginning qualitative researchers: A philosophical and practical guide. Washington, DC: Falmer. Available from:
 https://journals.sagepub.com/doi/full/10.1177/160940690900800105

McCann, K. 2016. Liberal Democrats promise to end SATs tests for primary school children. [Online]. [Accessed 23rd September 2018]. *The Daily Telegraph,* Available from: https://www.telegraph.co.uk/news/2016/09/19/liberal-democrats-promise-to-end-sats-tests-for-primary-school-c/

McCulloch, G. 1993. Judgement of the Teacher: Norwood Report and internal examinations. [Online]. [Accessed May 8th 2019]. *International Studies in Sociology of Education*, 3:1, 129-143. Available from:
https://www.tandfonline.com/doi/pdf/10.1080/0962021930030107

McIntosh, J. 2015. *Final report of the Commission on Assessment without Levels.* DfE, September 2015. Available from:

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment
_data/file/483058/Commission_on_Assessment_Without_Levels_-_report.pdf

Meadows, M. and Billington, L. 2005. *A Review of the Literature on Marking* Reliability. Manchester: AQA. [online]. [Accessed 15th Mat 2019]. Available from: https://orderline.education.gov.uk/gempdf/1849625344/QCDA104983_review_of_the _literature_on_marking_reliability.pdf

Meadows, M. and Black, B. 2018. Teachers' experience of and attitudes toward activities to maximise qualification results in England, [Online]. [Accessed 21st October 2018]. *Oxford Review of Education, 44:5, 563-580, DOI: 10.1080/03054985.2018.1500355.* Available from: https://doi.org/10.1080/03054985.2018.1500355

Miles, M.B. and Huberman, M. A. 1994. *Qualitative Data Analysis.* SAGE Publications.

Millard, W., Small, I. and Menzies, L. 2017. *Testing the Water. How assessment can underpin, not undermine, great teaching. A report by LKMco Think Tank.* Published by Pearson Education, November 2017. Available from: https://www.pearson.com/uk/about-us/news-and-policy/reports-and-campaigns/testing-the-water-report.html

Miller, D.M., Linn, R. L. and Gronlund, N.E. 2009. *Measurement and Assessment in Teaching.* Tenth Edition. Pearson International Edition. Pearson Education, Inc., Upper Saddle River, New Jersey 07458.

Ministry of Education, 1945. *The Nation's Schools. Their Plan and Purpose. Ministry of Education Pamphlet No. 1.* [Online]. [Accessed 18th August 2019]. London: His Majesty's Stationery Office 1945. Available from: http://www.educationengland.org.uk/documents/minofed/pamphlet-01.html

Ministry of Education, 1947. *The New Secondary Education (1947) Ministry of Education Pamphlet No. 9.* [Online]. [Accessed18th August 2019]. London, His Majesty's Stationery Office. Available from: http://www.educationengland.org.uk/documents/minofed/pamphlet-09.html

Mobley, M., Emerson, C., Goodwin. and Letch, R. 1986. All About GCSE. A clear and concise summary of all the basic information about GCSE. Heinemann Educational Books.

Moray House Tests. 1962. Moray House Tests archive. Available from the University of Edinburgh Archive Service.

Morgan, D. L. 2007. Paradigms lost and Pragmatism Regained, Methodological Implications of Combining Qualitative and Quantitative Methods. [Online] [Accessed January 14th 2019]. *Journal of Mixed Methods Research Volume 1 Number 1 January 2007 48-76* © 2007 Sage Publications. Available from:
 https://0-journals-sagepub-com.wam.leeds.ac.uk/doi/pdf/10.1177/2345678906292462

Molina Azorin, J.M. and Cameron, R. 2010. The application of Mixed Methods on Organisational Research: A Literature Review. *The Electronic Journal of Business Research Methods* Volume 8 Issue 2 2010. pp. 95-105. Available from:
 www.ejbrm.com

Moray House Tests, 1962. *Authorities using unpublished Moray House tests during year ended 31 March 1962.* Edinburgh University Archive. Shelf mark, EUA/IN1/ACU/G1/1/8.

Montague, J.B. 1958. The "Eleven-Plus" Battle in England, *The Clearing House: A Journal of Educational Strategies, Issues and Ideas,* 32:5, pp. 259-262. [Online]. [Accessed 11th July 2019]. Available from:
https://www.tandfonline.com/doi/abs/10.1080/00098655.1958.11476122

Moyser, G. 2011. Elite Interviewing. In: Jupp, R. ed. *SAGE Dictionary of Social Research Methods.* [Online]. SAGE Publications Ltd. London. [Accessed 14th April 2019]. Available from: https://dx.doi.org/10.4135/9780857020116

Murphy, R. 2003. A Review of the Changing Views of Public Examinations. In: *Setting the Standard, A century of public examining by AQA and its parent boards.* Tattersall, K., Day, J., James, H., Gillan, D. and Spencer, A. AQA, England.

NAHT. 2014. *Report of the NAHT Commission on Assessment. National Association of Headteachers*. Available from:

https://www.stem.org.uk/system/files/elibrary-resources/2016/04/Assessment%20commission%20report%20document%20%282%29.pdf

NAHT. 2016. *Advice: LA moderators diverging from STA guidelines for teacher assessment. What to do when LA moderators diverge from STA guidelines for teacher assessment.* National Association of Headteachers. Published: 14/06/2016.

National Archives. Special Place Examination. Un-named Education Department Draft Circular, 1945. National Archives [Online] [Accessed 23 October 2019]. Available from: https://www.nationalarchives.gov.uk/wp-content/uploads/2014/03/ed147-1341.jpg

NEAB, 1994. *A Report on Coursework Examinations in English and English Literature,* NEAB, 1994.

Navarro, S. M. 2008. *Assessment and School Reform* in Dwyer, C.A., (Editor) *The Future of Assessment – shaping teaching and learning.* Lawrence Erlbaum Associates, New York and Oxford.

Newton, P. E. 2005. Clarifying the purposes of educational assessment. Paper prepared for submission to *Assessment in Education*.

Newton. P. E. 2007. Clarifying the purposes of educational assessment. In: Baker, E., McGaw, B. and Peterson, P. eds. *International Encyclopedia of Education.* Third Edition. Elsevier.

Newton, P. E. 2010. *Review of Key Stage 2 Testing and Assessment. Summary paper for Evidence Gathering Session*. Cambridge Assessment.

NFER. 2013. *A review of literature on marking reliability research; Report for Ofqual.* Tisi, J., Whitehouse, G., Maughan, S. and Burdett, N. Report commissioned by the Office of Qualifications and Examinations Regulation. Ofqual/13/5285. June 2013.

Norwood. 1943. *The Report of the Committee of the Secondary Schools Examination Council on Curriculum and Examinations in Secondary Schools.* Board of Education. The Norwood Report. London: HM Stationery Office 1943.

NUT. 1949. *Transfer from Primary to Secondary Schools. Report of a Consultative Committee appointed by the Executive of the National Union of Teachers.* The National Union of Teachers, Evans Brothers Ltd., London.

Oates, T. 2013. *Tiering in GCSE – which structure holds most promise?* Discussion paper. Cambridge Assessment. April 2013.

Oates, T. 2017. *The Cambridge Approach to Assessment. Principles for designing, administering and evaluating assessment.* January 2009 – revised April 2017. Available from:
https://www.cambridgeassessment.org.uk/Images/cambridge-approach-to-assessment.pdf

OECD. 2007. *PISA 2006: Science Competencies for Tomorrow's World.* OECD: Paris.

OECD. 2010. *PISA 2009 Results: Executive Summary.*

OECD. 2010a. *PISA 2009 Results: What Makes a School Successful? Resources, Policies and Practice.* OECD: Paris.

OECD. 2014. *Governing Complex Education Systems; Trust: what it is and why it matters for governance and education.* Unclassified. EDU/CERI/CD/RD(2014)14. 30-31 October 2014. Available from:
http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=EDU/CERI/CD/RD(2014)14&docLanguage=En

OECD. 2019. *TALIS 2018 Results (Volume I): Teachers and School Leaders as Lifelong Learners*, TALIS, Organisation for Economic Co-operation and Development (OECD). OECD Publishing, Paris. Available from:
https://doi.org/10.1787/1d0bc92a-en

Office for National Statistics. 2014. *2011 Census Analysis, Local Area Analysis of Qualifications Across England and Wales.*

Ofsted. 2007. The Annual Report of Her Majesty's Chief Inspector of Education, Children's Services and Skills 2006/07. October 2007. Available from:

http://www.educationengland.org.uk/documents/pdfs/2007-ofsted-annual-report.pdf

Ofsted. 2011. *The impact of the 'Assessing Pupils' Progress' initiative.* The Office for Standards in Education, Children's Services and Skills (Ofsted). Reference no: 100226. April 2011.

Ofsted. 2015. *The Annual Report of Her Majesty's Chief Inspector of Education, Children's Services and Skills 2014/15.* 1st December 2015. Print ISBN 9781474125314 Web ISBN 9781474125321.

Ofsted. 2016. *The Annual Report of Her Majesty's Chief Inspector of Education, Children's Services and Skills 2015/16.* 1st December 2016. Print ISBN 9781474138758 Web ISBN 9781 474138765.

Ofsted. 2016b. Busting myths on Ofsted inspections. [Online] Ofsted blog: schools, early years, further education and skills. 20th January 2016. [Accessed 20th January 2016]. Available from: https://educationinspection.blog.gov.uk/2016/01/20/busting-myths-on-ofsted-inspections/

Ofsted. 2018. *Ofsted Inspection Myths*. [Online]. [Accessed 2nd August, 2019]. Office for Standards in Education, Children's Services and Skills. Available from: https://www.gov.uk/government/publications/school-inspection-handbook-from-september-2015/ofsted-inspections-mythbusting

Ofsted. 2019. School inspection handbook. Handbook for inspecting schools in England under section 5 of the Education Act 2005. Reference number: 190017. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/843108/School_inspection_handbook_-_section_5.pdf

Ofsted. 2020. *Building great teachers? Initial teacher education curriculum research: phase 2.* Office for Standards in Education, Children's Services and Skills. January, 2020. Available from: https://www.gov.uk/government/publications/initial-teacher-education-curriculum-research/building-great-teachers

Ofqual. 2011. *A focus on teacher assessment reliability in GCSE and GCE.* Ofqual/11/4807 January 2011. The Office of Qualifications and Examinations Regulation. Available from:

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment _data/file/578874/2011-03-16-a-focus-on-teacher-assessment-reliability-in-gcse-and-gce.pdf

Ofqual. 2011a. *OF127 Estimation of teacher assessment reliability specification.* Archived document. Available from:

https://www.google.com/search?client=safari&rls=en&q=data.gov.uk+%E2%80%BA+dat a+%E2%80%BA+contracts-finder-archive+%E2%80%BA+download&ie=UTF-8&oe=UTF-8

Ofqual. 2012. *GCSE English 2012 – Final Report.* Ofqual/12/5225.

Ofqual. 2013. *Reliability of assessment: compendium.* Published May, 2013. Available from:

https://www.gov.uk/government/publications/reliability-of-assessment-compendium

Ofqual. 2014. *Schools, colleges and children's services – press release.* Available from:

https://www.gov.uk/government/news/gcse-a-level-and-as-qualification-updates-from-ofqual

Ofqual. 2015. *GCSE (Graded A* to G) Controlled Assessment Regulations.* Available from:

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment _data/file/446029/gcse-graded-a-star-to-g-controlled-assessment-regulations.pdf

Ofqual. 2015a. *Grading new GCSEs from 2017.* Available from:

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/537147/Po stcard_-_Grading_New_GCSEs.pdf

Ofqual. 2016. *Summary of changes to GCSEs from 2015.* Available from:

https://www.gov.uk/government/publications/gcse-changes-a-summary/summary-of-changes-to-gcses-from-2015

Ofqual. 2017. *Teacher involvement in developing exam papers. Findings from our call for evidence.* The Office of Qualifications and Examinations Regulation. (Ofqual). Available from:

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/690002/Teacher_involvement_in_developing_exam_papers_Findings_from_our_call_for_evidence_.pdf

Ofqual. 2018. *Research and Analysis, Marking consistency metrics. An update.* The Office of Qualifications and Examinations Regulation. (Ofqual). Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/759207/Marking_consistency_metrics_-_an_update_-_FINAL64492.pdf

Ofqual. 2020a. *Summer 2020 grades for GCSE, AS and A level, Extended Project Qualification and Advanced Extension Award in maths: Information for Heads of centre, Heads of Department and teachers on the submission of Centre assessment grades.* The Office of Qualifications and Examinations Regulation. (Ofqual). Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/877930/Summer_2020_grades_for_GCSE_AS_A_level_EPQ_AEA_in_maths_-_guidance_for_heads_of_centres.pdf

Ofqual, 2020b. *Awarding technical and vocational qualifications this summer.* The Office of Qualifications and Examinations Regulation. (Ofqual). 9th April, 2020. Available from: https://www.gov.uk/government/news/awarding-vocational-and-technical-qualifications-this-summer

Ofqual. 2020c. *About us.* The Office of Qualifications and Examinations Regulation, Ofqual. Available from:
https://www.gov.uk/government/organisations/ofqual/about

O'Niell, O. 2002. Onora O'Neill, Reith Lecture 3, 2002. Cited in: Mansell, W., James, M. and The ARG Group. 2009. *Assessment in schools: Fit for purpose? Commentary by the Teaching and Learning Research Programme.* London. Economic and Social Research Council, Teaching and Learning Programme. Available from:
https://www.bbc.co.uk/programmes/p00gpzdf

Opposs, D. and He, Q. 2011. *The Reliability Programme. Final Report.* The Office of Qualifications and Examinations Regulation. (Ofqual). Available from:

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/578899/2011-03-16-the-reliability-programme-final-report.pdf

Opposs, D. and He, Q. eds. 2012. *Ofqual's Reliability Compendium.* The Office of Qualifications and Examinations Regulation. (Ofqual).

Opposs, D. 2016. Whatever happened to school-based assessment in England's GCSEs and A levels? *Perspectives in Education*, Volume 34 Number 4, Dec 2016, p. 52 – 61. Publisher; University of the Free State.

Oxford Dictionary 2019. [Online]. [Accessed 21st April 2019]. Available from:
 https://en.oxforddictionaries.com/definition/develop

Parliamentary Education Select Committee. 2020. Ofqual session, 2nd September, 2020. Available from:
https://parliamentlive.tv/Event/Index/a3d523ca-09fc-49a5-84e3-d50c3a3bcbe3

Patten, J. 1992. 'Education League Tables; School Report' in *The Times*, 19th November 1992.

Peal, R. 2014. *Playing the Game - The enduring influence of the preferred Ofsted teaching style.* Civitas: Institute for the Study of Civil Society.

Pellegrino, J.W. 1999. The Evolution of Educational Assessment: Considering the past and imagining the future. *The sixth William H. Angoff Memorial Lecture. Education and Testing Service (ETS),* Princeton, New Jersey, 7th November 1999.

Pellegrino, J.W. and Goldman, S.R. 2008. Beyond Rhetoric: Realities and Complexities in Integrating Assessment into Classroom Teaching and Learning. In: Dwyer, C.A. ed. *The Future of Assessment – shaping teaching and learning.* Lawrence Erlbaum Associates, New York and Oxford.

Percy, Lord. 1967. Article in the Times Educational Supplement, June 1967. In: Gillan, D. The Northern CSE Boards, p. 95. In: *Setting the Standard, A century of public examining by AQA and its parent boards.* Tattersall, K., Day, J., James, H., Gillan, D. and Spencer, A. AQA, England.

Pilgrim's Cross Primary School. 2019. *Assessment for Children Charter.* Available from: https://www.pilgrimscross.co.uk/assessment-for-children-charter/

PISA. 2011. *PISA in Focus: School autonomy and accountability: Are they related to student performance?* OECD.

Popham, W. J. 2001. Teaching to the Test? How can we dissuade schools from "item teaching" and return the focus to the curriculum? *Educational Leadership.* March 2001, Volume 58. Number 6. Helping All Students Achieve. pp. 16-20.

Popham, W. J. 2010. *Everything School Leaders Need to Know About Assessment.* SAGE.

QCA. 2002. *Assessment and reporting arrangements: key stage 1, 2002.* London.

QCA. 2003. *Financial Modelling of the English Exams System 2003-04. A report by PriceWaterhouseCoopers for the Qualifications and Curriculum Authority. (QCA).* London.

QCA. 2005. *A review of GCE and GCSE coursework arrangements.* Qualifications and Curriculum Authority. London.

QCA. 2006. *A review of GCSE coursework.* Qualifications and Curriculum Authority. London.

QCA. 2009. *Research into marking quality: studies to inform future work of National Curriculum assessment.* London: Qualifications and Curriculum Authority.

QCA. 2009a. *Developing whole-school assessment.* Qualifications and Curriculum Authority. March 2009 QCA/09/4110. Available from: https://dera.ioe.ac.uk/2621/7/Developing%20whole-school%20assessment%20March%202009_tcm8-14489_Redacted.pdf

QCDA. 2010. *Assessing Pupils' Progress: learners at the heart of assessment.* Qualifications and Curriculum Development Agency. QCDA/10/4768.

QCDA. 2011. *Day-to-day, periodic and transitional assessment.* Available from:

http://archive.teachfind.com/qcda/curriculum.qcda.gov.uk/key-stages-3-and-4/assessment/assessment_and_curriculum/day-to-day-periodic-and-transitional-assessment/index.html

Ratcliffe, R. 2014. The verdict on Ofsted: 'requires improvement'? [Online]. The Guardian. 28th October 2014. [Accessed 30th September 2017]. Available from: https://www.theguardian.com/education/2014/oct/28/-sp-verdict-ofsted-requires-improvement

Rethinking Assessment. 2020. *Open Letter to the Sunday Times explaining the reasons for Rethinking Assessment. Launching a broad coalition of state and independent schools, universities and academics, employers and stakeholders.* Rethinking Assessment Group. Available from:
 https://rethinkingassessment.com/rethinking-blogs/open-letter-to-the-sunday-times-explaining-the-reasons-for-rethinking-assessment/

Reynolds, J.M. 2004. "Pragmatic Humanism" in Foucault's Later Work. *Canadian Journal of Political Science / Revue canadienne de science politique*, Dec., 2004, Vol. 37, No. 4 (Dec., 2004), pp. 951-977. Available from: https://www.jstor.org/stable/pdf/25165736.pdf?refreqid=excelsior%3Aa31382f4382d57a0872007ed1029da19

Rimfeld, K., Malanchini, M., Hannigan., L J., Dale, P. J., Allen, R., Hart, S. A., and Robert Plomin, R. 2019. Teacher assessments during compulsory education are as reliable, stable and heritable as standardized test scores. In: *The Journal of Child Psychology and Psychiatry.* pp. 1278 -1288. Published by John Wiley and Sons Ltd, Oxford. doi:10.1111/jcpp. 13070. Available from: https://acamh.onlinelibrary.wiley.com/doi/full/10.1111/jcpp. 13070

Robbins Report. 1963. *Report of the Committee on Higher Education appointed by the Prime Minister.*

Robinson, W. 2006. Teacher Training in England and Wales: Past, Present and Future Perspectives*. Education Research and Perspectives* Vol. 33, No. 2, 2006.

Roberts, N. and Foster, D. 2016. *Initial teacher training in England.* House of Commons Briefing Paper Number 6710, 9 May 2016.

Rogers, T. 2017. Why Jessica quit teaching: She'd given her heart and soul to those kids but SLT didn't think it was enough. [Online]. [Accessed 17th November 2017]. *Times Educational Supplement. 11th November, 2017*. Available from: https://www.tes.com/news/why-jessica-quit-teaching-shed-given-her-heart-and-soul-those-kids-slt-didnt-think-it-was

Rosenkvist , M. A. 2010. *Using student test results for accountability and improvement: a literature review*. OECD Education Working Paper No. 54.

Rightmove. 2013. Strathmore Drive, Kirklevington, TS15 9NS. [Online]. [Accessed 4th January 2014]. Available from: https://www.rightmove.co.uk/house-prices/detailMatching.html?prop=39706915&sale=82232961&country=england

Rubin, H.J. and Rubin, I.S. 1995. Qualitative Interviewing. The Art of Hearing Data. Sage Publications, Inc.

Saidin, K. and Yaacob, A. 2016. Insider researchers: challenges & opportunities. Proceeding of ICECRS, 1 (2016) 849-854. International Seminar on Generating Knowledge Through Research, UUM-UMSIDA, 25-27 October 2016, Universiti Utara Malaysia, Malaysia. [Online]. [Accessed 1st December 2020]. Available from: https://core.ac.uk/download/pdf/154353144.pdf

Saldana, J. 2009. *The Coding Manual for Qualitative Researchers.* SAGE Publications Ltd.

Sarantakos, S. 1998*. Social Research.* Second Edition. London: Macmillan Press.

Saunders, M. N. K. 2012. Choosing Research Participants. In: Cassell, C. and Symon, G. eds. 2012. *Qualitative Organisational Research. Core methods and current challenges.* SAGE Publications. London.

Sauro, J. 2015. *5 Things to know about Liker scales.* [Online]. [Accessed 22nd May 2017]. Available from: https://measuringu.com/

Schirato, T., Danaher. G. and Webb, J. 2012. *Understanding Foucault; a critical introduction.* 2nd Edition. SAGE Publications Ltd.

Schools Week. 2018. Ofqual clamps down on teacher-examiner cheating. [Online]. *Schools Week on-line.* Pippa Allen-Cross, Wednesday 14th March, 2018. [Accessed 14th March 2018]. Available from: https://schoolsweek.co.uk/ofqual-clamps-down-on-teacher-examiner-cheating/

Scott, J. 1990. *A Matter of Record, Documentary Sources in Social Research.* Polity Press, Cambridge. UK.

SEC. 1986. *Working paper 3: policy and practice in school-based assessment*. Secondary Examinations Council. London.

Seidman, I. 2006. Interviewing as qualitative research: a guide for researchers in education and the social sciences. Third edition. New York; Teachers College Press.

Seidman, I. 2013. Interviewing as qualitative research: a guide for researchers in education and the social sciences. Fourth edition. New York; Teachers College Press.

Shaffner, M. 2008. *Assessment Professional Development guide.* Edutopia. Available from: https://www.edutopia.org/assessment-guide

Shannon-Baker, P. 2016 Making Paradigms Meaningful in mixed Methods Research. *Journal of Mixed Methods Research 2016, Vol. 10(4) 319–334 Ó* [Online]. [Accessed 15th January 2019]. Available from*:* https://0-journals-sagepub-com.wam.leeds.ac.uk/doi/pdf/10.1177/1558689815575861

Sharp, P. 1997. *The Development of the Vocational Curriculum for 16-19 year-olds in Colleges and Schools 1979-1995.* [Online]. School of Education, University of Leeds. [Accessed 14th August 2019]. Available from: http://www.leeds.ac.uk/educol/documents/00002218.htm

Sharp, W. 2019. *The Educated Guess. How to Challenge invisible bias and make better education choices.* Warwick Sharp: Creative works and print by Spiffing Covers.

Shaw, S. and Crisp, V. 2011. Tracing the evolution of validity in educational measurement: past issues and contemporary challenges. *Research Matters: A Cambridge Assessment publication.* UCLES 2011. Available from:

https://www.cambridgeassessment.org.uk/Images/471470-tracing-the-evolution-of-validity-in-educational-measurement-past-issues-and-contemporary-challenges.pdf

 Sheffield Institute of Education. 2019. *The Curriculum for Initial Teacher Education: Literature Review.* Sheffield Hallam University. March 2019. Authors: Dr Emily Perry, Dr Josephine Booth, Dr David Owen, Dr Kim Bower. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment _data/file/839656/Literature_Review_of_Initial_Teacher_Education_Curriculum_- _final.pdf

Shorrocks-Taylor, D. 1999. *National Testing: Past, Present and Future*. British Psychological Society. UK.

Silverman, D. 2006. *Interpreting Qualitative Data*. Third Edition. SAGE Publications.

Simon, B. 1954. *Intelligence Testing and the Comprehensive School.* Lawrence & Wishart Ltd. London.

Simons, M. and Masschelien, J. 2006. The permanent quality tribunal in education and the limits of education policy. *Policy Futures in Education.* 4(3). pp. 294–307.

Smart, B. 2002. *Michel Foucault.* Revised edition (2002) Routledge, New York. First published 1985 by Ellis Horwood Limited and Tavistock Publications Ltd.

Socialist review. 1996. The Intelligence Fraud. *Socialist review. Issue 196.* [Online] Published April 1996. [Accessed 16th August 2019]. Available from: http://pubs.socialistreviewindex.org.uk/sr196/parrington.htm

Special Place Examination. 1945. Extract from an un-named Education Department draft Circular on the introduction of the 11 plus examination, July 1945. [Online]. National Archive. [Accessed 1 September 2019]. Available from: https://www.nationalarchives.gov.uk/wp-content/uploads/2014/03/ed147-1341.jpg

Spencer, A. 2003. Changes in the Nature of Assessment. In: *Setting the Standard, A century of public examining by AQA and its parent boards.* Tattersall, K., Day, J., James, H., Gillan, D. and Spencer, A. AQA, England.

Spens Report. 1938. Report of the Consultative Committee on secondary education, with special reference to Grammar Schools and Technical High Schools ... (Spens Report), London, H.M.S.O.

Spielman, A. 2017. *Speech at the Festival of Education. 23 June 2017*. Available from: https://www.gov.uk/government/speeches/amanda-spielmans-speech-at-the-festival-of-education

STA. 2013. National Curriculum Optional Tests. Bulk Orders 2008 to 2012. Source: Standards and Testing Agency (STA). 2012. Internal document.

STA. 2018. *2018 Assessment and Reporting Arrangements (ARA).* Standards and Testing Agency, October 2018. Reference: STA/18/8180/e ISBN: 978-1-78644-940-5. Available from:
www.gov.uk/government/publications

STA. 2019. *2018 Maladministration Report. National curriculum assessments at key stage 1 and key stage 2.* Standards and Testing Agency, Available from:
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/858413/2018_maladministration_report.pdf

Stacey, G. 2014. *GCSE, A level and AS qualification updates.* Ofqual press release. April 9th, 2014.

Stanley, G., MacCann, R., Gardner, J., Reynolds, L. and Wild, I. 2009. *Review of teacher assessment: evidence of what works best and issues for development.* Oxford University Centre for Educational Assessment for the Qualifications and Curriculum Authority (QCA). London.

Starch, D. and Elliot, E. C. 1912. Reliability of the Grading of High-School Work in English. *The School Review*, Vol. 20, No. 7 (Sep. , 1912), pp. 442-457. [Online] Published by: The University of Chicago Press. [Accessed 7th April 2020]. Available from:
 https://www.jstor.org/stable/1076706

Stickney, J. A. 2012. Judging Teachers: Foucault, governance and agency during educational reforms. *Educational Philosophy and Theory*. Available from:

https://www.tandfonline.com/doi/abs/10.1111/j.1469-5812.2011.00752.x

Stiggins, R.J. 1991. *Assessment literacy.* Phi Delta Kappan.

Stiggins, R.J. and Conklin, N.F. 1992. *In Teachers' Hands – Investigating the Practices of Classroom Assessment*. State University of New York Press, Albany.

Stiggins, R.J. 2002. *Assessment Crisis: The Absence of Assessment FOR Learning*. PHI Delta Kappan.

Stiggins, R. 2008. Correcting "Errors of Measurement" That Sabotage Student Learning. In: Dwyer, C.A. ed. *The Future of Assessment – shaping teaching and learning.* Lawrence Erlbaum Associates, New York and Oxford.

Stiggins, R. 2014. *Improve assessment literacy outside of school too.* PHI Delta Kappan. V96 N2. 67-72.

Stobart, G. 2008. *Testing Times – The uses and abuses of assessment*. Routledge, Abingdon, Oxon 7 New York.

St. Pierre, E. and Youngblood Jackson, A. 2014. Qualitative Data Analysis After Coding. *Qualitative Inquiry. June 2014.* [Online]. DOI: 10.1177/1077800414532435. [Accessed 9th June 2019]. Available from:
https://journals.sagepub.com/doi/full/10.1177/1077800414532435

Stringer, N. 2014. *The Achieved Weightings of Assessment Objectives as a Source of Validity Evidence*. The Office of Qualifications and Examinations Regulation. Ofqual/14/5375.

Survey Monkey. 2019. [Online]. [Accessed 15th April 2019]. Available from:
https://www.surveymonkey.com/curiosity/how-often-should-you-use-email-reminders/

Sutherland, S. 2008. *An independent inquiry into the delivery of National Curriculum tests in 2008. A report to Ofqual and the Secretary of State for Children, Schools and Families.* The Sutherland Inquiry. Ordered by the House of Commons to be printed on 8th December 2008. London. The Stationery Office. Ref: ISBN 9780102958393, HC 62 2008-09. Available from:

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment
_data/file/229092/0062.pdf

Tashakkori, A. and Teddlie, C. 1998. *Mixed Methodology Combining; Qualitative and Quantitative Approaches*. Applied Social Research Methods Series, Volume 46. SAGE Publications.

Tattersall, K. 2003. Ringing the changes: educational and assessment policies, 1900 to the present. In: *Setting the Standard, A century of public examining by AQA and its parent boards.* Tattersall, K., Day, J., James, H., Gillan, D. and Spencer, A. Assessment and Qualifications Alliance (AQA), England.

Tattersall, K. 2007. A brief history of policies, practices and issues relating to comparability. In Techniques for monitoring the comparability of examination standards. A collection of research papers for QCA. In: Newton, P., Baird, J., Goldstein, H., Patrick, H. and Tymms, P. eds. Qualifications and Curriculum Authority. Available from:
https://www.gov.uk/government/publications/techniques-for-monitoring-the-comparability-of-examination-standards

Taylor, J. 2015. Merseyside teacher banned for life after helping pupils cheat. [Online*]. Liverpool Echo, 29th May, 2015*. [Accessed 14th August 2019]. Available from:
https://www.liverpoolecho.co.uk/news/liverpool-news/merseyside-teacher-banned-pupils-cheat-9351572

Taylor, M. 1992. *The reliability of judgements made by coursework assessors.* Associated Examining Board internal report.

TDA. 2007. *Professional Standards for Teachers. Why sit still in your career? Training and Development Agency for Schools.* Available from:
https://www.rgs.org/NR/rdonlyres/13C47A9B-633C-436F-8617-668966AEAEB7/0/CGT_Online_TDA_standards2007.pdf

Teach First. 2017. *Get into Teaching.* DfE. Available from:
https://getintoteaching.education.gov.uk/explore-my-options/teacher-training-routes/specialist-training-options/teach-first

Teaching Schools Council. 2020. Post Graduate Teaching Apprenticeship. Available from: https://tscouncil.org.uk/postgraduate-teaching-apprenticeship/

Teddlie, C. and Tashakkori, A. 2009. *Foundations of Mixed Methods Research. Integrating Quantitative and Qualitative Approaches to the Social and Behavioural Sciences.* SAGE Publications, Inc.

TGAT Report, 1988. *National Curriculum Task Group on Assessment and Testing.* Department of Education and Science and the Welsh Office.

Thane, P. 1982. The Foundations of the Welfare State, London, Longman, 1982, p. 204. In: Whitty, G. and Power, S. 2015. *'Selective, comprehensive and diversified secondary schooling in England: a brief history.'* In: de Waal, A, ed. *The ins and outs of selective secondary schools: a debate.* Civitas, London, pp. 9-28. ISBN. [Online]. [Accessed 15th September 2019]. Available from:
https://www.civitas.org.uk/pdf/theselectiondebate

The Daily Telegraph, 2016. Compare your school's performance: primary, GCSE and A-level results. *The Daily Telegraph,* 5th May, 2016. Available from:
https://www.telegraph.co.uk/education-and-careers/league-tables/

The Huddersfield Examiner. 2020. The climb to the top of the class. Published Thursday January 9th 2020.

The Independent. 2017. *Public school cheating scandal sparks review into exam rules.* Benjamin Kentish, 17th August, 2017. [Online]. [Accessed 17th August 2017]. Available from:
https://www.independent.co.uk/news/education/public-school-cheating-exams-questions-papers-ofqual-eton-college-winchester-radley-charterhouse-a7922976.html

The Mail Online. 2005. 2005 primary school performance tables. Published 5th December 2005.

The National Strategies. 2011. *The National Strategies 1997–2011 A brief summary of the impact and effectiveness of the National Strategies.* Date of issue: 03-2011. Ref: 00032-2011PDF-EN-01**.** Published by the Department for Education, England. Available from:

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/175408/DFE-00032-2011.pdf

The Sunday Times. 2017. Reaching new heights of SATisfaction. Article by Professor Alan Smithers. 19th November, 2017.

The Sunday Times. 2019. *The school is free – but it's £290,000 to live nearby.* Published November 24th 2019.

The Telegraph, 2017. *Exclusive: Three of Britain's leading independent schools caught up in an exam 'cheating' scandal.* Harry Yorke, On-line editor. 27th August, 2017 9:30pm. [Online]. [Accessed 28th August 2017]. Available from: https://www.telegraph.co.uk/education/2017/08/27/exclusive-three-britains-leading-independent-schools-caught/

The Times. 2020. *University leaders call for next year's A-levels to be axed*. Nicola Woolcock, Education Correspondent, The Times, Wednesday 30th September 2020.

Thom, D. 1986. The 1944 education Act: the 'art of the possible'? In: Smith, H.L. 1986. ed. *War and Social Change British Society in the Second World War.* Manchester University Press.

Thomas, G. 2011. How to do your Research Project. A guide for students and researchers. SAGE Publications Ltd.

Thomas, G. 2013. How to do your Research Project. A guide for students in applied social sciences. 2nd Edition. SAGE Publications Ltd.

Tooley, J. and Darby, D. 1998. Educational research a critique: a survey of published educational research. London. Report presented to OFSTED.

Turkel, G. 1990. Michel Foucault: Law, Power and Knowledge. *Journal of Law and Society, Vol 17 No.2 (Summer, 1990) pp. 170-193.* Published by Wiley on behalf of Cardiff University. Available from: http://www.jstor.org/stable/1410084

Turner, C. 2018. Teachers prepare to strike in protest at 'immoral' exams for four-year-olds. [Online]. [Accessed 3rd April 2018]. *The Telgraph,* 3rd April 2018. Available from:

https://www.telegraph.co.uk/education/2018/04/03/teachers-prepare-strike-protest-immoral-exams-four-year-olds/

Treadaway. M. 2015. *We worry about teachers inflating results; we should worry more about depression of baseline assessments.* FFT Datalab. 5th March, 2015. Available from: https://ffteducationdatalab.org.uk/2015/03/we-worry-about-teachers-inflating-results-we-should-worry-more-about-depression-of-baseline-assessments/

Twiselton, S. 2000. Seeing the Wood for the Trees: The National Literacy Strategy and Initial Teacher Education; pedagogical content knowledge and the structure of subjects. *Cambridge Journal of Education, 30:3. pp. 391-403*, DOI: 10.1080/713657157. Published online: 01 Jul 2010. Available from: https://doi.org/10.1080/713657157

Twiselton, S. 2004. The role of teacher identities in learning to teach primary literacy, *Educational Review, 56:2.pPp. 157-164,* DOI: 10.1080/0031910410001693245. Published online: 07 Aug 2006. Available from: https://www.tandfonline.com/doi/abs/10.1080/0031910410001693245

Twiselton, S. 2007. Seeing the wood for the trees: learning to teach beyond the curriculum. How can student teachers be helped to see beyond the National Literacy Strategy?, *Cambridge Journal of Education, 37:4. pp. 489-502,* DOI: 10.1080/03057640701705849. Published on line: 2 September 2010. Available from: https://doi.org/10.1080/03057640701705849

Twiselton, S. 2019. Seeing the wood for the trees. The Development of Teacher Expertise – conditions, and implications. *Presentation, JUSCO Conference, Liverpool. February, 2019. Junior School Collaboration (JUSCO).* Available on-line in the members area from: https://www.juniorschoolcollaboration.co.uk

UCAS. 2020. *Predicted grades – what you need to know.* [Online]. The Universities and Colleges Admissions Service. [Accessed 11th January 2020]. Available from: https://www.ucas.com/advisers/managing-applications/predicted-grades-what-you-need-know

UCET. 2017. UCET News, Jobs and Events: weekly bulletin w/e 8th September 2017.

UKRLP. 2017. United Kingdom Register of Learning Providers (UKRLP). [Online]. [Accessed 2nd May 2018]. Available from: https://www.ukrlp. co.uk/ukrlp/ukrlp. first

UN. 1948. The Universal Declaration of Human Rights. United Nations, 10th December 1948, Paris. Available from: https://www.un.org/en/universal-declaration-human-rights/

UNESCO, 2019. Glossary. UNESCO Institute for Statistics (UIS).[Online]. [Accessed November 23rd 2019]. Available from: http://uis.unesco.org/en/glossary-term/high-stake-assessment

van Rijn, P. W., Béguin, A., A. and Verstralen. H..H.F. M. 2012. Educational Measurement Issues and Implications of High-stakes Decisions Making in Final Examinations in Secondary Education in the Netherlands. *Assessment in Education 19: 117–136.*

Vaughan, R. 2015. UK among world's worst for 'teaching to the test', research finds. *TES, on-line. 18 December 2015.* Available from: https://www.tes.com/news/uk-among-worlds-worst-teaching-test-research-finds

Waddell Report, 1978. *School Examinations. Report of the Steering Committee established to consider proposals for replacing the General Certificate of Education Ordinary Level and Certificate of Secondary Education examinations by a common system of examining*. London: Her Majesty's Stationery Office 1978. [Online]. Accessed 11th March 2019]. Available from:
http://www.educationengland.org.uk/documents/waddell/waddell1978.html#105

Wang, C. 2011. Power/Knowledge for Educational Theory: Stephen Ball and the Reception of Foucault. *Journal of Philosophy of Education. Vol, 45, No. 1, 2011.* Blackwell Publishing. Available from:
https://onlinelibrary.wiley.com/doi/full/10.1111/j.1467-9752.2011.00789.x

Ward, H. 2017. Investigation into double entry of maths GCSE. *TES. 25th September 2017.* Available from:
https://www.tes.com/news/school-news/breaking-news/exclusive-investigation-double-entry-maths-gcse-candidates

Ward, H. 2017. Government announces details of postgraduate teaching apprenticeships. *TES. 19th October 2017*. Available from:

https://www.tes.com/news/school-news/breaking-news/government-announces-details-postgraduate-teaching-apprenticeships

Wellington, J. 2015. *Educational Research: Contemporary Issues and Practical Approaches.* Bloomsbury Publishing.

Whetton, C. 2009. A brief history of a testing time: National Curriculum assessment in England 1989–2008. *Educational Research, 51:2, pp. 137-159,* [Online]. [Accessed 20th May 2019]. DOI: 10.1080/00131880902891222. Available from: https://vle.canetwork.org.uk/pluginfile.php/5916/mod_label/intro/A%20brief%20history%20of%20a%20testing%20time%20-%20Chris%20Whetton.pdf

White Paper. 1943.Educational Reconstruction (1943). London: Board of Education 1943.

White Paper. 1978. *Secondary School Examinations - A single system at 16 plus.* London: Her Majesty's Stationery Office 1978.

Whitty, G. 2001. Teacher professionalism in new times. In: Gleeson, D. and Husbands, C. eds. *The Performing School. Managing teaching and learning in a performance culture.* Routledge. London and New York.

Whitty, G. and Power, S. 2015. *Selective, comprehensive and diversified secondary schooling in England: a brief history* In: de Waal, A, ed. *The ins and outs of selective secondary schools: a debate.* Civitas, London, pp. 9-28. . [Online]. ISBN 9781906837716. [Accessed 15th September 2019]. Available from: http://www.civitas.org.uk/content/files/theselectiondebate.pdf

Wiliam, D. 1998. Paper presented to Working Group 6 (Research on the Psychology of Mathematics Teacher Development) of *The 22nd annual conference of the International Group for the Psychology of Mathematics Education;* Stellenbosch, South Africa, July 1998.

Wiliam, D. 2001. Reliability, validity, and all that jazz, Education 3-13, 29:3, pp. 17-21. DOI: 10.1080/03004270185200311. Available from: https://doi.org/10.1080/03004270185200311

Wiliam, D. and Thompson, M. 2008. Integrating assessment with learning: What will it take to make it work? In: C.A. Dwyer. ed. *The future of assessment: Shaping teaching and learning,* pp. 53–82. New York: Erlbaum.

Wiliam, D. 2013. Assessment. The Bridge between teaching and Learning. In; *Voices from the Middle, Volume 2, December 2013.* National Council of Teachers of English. Available from:
http://www.ncte.org/library/NCTEFiles/Resources/Journals/VM/0212-dec2013/VM0212Assessment.pdf

Wilkins, C. 2011. Professionalism and the post-performative teacher: new teachers reflect on autonomy and accountability in the English school system*, Professional Development in Education, 37:3. pp. 389-409,* DOI: 10.1080/19415257.2010.514204. Available from**:**
https://doi.org/10.1080/19415257.2010.514204

Williamson, G. 2020. Impact of Covid-19 on Summer Exams: Written statement - HLWS170. Gavin Williamson, Secretary of State for Education. Available from:
https://www.parliament.uk/business/publications/written-questions-answers-statements/written-statement/Lords/2020-03-23/HLWS170/

Williamson, G. 2020a. Statement by the Secretary of State for Education to the Education Select Committee on *The effects of COVID-19 on Education and Children's Services*. Wednesday April 2nd 2020. Unpublished as of August 2020.

Williamson, G. 2020b. Extra measures to support students ahead of next summer's exams. DfE Press release, 3 December 2020. Department for Education, England. Available from:
https://www.gov.uk/government/news/extra-measures-to-support-students-ahead-of-next-summers-exams

Wilshaw, M. 2014. *Letter from Sir Michael Wilshaw (HMCI) to all schools*. Ofsted. July, 2014.

Wilson, M. 2014. Quality educational research outputs and significance of impact: enduring dilemma or stimulus to learning transformations between multiple communities of practice?

Hillary Place Papers, University of Leeds, 2014.

Wilson, M. 2017. *Session 1, EDUC 5061M Philosophical Underpinnings of Educational Research. Knowledge Claims and the Postivistic Paradigm.* University of Leeds School of Education.

Wilson, R. 2020. It takes five years to become a teacher. P and R Education Ltd.

Wolf, A. 2002. Qualifications and Assessment. In: Aldrich, A. ed. *A Century of Education*. Routledge.

Wolf, A. 2008. Looking for the best results. In*: Make the Grade,* Chartered Institute of Educational Assessors. (CIEA).

Wolf, A. 2011. Review of Vocational Education. The Wolf Report. DfE. Available from: https://www.gov.uk/government/publications/review-of-vocational-education-the-wolf-report

Wong, L, F. and Kaur, B. 2015. A Study of Mathematics Written Assessment in Singapore Secondary Schools. In: *The Mathematics Educator, 2015 Vol.16, No.1, pp. 19-44.*

Woods, H. 2019. Examining the Implications of Insider-Outsider Positioning. [Online]. [Accessed 30th September, 2019]. Available from:
 https://heatherawoods.ca/research/examining-the-implications-of-insider-outsider-positioning/

Yates, A. and Pidgeon, D. A. 1957. *Admission to Grammar Schools. Third Interim Report on allocation of Primary School Leavers to course of Secondary Education.* National Foundation for Educational Research. Newnes Educational Publishing Co., Ltd. London.

Appendix 1: Main routes to qualified teacher status in England

Route[1]	Number started 2014-15	Number started 2015-16	Number started 2016-17	Recruitment & training design	Training delivery	Student or employed	Qualification
University-led (undergraduate)	5,936[2] (18)	5,440[2,3] (16)	5,195[4] (16)	University	University	Student	BA, BSc or BEd with QTS
University-led (post-graduate)	14,695[2] (44)	13,561[2,3] (41)	11,992[4] (37)	University	University	Student	QTS and PGCE
School Direct – post-graduate (fee)[7]	6,311[2] (19.5)	7,086[2,3] (21)	7,470[4] (23)	School	School-centred, university mix	Student	QTS usually with PGCE
School Direct – post-graduate (salaried)[8]	2,759[2] (8)	3,166[2,3] (9.5)	3,159[4] (10)	School	School-centred, university mix	Employee	QTS usually with PGCE
School centred ITT (post-graduate)	1,988[2] (6)	2,372[2,3] (7)	3,057[4] (9.5)	School-centred provider	School-centred	Student	QTS usually with PGCE
Teach First (post-graduate)	1,426[5] (4)	1,584[2,3] (5)	1,375[4] (4)	Teach First	Teach First and university	Employee	QTS and PGDE
Total	33,115	33,209	32,248				

Notes:

1. Table based on Allen et al., Institute of Fiscal Studies Report 118 – Table 1.1 p. 8, and DfE ITT data – see links below.
2. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/477891/Main_tables_SFR_46_2015_to_2016.xls
3. Data for 2015–16 are provisional.
4. https://www.gov.uk/.../initial-teacher-training-trainee-number-census-2016-to-2017
5. https://graduates.teachfirst.org.uk/sites/graduates.teachfirst.org.uk/files/2014%20cohort%20profile.pdf
6. Numbers in parenthesis show the rounded percentage share of the total number of trainees.
7. School Direct (fee), a school-led route where participating schools contract accredited training providers and then recruit, select and employ their own trainees; see Allen et al. (2016) for more details.
8. School Direct unsalaried routes receive payments from ITT providers. These vary depending on the route and the length of placement; see Allen et al. (2014) for more details.

References

Allen, R., Belfield, C., Greaves, E., Sharp, C. and Walker, M. 2014. *The costs and benefits of different initial teacher training routes*. Institute for Fiscal Studies (IFS) Report R100. Available from:
 https://www.ifs.org.uk/uploads/publications/comms/r100.pdf

Allen, R., Belfield, C., Greaves, E., Sharp, C. and Walker, M. 2016. *The Longer-Term Costs and Benefits of Different Initial Teacher Training Routes.* Institute for Fiscal Studies (IFS) Report R118. Available from:
https://www.ifs.org.uk/uploads/publications/comms/R118.pdf

Appendix 2: Initial Teacher Training Providers – top, bottom and middle in terms of trainee numbers 2016-17[1]

Top 20 ITT providers in terms of numbers of trainees 2016-17				
Provider name	Region	Total	Postgraduate	Undergraduate
Total	**England**	**32,248**	**27,053**	**5,195**
UCL Institute of Education	London	1,333	1,333	0
Canterbury Christ Church University	South East	1,212	972	240
Edge Hill University	North West	1,154	652	502
Sheffield Hallam University	Yorkshire and The Humber	1,039	809	230
Manchester Metropolitan University	North West	1,003	761	242
University of Cumbria	North West	773	594	179
Birmingham City University	West Midlands	715	537	178
Liverpool Hope University	North West	572	421	151
University of Brighton, School of Education	South East	553	316	237
St Mary's University College	London	552	389	163
Liverpool John Moores University	North West	479	425	54
University of Chester	North West	441	339	102
University of Worcester	West Midlands	434	321	113
University of Manchester	North West	423	423	0
Bath Spa University	South West	410	410	0
Leeds Trinity University	Yorkshire and The Humber	410	277	133
Roehampton University	London	407	279	128
Bishop Grosseteste University	East Midlands	381	263	118
Goldsmiths University	London	350	350	0
University of Birmingham	West Midlands	344	344	0
Total top 20		**12,985**	**10,215**	**2,770**

---

[1] DfE (2016b).

Bottom 20 ITT providers in terms of numbers of trainees 2016-17				
Provider name	Region	Total	Postgraduate	Undergraduate
Compton SCITT	London	19	19	0
Cumbria Primary Teacher Training	North West	19	19	0
NELTA (North East London Teaching Alliance)	London	18	18	0
The Solent SCITT	South East	18	18	0
Devon Secondary Teacher Training Group (DSTTG)	South West	17	17	0
Harefield SCITT	South East	17	17	0
Chepping View Primary Academy SCITT	South East	16	16	0
Buile Hill Visual Arts College SCITT	North West	14	14	0
Landau Forte College Derby SCITT	East Midlands	14	14	0
North Lincolnshire SCITT Partnership	Yorkshire and The Humber	14	14	0
HART of Yorkshire	Yorkshire and The Humber	13	13	0
West Berkshire Training Partnership	South East	13	13	0
The John Taylor SCITT	East Midlands	12	12	0
University of Buckingham	London	11	11	0
Royal Academy of Dance	London	10	10	0
Cheshire East SCITT	North West	9	9	0
Services For Education SCITT	West Midlands	9	9	0
Catalyst Teaching School Alliance SCITT	East of England	8	8	0
Merseyside, Manchester and Lancashire (MML) SCITT	North West	7	7	0
The Beauchamp ITT Partnership	East Midlands	7	7	0
Total bottom 20 ITT providers		**265**	**265**	**0**

Middle 20 ITT providers in terms of numbers of trainees 2016-17 (10 either side of the median - 46 trainees)				
Provider name	Region	Total	Postgraduate	Undergraduate
Astra SCITT	South East	50	50	0
Mid Somerset Consortium for Teacher Training	South West	50	50	0
The Kemnal Academies Trust (TKAT)	London	50	50	0
Cabot Learning Federation SCITT	South West	49	49	0
Stockton-on-Tees Teacher Training Partnership	North East	49	49	0
The Deepings SCITT	East Midlands	49	49	0
The Grand Union Training Partnership	East Midlands	47	47	0
The Havering Teacher Training Partnership	London	47	47	0
Haybridge Alliance SCITT	West Midlands	46	46	0
Kingsbridge EIP SCITT	North West	46	46	0
Peninsula Teacher Training Cornwall SCITT	South West	46	46	0
Sutton SCITT	London	46	46	0
The Tommy Flowers SCITT Milton Keynes	South East	45	45	0
Leicester and Leicestershire SCITT	East Midlands	44	44	0
University College Birmingham	West Midlands	44	44	0
Yorkshire and Humber Teacher Training	Yorkshire and The Humber	44	44	0
Bourton Meadow Initial Teacher Training Centre	South East	43	43	0
George Spencer Academy SCITT	East Midlands	43	43	0
SCITTELS	London	42	42	0
London East Teacher Training Alliance	London	41	41	0
Total Mid 20 around the median (46)		**920**	**920**	**0**

UNIVERSITY OF LEEDS

ITT providers educational assessment course content

Page 1: The theoretical, technical and practical aspects of educational assessment delivered in initial teacher training (ITT) courses

As part of a wider study into trust in teachers' ability to deliver reliable assessments in mainstream education in England, this survey looks to provide information on the theoretical, technical and practical aspects of educational assessment delivered in initial teacher training (ITT) courses and in the further professional development of teachers. The survey will be open until the 31st August and should take about 15 minutes to complete.

The term *'educational assessment'* is used to define **all** assessment activities used in the compulsory phase of education in England designed to measure pupil performance for the purpose of informing teaching and learning, accountability measures and general qualifications.

The information produced as a result of the survey will be used to present an analysis of current provision and the formulation of a syllabus and associated resources of what teachers need to know about educational assessment in their initial training. **The syllabus and supporting resources produced as a result of this study will be made freely available to all contributors who provide contact details in the survey.**

This questionnaire is being circulated to a sample of providers of initial teacher training as contained in the Initial Teacher Training (ITT) census: 2016 to 2017 England. (DfE, 2017)

If you have any questions or concerns please feel free to contact me via e-mail at edmgw@leeds.ac.uk or by mobile on 07976 874171 or contact my research supervisor, Dr. Matt Homer at the University of Leeds, e-mail m.s.homer@leeds.ac.uk

Thank you for your generosity and time.

Mick Walker

University of Leeds, February 2018.

Page 2: Data protection

All data produced in the survey will be used for the sole purpose of the research. Your name will not be linked with the research materials and you will not be identified or be identifiable in any report that results from this research. Your participation in this study is entirely voluntary and you are under no obligation to answer all of the questions. You can withdraw from the questionnaire at any point but once the survey has been submitted you are in effect giving consent that the information provided can be used for the sole purpose of the study.

The questionnaire offers the option for a voluntary follow-up interview to probe deeper into aspects contained in the survey. If you do volunteer to take part in a follow-up interview, you can withdraw from the process at any time and withdraw your interview data up to two weeks following the interview.

Page 3: Background information

This section asks for general background information on your institution.

1. **As a provider of ITT, which of the following routes into teaching do you offer?** *Please select as appropriate.*

☐ HEI-led Undergraduate   ☐ HEI-led Postgraduate   ☐ Teach First
☐ School Direct (Fee)   ☐ School Direct (Salaried)   ☐ School Centred ITT (SCITT)

2. **If you are a School-led provider, do you partner with a Higher Education Institute?**

○ Yes
○ No
○ Not applicable

3. **Which phase of education and/or subjects do you cover in your ITT course?** *Please select the phase of education and/or subjects you cover in the list below and insert the approximate number of students currently on the course. If you support subjects that are not on the list, you can add these in question 3.a.*

	Please select as appropriate.	Approximate number of current students.
Primary phase	☐	
Secondary phase -	☐	
• Art & Design	☐	
• Biology	☐	
• Business Studies	☐	
• Chemistry	☐	
• Classics	☐	
• Computing	☐	

	Please select as appropriate.	Approximate number of current students.
• Dance	☐	
• Design & Technology – including Food	☐	
• Drama	☐	
• Economics	☐	
• English	☐	
• Geography	☐	
• History	☐	
• Mathematics	☐	
• Modern Foreign Languages	☐	
• Music	☐	
• Physical Education	☐	
• Physics	☐	
• Psychology	☐	
• Religious Education	☐	
• Science	☐	
• Social Studies	☐	

3.a. Other subjects: please specify and provide the approximate number of current students.

4. **Which of the following qualifications do you offer?**

	Please select as appropriate.	Approximate number of current students.

Bachelor's Degree	☐	
Qualified Teacher Status (QTS)	☐	
Post Graduate Certificate in Education (PGCE)	☐	
Post Graduate Diploma in Education (PGDE)	☐	

4.a. *Other - please specify.*

4.a.i. *Approximate number of current students.*

Page 4: The Carter Review of Initial Teacher Training (ITT)

This section of the survey asks for your views on two conclusions reached by the Carter Review that refer to gaps in the capacity of schools and ITT providers in the theoretical and technical aspects of educational assessment.

5. **The Carter Review of Initial Teacher Training (ITT) concluded;** *'Of all areas of ITT content, we believe the most significant improvements are needed for training in assessment'.* **(Carter, 2015. P9).**

Please don't select more than 1 answer(s) per row.

	Strongly disagree	Disagree	Agree	Strongly agree
How strongly do you agree or disagree with this finding?	□	□	□	□

5.a. *If you wish to add a comment on this finding of the Carter Review, please use the space below.*

6. **Carter also concluded that:** *'Findings from the NAHT Commission (2014) as well as Ofsted have also found weaknesses in assessment training. We believe that there are significant gaps in both the capacity of schools and ITT providers in the theoretical and technical aspects of assessment' (Ibid).*

Please don't select more than 1 answer(s) per row.

	Strongly disagree	Disagree	Agree	Strongly agree
How strongly do you agree or disagree that there are significant gaps in the capacity of schools in the theoretical and technical aspects of assessment?	□	□	□	□

How strongly do you agree or disagree that there are significant gaps in the capacity of ITT providers in the theoretical and technical aspects of assessment?	☐	☐	☐	☐

6.a. *Use this space to add any supporting comments.*

Page 5: The educational assessment content of your ITT course - Page 1 of 6

*Questions 7 to 12 of the survey ask about a range of aspects relating to educational assessment to ascertain which aspects are included and which aspects are not included in your current ITT course. For each of the aspects that **are** covered, you are asked to state your level of confidence that your course fully prepares students for their first years in teaching.*

For each of the aspects that are covered, please state your level of confidence that your course fully prepares students for their first years in teaching where VC = very confident, MC = moderately confident, NVC = not very confident, NAC = not at all confident.

Models of learning.

7. **With reference to the aspects of educational assessment in the table below: a) which aspects are included in your ITT course; b) which aspects are not included in your ITT course?** *For each of the aspects that are covered, please state your level of confidence that your course fully prepares students for their first years in teaching.*

	Aspects included in our course.	Aspects not included in our course.	VC	MC	NVC	NAC
Knowledge and understanding of how pupils learn and how this impacts on teaching.	○	○	○	○	○	○
Stages of development within subjects.	○	○	○	○	○	○
Approaches for strengthening pupil memory, such as short tests and making effective use of questioning.	○	○	○	○	○	○

Page 6: The educational assessment content of your ITT course - Page 2 of 6

Understanding progression.

8. **With reference to the aspects of educational assessment in the table below: a) which aspects are included in your ITT course; b) which aspects are not included in your ITT course?** *For each of the aspects that are covered, please state your level of confidence that your course fully prepares students for their first years in teaching.*

	Aspects included in our course.	Aspects not included in our course.	VC	MC	NVC	NAC
Knowledge and understanding of how pupils are expected to make progress within different subjects across each relevant Key Stages.	○	○	○	○	○	○
The use of relevant data to monitor progress.	○	○	○	○	○	○
The use of relevant data to set targets.	○	○	○	○	○	○
The use of relevant data to plan subsequent lessons.	○	○	○	○	○	○
The use of assessment to give pupils regular feedback, both orally and through accurate marking, and encourage pupils to respond to the feedback.	○	○	○	○	○	○
The use of formative and summative assessment to secure pupils' progress.	○	○	○	○	○	○

Page 7: The educational assessment content of your ITT course - Page 3 of 6

Principles of educational assessment.

9. With reference to the aspects of educational assessment in the table below: a) which aspects are included in your ITT course; b) which aspects are not included in your ITT course? *For each of the aspects that are covered, please state your level of confidence that your course fully prepares students for their first years in teaching.*

	Aspects included in our course.	Aspects not included in our course.	VC	MC	NVC	NAC
Theory of assessment – why, when, how to assess and report.	○	○	○	○	○	○
The fundamental principles of assessment, including: the differences between formative and summative uses of assessment;	○	○	○	○	○	○
• bias;	○	○	○	○	○	○
• validity;	○	○	○	○	○	○
• reliability;	○	○	○	○	○	○
• utility;	○	○	○	○	○	○
• criterion- and norm-referencing;	○	○	○	○	○	○
• setting standards of educational performance;	○	○	○	○	○	○
• standardisation – purpose and process;	○	○	○	○	○	○
• moderation - purpose and process;	○	○	○	○	○	○
• the design of standardised tests (such as those that produce a reading age).	○	○	○	○	○	○
How nationally standardised summative assessment can be used to help teachers to understand national expectations and assess their own performance.	○	○	○	○	○	○
Using externally set standards of educational performance for example Interim Teachers Assessment Frameworks, exemplification of standards from STA or awarding bodies.	○	○	○	○	○	○

Page 8: The educational assessment content of your ITT course - Page 4 of 6

The uses of educational assessment outcomes.

10. **With reference to the aspects of educational assessment in the table below: a) which aspects are included in your ITT course; b) which aspects are not included in your ITT course?** *For each of the aspects that are covered, please state your level of confidence that your course fully prepares students for their first years in teaching.*

	Aspects included in our course.	Aspects not included in our course.	VC	MC	NVC	NAC
The use and misuse of assessment outcomes and the ethics of assessment.	○	○	○	○	○	○
Common inferences/uses of assessment.	○	○	○	○	○	○
Reporting and using assessment outcomes.	○	○	○	○	○	○
Using subject knowledge, progression, assessment and skilful questioning to inform planning.	○	○	○	○	○	○
The effective use of assessment data to ensure that teaching is both supportive and challenging.	○	○	○	○	○	○
The effective use of assessment to give effective and efficient oral and written feedback to pupils and parents.	○	○	○	○	○	○
Using knowledge of how pupils make progress and assessment outcomes for teachers to self-evaluate.	○	○	○	○	○	○

Page 9: The educational assessment content of your ITT course - Page 5 of 6

Assessment design.

11. **With reference to the aspects of educational assessment in the table below: a) which aspects are included in your ITT course; b) which aspects are not included in your ITT course?** *For each of the aspects that are covered, please state your level of confidence that your course fully prepares students for their first years in teaching.*

	Aspects included in our course.	Aspects not included in our course.	VC	MC	NVC	NAC
The relationship between what is taught and what is assessed.	○	○	○	○	○	○
How to identify key constructs and valid approaches to assessment.	○	○	○	○	○	○
Knowledge and understanding of how to assess the relevant subject and curriculum areas, including statutory assessment requirements.	○	○	○	○	○	○
Designing appropriate assessments.	○	○	○	○	○	○
Setting effective questions.	○	○	○	○	○	○
The use of continuous assessment of pupils' work and how approaches to assessment differ across subjects.	○	○	○	○	○	○
Assessment systems – purpose, design, implementation and uses including: school based; local/regionally derived systems; national or large scale based systems.	○	○	○	○	○	○

Page 10: The educational assessment content of your ITT course - Page 6 of 6

Educational assessment and accountability.

12. **With reference to the aspects of educational assessment in the table below: a) which aspects are included in your ITT course; b) which aspects are not included in your ITT course?** *For each of the aspects that are covered, please state your level of confidence that your course fully prepares students for their first years in teaching.*

	Aspects included in our course.	Aspects not included in our course.	VC	MC	NVC	NAC
Knowledge and understanding of the statutory accountability measures.	○	○	○	○	○	○
The requirements of the national curriculum, national Key Stage assessments and/or specifications for public examinations for the subject(s) and phase(s) they will be teaching.	○	○	○	○	○	○
Knowledge and understanding of the National Curriculum assessment system and administrative requirements.	○	○	○	○	○	○
Knowledge and understanding of the assessment system and administrative requirements for GCSE and GCE examination courses.	○	○	○	○	○	○

Page 11: Course development

13. Are there any aspects of educational assessment listed in the previous sections that you think should *NOT* be included in a course of initial teacher training? *Please say why.*

	Please use the space belowto name the aspects.	*Please use the space belowto say why.*
Aspect		
Aspect		
Aspect		
Aspect		
Aspect		

13.a. *If you wish to add further aspects that should NOT be included, please use the space below.*

14. Are there any aspects of educational assessment not listed in the previous sections that you think should be ADDED to those listed earlier? *Please say why.*

	Please us the space belowto name the aspects.	*Please use the space belowto say why.*
Aspect		

Aspect		
Aspect		
Aspect		
Aspect		

14.a. *If you wish to add further aspects, please use the space below.*

15. **Would you be prepared to share with me the part of your curriculum/course specification that provides details of the theory and practice of educational assessment content to assist in achieving the aims of this survey?**

Please select no more than 1 answer(s).

☐ Yes - it is already in the public domain.

☐ Yes - but only if the content is treated confidentially

☐ No

16. **How confident are you that your institution has the right level of expertise to deliver all aspects of educational assessment theory and practice to the same depth and quality as all other aspects of your ITT course?** *Please specify:*

Please don't select more than 1 answer(s) per row.

	Very confident	Moderately confident	Not very confident	Not at all confident
Level of confidence.	☐	☐	☐	☐

16.a. *If you wish to add a supporting comment, please use the space below.*

17. **Are there any constraints that prevent your institution from delivering what you believe would be an ideal course of instruction in educational assessment theory and practice?** *Please specify and provide a suggestion for how the constraint could be alleviated if possible.*

	Please use the space below to list the constraints.	How the constraint could be alleviated.
Constraint		

Constraint		
Constraint		
Constraint		
Constraint		

18. Does your institution or do any members of your institution engage in research on educational assessment theory and practice? *Please specify as appropriate and if possible add supporting information.*

☐ Institution

☐ Individual

18.a. *If you wish to add supporting information, please use the space below.*

Page 13: Further contact permission

19. **Would you be prepared to be contacted for a follow-up interview and/or receive a copy of the outcomes of this research?**

Please select between 1 and 2 answers.

☐ I would be prepared to be contacted for a follow-up interview.

☐ I would like to receive a copy of the research outcomes

☐ No, I would not like any further contact

19.a. *Please add your contact details here if appropriate.*

20. *If you would like to add any further comments or additional information relating to this survey, please use the space below.*

Page 14: Thank you very much for completing this survey!

Appendix 4: Teacher Questionnaire Survey

UNIVERSITY OF LEEDS

ITT educational assessment course content - teacher survey

Page 1: Trust in teachers' assessment.

Introduction and background

Since retiring as Executive Director of Education at the Qualifications and Curriculum Development Agency, I have continued my interest and involvement in educational assessment with a particular focus on how we can increase trust in teachers' assessments. I am now studying for a PhD at the University of Leeds and writing to you now to ask for your help and assistance.

As you are probably aware, teacher assessment has been greatly reduced in the new general qualifications, its place in national curriculum assessments is under review and the variability and quality of initial training in aspects of assessment has been questioned. For example, the Carter Review of Initial Teacher Training (ITT) concluded that significant improvements are needed for training in assessment and that there are significant gaps in schools and ITT providers in the theoretical and technical aspects of assessment (Carter, 2015).

As part of a wider study into trust in teachers' ability to deliver reliable assessments in mainstream education in England, this survey looks to provide information on the theoretical, technical and practical aspects of educational assessment delivered in ITT. The term *'educational assessment'* is used to define **all** assessment activities used in the compulsory phase of education in England designed to measure pupil performance for the purpose of informing teaching and learning, accountability measures and general qualifications.

I would be delighted if you would take a little of your precious time to complete and submit the questionnaire. It should take about 15 to 20 minutes to complete the survey.

The information produced as a result of the survey will be used to present an analysis of current provision and the formulation of a syllabus and associated resources of what teachers need to know about educational assessment in their initial training. **The syllabus and supporting resources produced as a result of this study will be made freely available to all contributors who provide contact details in the survey.**

This questionnaire is open until August 31st and is being circulated to teachers in the primary and

secondary phases of education. Please feel free to pass the survey link to your teaching colleagues.

If you have any questions or concerns please contact me via e-mail at edmgw@leeds.ac.uk or by mobile on 07976 874171 or contact my research supervisor, Dr. Matt Homer at the University of Leeds, e-mail m.s.homer@leeds.ac.uk

Thank you for your generosity and time.

Mick Walker

University of Leeds. February 2018.

Page 2: Data protection

All data produced in the survey will be used for the sole purpose of the research. Your name will not be linked with the research materials and you will not be identified or be identifiable in any report that results from this research. Your participation in this study is entirely voluntary and you are under no obligation to answer all of the questions. You can withdraw from the questionnaire at any point but once the survey has been submitted you are in effect giving consent that the information provided can be used for the sole purpose of the study.

The questionnaire offers the option for a voluntary follow-up interview to probe deeper into aspects contained in the survey. If you do volunteer to take part in a follow-up interview, you can withdraw from the process at any time and withdraw your interview data up to two weeks following the interview.

Page 3: Background information

1. **Which of the following routes into teaching did you take?**

Please select no more than 1 answer(s).

☐ HEI-led Undergraduate   ☐ HEI-led Postgraduate   ☐ Teach First

☐ School Direct (Fee)   ☐ School Direct (Salaried)   ☐ School Centred ITT (SCITT)

1.a. ***Other route - please state.***

2. ***Please state the year in which completed your initial teacher training.***

3. ***Please provide the name of the institution where you completed your initial teacher training.***

4. **In which phase of education and/or subjects were you trained?**

	Please select as appropriate.
Primary phase	☐
Secondary phase -	☐
• Art & Design	☐
• Biology	☐
• Business Studies	☐
• Chemistry	☐

• Classics	☐
• Computing	☐
• Dance	☐
• Design & Technology – including Food	☐
• Drama	☐
• Economics	☐
• English	☐
• Geography	☐
• History	☐
• Mathematics	☐
• Modern Foreign Languages	☐
• Music	☐
• Physical Education	☐
• Physics	☐
• Psychology	☐
• Religious Education	☐
• Science	☐
• Social Studies	☐

4.a. *Other subject: please specify.*

```
┌─────────────────────────────┐
│                             │
│                             │
│                             │
└─────────────────────────────┘
```

5. **Which of the following qualifications do you hold?**

	Please select as appropriate.
Bachelor's Degree	☐
Qualified Teacher Status (QTS)	☐
Post Graduate Certificate in Education (PGCE)	☐
Post Graduate Diploma in Education (PGDE)	☐

| Masters | ☐ |
| PhD | ☐ |

5.a. *If you hold another qualification, please specify.*

Page 4: The Carter Review of Initial Teacher Training (ITT)

This section of the survey aks for your views on two conclusions reached by the Carter Review of Initial Teacher Training that refers to gaps in the capacity of schools and ITT providers in the theoretical and technical aspects of educational assessment.

6. **The Carter Review of Initial Teacher Training (ITT) concluded;** *'Of all areas of ITT content, we believe the most significant improvements are needed for training in assessment'*. **(Carter, 2015. P9).**

Please don't select more than 1 answer(s) per row.

	Strongly disagree	Disagree	Agree	Strongly agree
How strongly do you agree or disagree with this finding?	☐	☐	☐	☐

6.a. *If you wish to add a comment on this finding of the Carter Review, please use the space below.*

7. **Carter also concluded that:** *'Findings from the NAHT Commission (2014) as well as Ofsted have also found weaknesses in assessment training. We believe that there are significant gaps in both the capacity of schools and ITT providers in the theoretical and technical aspects of assessment'* **(Ibid).**

Please don't select more than 1 answer(s) per row.

	Strongly disagree	Disagree	Agree	Strongly agree
How strongly do you agree or disagree that there are significant gaps in the capacity of SCHOOLS in the theoretical and technical aspects of assessment?	☐	☐	☐	☐

How strongly do you agree or disagree that there are significant gaps in the capacity of ITT PROVIDERS in the theoretical and technical aspects of assessment?	☐	☐	☐	☐

7.a. *Use this space to add any supporting comments.*

Page 5: Trust in teachers' assessment.

8. The NAHT Commission on Assessment stated that there is a lack of trust in teacher assessment at the present time. As a teacher, do you think that your assessment judgments are trusted by:

	Yes	No
Your immediate colleagues?	○	○
Colleagues in other schools?	○	○
Ofsted?	○	○
Local Education Authority?	○	○
Examination boards?	○	○
Standards and Testing Agency?	○	○
Pupils?	○	○
Parents?	○	○
Politicians?	○	○
Employers?	○	○
Parents?	○	○

8.a. Please use the space below if you wish to add a comment on trust in teachers' assessments.

Page 6: The educational assessment content of your ITT course.
Page 1 of 6

*Questions 8 to 13 of the survey ask about a range of aspects relating to educational assessment to ascertain which aspects were included and which aspects were not included in your ITT course. For each of the aspects that were covered, you are asked to state your level of confidence that your course **fully prepared you** for your first years in teaching.*

For each of the aspects that were covered, please state your level of confidence that your course fully prepared you for your first years in teaching where VC = very confident, MC = moderately confident, NVC = not very confident, NAC = not at all confident.

Models of learning.

9. **With reference to the aspects of educational assessment in the table below: a) which aspects were included in your ITT course; b) which aspects were not included in your ITT course?** *For each of the aspects that were included, please state your level of confidence that your course fully prepared you for your first years in teaching.*

	Aspects included in my ITT course.	Aspects not included in my ITT course.	VC	MC	NVC	NAC
Knowledge and understanding of how pupils learn and how this impacts on teaching.	○	○	○	○	○	○
Stages of development within subjects.	○	○	○	○	○	○
Approaches for strengthening pupil memory, such as short tests and making effective use of questioning.	○	○	○	○	○	○

Page 7: The educational content of your ITT course. Page 2 of 6.

Understanding progression.

10. **With reference to the aspects of educational assessment in the table below: a) which aspects were included in your ITT course; b) which aspects were not included in your ITT course?** *For each of the aspects that were included, please state your level of confidence that your course fully prepared you for your first years in teaching.*

	Aspects included in my ITT course.	Aspects not included in my ITT course.	VC	MC	NVC	NAC
Knowledge and understanding of how pupils are expected to make progress within different subjects across each relevant Key Stages.	○	○	○	○	○	○
The use of relevant data to monitor progress.	○	○	○	○	○	○
The use of relevant data to set targets.	○	○	○	○	○	○
The use of relevant data to plan subsequent lessons.	○	○	○	○	○	○
The use of assessment to give pupils regular feedback, both orally and through accurate marking, and encourage pupils to respond to the feedback.	○	○	○	○	○	○
The use of formative and summative assessment to secure pupils' progress.	○	○	○	○	○	○

Page 8: The educational content of your ITT course. Page 3 of 6

Principles of educational assessment.

11. **With reference to the aspects of educational assessment in the table below: a) which aspects were included in your ITT course; b) which aspects were not included in your ITT course?** *For each of the aspects that were included, please state your level of confidence that your course fully prepared you for your first years in teaching.*

	Aspects included in my ITT course.	Aspects not included in my ITT course.	VC	MC	NVC	NAC
Theory of assessment – why, when, how to assess and report.	○	○	○	○	○	○
The fundamental principles of assessment, including: the differences between formative and summative uses of assessment;	○	○	○	○	○	○
• bias;	○	○	○	○	○	○
• validity;	○	○	○	○	○	○
• reliability;	○	○	○	○	○	○
• utility;	○	○	○	○	○	○
• criterion- and norm-referencing;	○	○	○	○	○	○
• setting standards of educational performance;	○	○	○	○	○	○
• standardisation – purpose and process;	○	○	○	○	○	○
• moderation - purpose and process;	○	○	○	○	○	○
• the design of standardised tests (such as those that produce a reading age).	○	○	○	○	○	○
How nationally standardised summative assessment can be used to help teachers to understand national expectations and assess their own performance.	○	○	○	○	○	○
Using externally set standards of educational performance for example Interim Teachers Assessment Frameworks, exemplification of standards from STA or awarding bodies.	○	○	○	○	○	○

Page 9: The educational content of your ITT course. Page 4 of 6

The uses of educational assessment outcomes.

12. **With reference to the aspects of educational assessment in the table below: a) which aspects were included in your ITT course; b) which aspects were not included in your ITT course?** *For each of the aspects that were included, please state your level of confidence that your course fully prepared you for your first years in teaching.*

	Aspects included in my ITT course.	Aspects not included in my ITT course.	VC	MC	NVC	NAC
The use and misuse of assessment outcomes and the ethics of assessment.	○	○	○	○	○	○
Common inferences/uses of assessment.	○	○	○	○	○	○
Reporting and using assessment outcomes.	○	○	○	○	○	○
Using subject knowledge, progression, assessment and skilful questioning to inform planning.	○	○	○	○	○	○
The effective use of assessment data to ensure that teaching is both supportive and challenging.	○	○	○	○	○	○
The effective use of assessment to give effective and efficient oral and written feedback to pupils and parents.	○	○	○	○	○	○
Using knowledge of how pupils make progress and assessment outcomes for teachers to self-evaluate.	○	○	○	○	○	○

Page 10: The educational assessment content of your ITT course. Page 5 of 6

Assessment design.

13. **With reference to the aspects of educational assessment in the table below: a) which aspects were included in your ITT course; b) which aspects were not included in your ITT course?** *For each of the aspects that were included, please state your level of confidence that your course fully prepared you for your first years in teaching.*

	Aspects included in my ITT course.	Aspects not included in my ITT course.	VC	MC	NVC	NAC
The relationship between what is taught and what is assessed.	○	○	○	○	○	○
How to identify key constructs and valid approaches to assessment.	○	○	○	○	○	○
Knowledge and understanding of how to assess the relevant subject and curriculum areas, including statutory assessment requirements.	○	○	○	○	○	○
Designing appropriate assessments.	○	○	○	○	○	○
Setting effective questions.	○	○	○	○	○	○
The use of continuous assessment of pupils' work and how approaches to assessment differ across subjects.	○	○	○	○	○	○
Assessment systems – purpose, design, implementation and uses including: school based; local/regionally derived systems; national or large scale based systems.	○	○	○	○	○	○

Page 11: The educational content of your ITT course. Page 6 of 6

Educational assessment and accountability.

14. **With reference to the aspects of educational assessment in the table below: a) which aspects were included in your ITT course; b) which aspects were not included in your ITT course?** *For each of the aspects that were included, please state your level of confidence that your course fully prepared you for your first years in teaching.*

	Aspects included in my ITT course.	Aspects not included in my ITT course.	VC	MC	NVC	NAC
Knowledge and understanding of the statutory accountability measures.	○	○	○	○	○	○
The requirements of the national curriculum, national Key Stage assessments and/or specifications for public examinations for the subject(s) and phase(s) they will be teaching.	○	○	○	○	○	○
Knowledge and understanding of the National Curriculum assessment system and administrative requirements.	○	○	○	○	○	○
Knowledge and understanding of the assessment system and administrative requirements for GCSE and GCE examination courses.	○	○	○	○	○	○

Page 12: Reflections on your ITT course.

15. **Based on your own experience as a practising teacher, do you think any of the aspects identified in the previous sections are NOT necessary for a course in initial teacher training?** *Please say why.*

	Please use the space below to name the aspects.	*Please use the space below to say why.*
Aspect		
Aspect		
Aspect		
Aspect		
Aspect		

15.a. *If you wish to add further aspects that should NOT be included, please use the space below.*

16. **Would you have liked to cover other aspects of educational assessment theory and practice that were not included in your ITT course?** *Please specify and say why.*

	Please us the space below to specify the aspects.	*Please use the space below to say why.*

Aspect		
Aspect		
Aspect		
Aspect		
Aspect		

16.a. *If you wish to add further aspects, please use the space below.*

Page 13: Educational assessment course delivery.

17. **Based on the experience of your initial teacher training, how confident are you that your ITT provider had the right level of expertise to deliver all aspects of educational assessment theory and practice to the same depth and quality as all other aspects of your ITT course?**

Please don't select more than 1 answer(s) per row.

	Very confident	Moderately confident	Not very confident	Not at all confident
Level of confidence.	☐	☐	☐	☐

17.a. **If you wish to add a supporting comment, use the space below.**

18. **If you had access to a curriculum, course specification or other documents that provided the content of your ITT course, would you be prepared to share a copy with me to assist in achieving the aims of this research?**

Please select no more than 1 answer(s).

☐ Yes - it is already in the public domain.

☐ Yes - but only if the content is treated confidentially

☐ No

19. **Since your initial teacher training have you engaged in research on educational assessment theory and practice?** *Please specify as appropriate and if possible add supporting information.*

☐ Yes

☐ No

19.a. *If you wish to add supporting information, please use the space below.*

Page 14: Further contact permission.

20. **Would you be prepared to be contacted for a follow-up interview and/or receive a copy of the outcomes of this research?**

Please select between 1 and 2 answers.

☐ I would be prepared to be contacted for a follow-up interview.

☐ I would like to receive a copy of the research outcomes

☐ No, I would not like any further contact

20.a. *If you are prepared to be contacted for a follow-up interview or you would like a copy of the research outcomes, please add your contact details here.*

21. *If you would like to add any further comments or additional information relating to this survey, please use the space below.*

Page 15: Thank you very much for completing the survey!

Appendix 5: Ethical approval

UNIVERSITY OF LEEDS

Michael Gordon Walker
School of Education
University of Leeds
Leeds, LS2 9JT

ESSL, Environment and LUBS (AREA) Faculty Research Ethics Committee
University of Leeds

8 November 2017

Dear Michael

Title of study: Can we increase trust in teachers' assessments?
Ethics reference: LTEDUC-091

I am pleased to inform you that the above application for light touch ethical review has been reviewed by a School Ethics Representative of the ESSL, Environment and LUBS (AREA) Faculty Research Ethics Committee. I can confirm a favourable ethical opinion on the basis of the application form as of the date of this letter. The following documentation was considered:

Document	Version	Date
LTEDUC-091 Signed Light touch ethical approval-2.docx	1	06/06/17

The reviewer made the following comments about your application:
- The questionnaire is very straightforward and all ethical precautions have been carefully considered.
- There is just one omission – the completion of Section A4 by the supervisor by ticking the two boxes to approve of the research and the student's capability to carry it out – a formality that appears to have been overlooked.

Please notify the committee if you intend to make any amendments to the original research as submitted at date of this approval, including changes to recruitment methodology. All changes must receive ethical approval prior to implementation. The amendment form is available at http://ris.leeds.ac.uk/EthicsAmendment.

Please note: You are expected to keep a record of all your approved documentation, as well as documents such as sample consent forms, and other documents relating to the study. This should be kept in your study file, which should be readily available for audit purposes. You will be given a two week notice period if your project is to be audited. There is a checklist listing examples of documents to be kept which is available at http://ris.leeds.ac.uk/EthicsAudits.

We welcome feedback on your experience of the ethical review process and suggestions for improvement. Please email any comments to ResearchEthics@leeds.ac.uk.

Yours sincerely

Jennifer Blaikie
Senior Research Ethics Administrator, the Secretariat
On behalf of Dr Kahryn Hughes, Chair, AREA Faculty Research Ethics Committee
CC: Student's supervisor(s)

Appendix 6: Teacher Interview – introduction, themes and question prompts

Interview Schedule - teachers

Interviews with teachers

General introduction

The aim of the interview(s) is to gather factual information about your experience, feelings and actions. It seeks to understand your experience and views on educational assessment.

There is no judgment about what is right or wrong in the interview.

Please describe as a precisely as possible; your experience, feelings and actions.

There are three interviews in the series with time between each one to reflect on the experience and gather thoughts on the next interview. However, there is no need to prepare for the interviews.

Interview 1

Setting the context: establishing the background, that is their experiences up to the present time; this stage should provide an insight into their initial teacher education phase.

Interview 2

Establishing the details of their present lived experience, that is to focus on how educational assessment impacts and influences their current practice in order to reconstruct their day-to-day experiences of educational assessment be that through external expectations or the use of assessment in their professional practice.

Interview 3

A focus on establishing perceptions of the function of educational assessment and how they see future possibilities in terms of their role in high stakes assessment and the use of assessment in what might be described as 'low stakes' assessment.

Survey link

https://leeds.onlinesurveys.ac.uk/itt-educational-assessment-course-content-teacher-survey-6

The initial Lead Questions are based on themes generated to support the study:

Whither teachers Assessment: an evaluation of trust in teachers' assessment since the 1944 Education Act.

Follow-up Lead Questions will respond to the initial answers and areas of interest. Before each interview, the themes will be highlighted to inform you of the intended structure. This may of course change depending on your replies.

Interview 1

Themes

1T1. Schooling background and recollections of educational assessment

1T1.1 Early years and the primary phase of education

Lead Question: Can you describe your early memories of assessment when you were at school as a pupil?

> Own school experience as a pupil – what part did assessment play – including on-going classroom based, end of term exams, qualifications and tests?
>
> What did assessment outcomes actually mean to you as a pupil?
>
> Was assessment noticeable?
> Was it a source of pressure or anxiety?
> A joy?
> A way to assist learning?
> Did it create competition with others?
> Did your parents have any views or input?
>
> Any examples of the above in practice?

1T1.2 Student – Secondary, high school phase of education

Lead Question: Can you describe your early memories of assessment when you were a student at high school?

> Own school experience as a student – what part did assessment play – including on-going classroom based, end of term exams, qualifications and tests?
>
> What did assessment outcomes actually mean to you as a pupil?
>
> Was assessment noticeable?
> Was it a source of pressure or anxiety?
> A joy?
> A way to assist learning?

Did it create competition with others?
Did your parents have any views or input?

Any examples of the above in practice?

1T1.3 Student, university phase of education

Lead Question: Can you describe your early memories of assessment when you were a student at university?

Own experience as a student – what part did assessment play – including on-going classroom based, end of term exams, qualifications and tests?

What did assessment outcomes actually mean to you as a student?

Was assessment noticeable?
Was it a source of pressure or anxiety?
A joy?
A way to assist learning?
Did it create competition with others?

Any examples of the above in practice?

1T2. Any other professional experience

1T2.1 Work outside of education

Lead Question: Do you have experience of assessment working outside of teaching?

If other professions or jobs were experienced, did assessment play a part?

1T3.Route into teaching

Lead Question: How and why did you get into teaching?

What route into teaching, did you have a choice?
Why did you select this route?
Why did you select this provider?

1T4. Initial Teacher Education/Training stage

1T4.1 Education/Training course format

Lead Question: Can you describe the structure and content of your initial teacher training?

> What did your training cover?
> How much was university v classroom based?
> What did you experience re educational assessment education/training?

1T4.2 Point of transfer into teaching

Lead Question: When you entered the profession in your first job, with reference to educational assessment, did you feel prepared?

> Plus
> Lead Questions specific to completed Lead Questionnaire

1T5. Perceptions of educational assessment over time

Lead Question: Have your perceptions or experience of educational assessment changed since you began teaching?

> If your view has changed, what has caused the change?

Interview 2

Themes

2T1. Early years of teaching

2T1.1 Level of preparation

Lead Question: Looking back at when you started teaching, were you adequately prepared in terms of assessment knowledge and understanding?

Where did you gain your knowledge of assessment practice?
Have you had any CPD or other support focused on assessment since you started teaching? Give examples.
Was it of sufficient or good quality?

2T2. Defining educational assessment

Lead Question: How would you define or describe educational assessment?

Prompts:
What's its purpose or purposes?
What part does assessment play in your daily life as a teacher?
Do you and other teachers have sufficient knowledge and understanding of educational assessment?
What is the impact of the accountability system on assessment practice in school?

On teachers
On managers
On pupils
On parents
On governors

2T3. Assessment in the English education system

2T3.1 System wide use of educational assessment

Lead Question: *What are your views on the way educational assessment is used in the English education system?*

> Are you happy with the way assessment is used in school?
> What do you know or think about assessment in other schools?

2T3.2 Potential changes to the system.

Lead Question: *How would you change the current system?*

> What are the benefits or downside of the current assessment system?
> Does assessment practice vary across the school?

2T4. Trust in teachers' assessment

2T4.1 Perceived level of trust

Lead Question: *Can we trust teachers' assessment?*

> In general terms – across the system
> Within this school
> Local schools
> In tests and examinations
> In coursework and controlled assessments

2T4.2 The role of external agencies

Lead Question: *What's your view on the role of external agencies?*

> STA/Exam boards/Ofsted/LEA/MAT/DfE/ Politicians in respect of their
> influence on assessment practice in the school?
>
> How would you compare assessment in this school with the
> expectations of external agencies?
>
> Do you trust external agencies – explain.

2T5. Changes to educational assessment practice

Lead Question: *How has educational assessment practice changed over your lifetime?*

What's the state of assessment in schools in England at the current time and how has it changed during your career?
What are the changes?
Are they for the good?

Has teachers' assessment changed over the time you have been teaching?

If so, in what ways?
In coursework or controlled assessments?
Teacher assessments in national curriculum assessments?
In day to day classroom practice?

What has been the impact of any change?

On teachers?
On pupils?
Public impact?
Politicians?
Performance tables?

Interview 3

Themes

3T1. The function of educational assessment

3T1.1 Use(s) of educational assessment in the English education system?

Lead Question: *How do you think assessment should be used in the English education system?*

> At national level?
> At school level?
> At classroom level?

3T2. Teachers' assessment and the accountability system

3T2.1 The place of teachers' assessment in the accountability system

Lead Question: *Should teachers' assessment form a part of the accountability system?*

> If it should, how should this be done?
> How can its quality be assured?
> What support would teachers need?
> Can we trust teachers' assessment? In this school/ other local schools/nationally
> Why/why not?
> Has impacted on the use of teachers' assessment changed over time?

3T3. School accountability

3T3.1 General impact

Lead Question: *What is the impact of the school accountability measures?*

3T3.2 Impact on schools

Lead Question: *How does school accountability impact on teachers' practice?*

How should schools' performance be measured and what part would assessment outcomes play?

Should government have any part to play in schools assessment?

3T3.3 Impact on learners

Lead Question: What is the impact of educational assessment on learners?

3T3.4 Workload implications

Lead Question: What is the impact of educational assessment on workload?

3T5. Expected future of assessment in England

3T5.1 Possible change to the system

Lead Question: Do you see the use of assessment changing over the next five years?

3T5.2 Government approach

Lead Question: Are current and planned approaches for the use of assessment by government on the right lines?

3T5.3 The future of teachers' assessment.

Lead Question: What's the future for teachers' assessment?

3T5.4 Professionalisation of assessment

Lead Question: Should external assessment, including marking, be carried out by professionally qualified persons? For example Chartered Assessors?

3T6. Knowledge and expertise

3T6.1 Teachers' knowledge of educational assessment

Lead Question: Do teachers have sufficient knowledge and understanding of educational assessment?

3T6.2 Developing expertise and knowledge

Lead Question: How do you, as a practicing teacher, develop your assessment expertise?

Do you have access to CPD?
Who provides training – internal or external?

Do those who educate/train teachers have sufficient expertise to provide the right quality of instruction?

3T7. What do teachers really need to know about assessment when they enter the profession and over the time of their career development?

3T7.1 On entry into the profession

Lead Question: Based on your own experience, do you have any views about what teachers really need to know about educational assessment when they enter the profession?

3T7.2 Early career development

Lead Question: What support or development do teachers need in the early stages of teaching?

3T7.3 On-going professional development

Lead Question: What support or development do teachers need as part of on-going CPD?

Appendix 7: Key Influencer Interview – themes and question prompts

Interview Schedule: Interviews with key influencers

Interviews with assessment specialists (academics), politicians, regulatory body officials, professional body representatives, government officials (local and national).

To note, the focus is on assessment in English schools only.

Themes

T.1 Background/expertise/interest in education – particularly around educational assessment.

T.2 Describing, defining educational assessment

Lead question: How would you describe educational assessment?

What's its purpose or purposes?

T.3 Assessment and accountability

Lead question: What's the general state of educational assessment in English schools at the current time?

In terms of the general state of educational assessment in schools in England *at the current time*, what's your view on how assessment is used as a tool for accountability?

Should assessment be used for accountability purposes?

- Should high stakes external assessment, including marking, be carried out by professionally qualified persons? For example full time chief examiners or Chartered Assessors?

- Should teachers' assessments form a part of the accountability system?

How should our schools be judged for accountability purposes?

What role should the government take?

T.4 Trust in teachers' assessments

Lead question: Can we trust teachers' assessments in high stake tests, coursework etc.?

- Why/why not?
- What's the evidence base?

T.5 Teachers' assessments in more general terms

Lead question: *In terms of the general state of assessment in schools in England at the current time, what's your view on how assessment is used as a tool for teaching and learning?*

What's the evidence base?

How has the use of teachers' assessment changed over time?

T.6 Teachers/ITT knowledge base

Lead question: *Do teachers have sufficient knowledge and understanding of educational assessment:*

- *at the end of their period of initial teacher training?*
- *as they progress through their careers?*

Do those who educate teachers have sufficient expertise to provide the right quality of instruction?

What's the evidence base?

T.7 How *should* assessment be used in the English education system?

Lead question: *What do teachers **really** need to know about assessment when they enter the profession?*
- *within the early years of their teaching careers; and*
- *for continuous professional development over the time of their careers?*

T.8 Whither teachers' assessment?

Lead question: *What's the future for teachers' assessment in the English education system?*

Does that differ from what you would like to see as an ideal?

Appendix 8: Contemporaneous interview notes example

(Name)   Interview 3.   30 April [?]

Assessment = to inform teachers planning. Identify gaps in pupils' knowledge

External assessment — schools have to be accountable. "You can't live in your own bubble."

2.00 A single test is not appropriate. They are too young — don't have test skills. It doesn't give a true picture

No problem with having a test — but it should be used to inform teacher judgement.

Tests have a long term impact on pupils which is not appreciated. It lays down expectations for GCSE.

This puts pressure on the high school & the child and does not recognise the range of possible change between age 11 and 16.

The assessment and benchmarking and expectations of progress take no account of real life. Children['s] circumstances change.

This isn't against having high expectations, but understanding real life.

5.30 Children become a number or a commodity (Rate — productivity, bureaucracy of children)

Removing TA in the public — done under the guise of workload — 'is a shame'. Teacher would have preferred TA & get rid of the test.

22.00

23.00 The entire summative of primary school comes down to five tests over four years.

They are 10 years old. Some of them

Tests & GCSEs, put a child down to be a number — 99, 100, 8/7 at GCSE.

Role Objectivity.

Not impressed by Ofsted. — Looking at data and forming pre-conceived ideas. Why have you got this poor data? Not looking wider.

Role Doesn't see computer as a better model of accountability because it's data driven.

Use item analysis and example of questions to illustrate to staff what children struggled. But these children have gone. But it could be an area of weakness within the school.

Role. From my experience this is rare?

Table tests shouldn't be necessary; teachers should know that. There is no need to publish all these things

Appendix 9: Coding grid example

Topics	Nodes	Sub-nodes	Example quote	Reference
1T1. Schooling background and recollections of educational assessment **1T1.1 Early years and the primary phase of education** *Lead Question: Can you describe your early memories of assessment when you were at school as a pupil?*	Setting	Moving tables	it was classes of the ongoing assessment, you'd have to get a certain amount right on the particular page you were doing before you could move on to the next page, and if you didn't get them all right, you then went on to a different table, and you had to stay on that table until you did understand it, then you did the page again,	NW1 P1
			I don't remember sitting in different table groups either, we were just sat mixed around, there didn't seem to be a higher or a lower table, and I remember in Year 6 we were in rows, I'm sure we were in rows, so there can't have been separate tables for ability.	ID1 P4
		Awareness of setting	it was only the odd time where I had to move tables, but you were aware of the children that were always on that table, who didn't move to the next book, so you were aware that there were some children that were not doing very well. I never remember the teachers as being mean about it, it's just that's how it was, you went to the teacher's desk, you got them marked, and then they told you whether to go on to the next page or whether to go and sit with the classroom support.	NW1 P2
		Moving class/set Secondary	...yes, secondary school differed, because then you were getting your As and your Bs, or your percentages of your tests and things, and setting as well, we setted at my high school, we setted from straight away basically, for English, Maths and Languages, we setted for. So I was always in the top, so I was in the top set for English, I was in the second set for Maths, and we did have quite a lot of test practices and things, and then you were	NW1 P5

Appendix 10: Salendine Nook High School Mode 3 GCSE
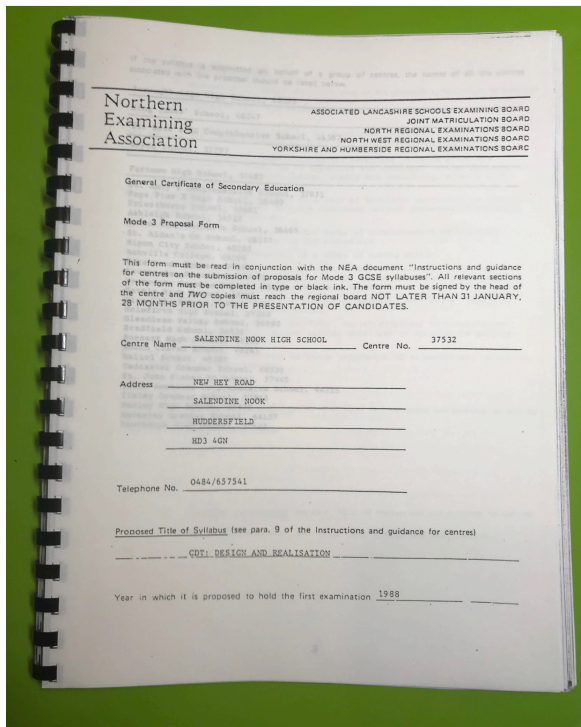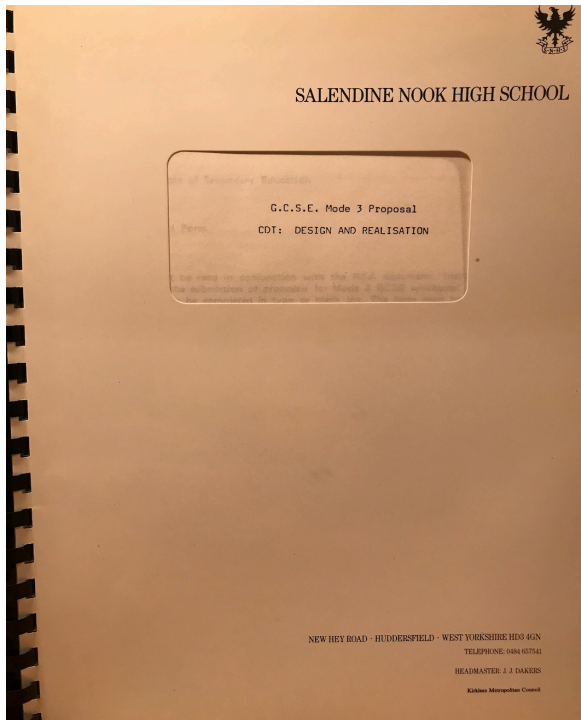
Appendix 11: Statistical test data

a) A summary of the percentage of respondents who answered that each aspect of educational assessment had been included in their ITT course by survey: teachers versus ITT providers' views (see Chapters 5 and 6 for details of the data analysis)

Aspect of educational assessment	ITT providers survey response	Teacher survey responses	Statistical Test. All of these differences in percentage responses between groups are statistically significant at the 5% level (using a chi-square test of [1] association).
The fundamental principles of assessment	100% (n=47)	73.9% (n=199)	Difference: 26.1% 95% CI: 16.6904% to 32.6134% Chi-squared: 15.487 DF; 1 Significance level: P = 0.0001
The effective use of assessment data to ensure that teaching is both supportive and challenging.	100% (n=49)	51% (n=202)	Difference: 49% 95% CI: 39.0385% to 55.8478% Chi-squared: 39.485 DF; 1 Significance level: P = <0.0001
Bias	68.1% (n=47)	17.8% (n=197)	Difference: 50.3% 95% CI: 34.8589% to 62.7428% Chi-squared: DF; 1 Significance level: P = <0.0001
Validity	72.3% (n=47)	24.2% (n=198)	Difference: 48.1% 95% CI: 32.6063 to 60.1303% Chi-squared: 39.329 DF; 1 Significance level: P = <0.0001
Reliability	70.2% (n=47)	26.8% (n=198)	Difference: 43.4% 95% CI: 27.7629% to 55.9052% Chi-squared: 31.268 DF; 1 Significance level: P = <0.0001
Setting standards of educational performance	79.2% (n=48)	26.3% (n=198)	Difference: 52.9% 95% CI: 37.9687% to 63.6012% Chi-squared: 39.485 DF; 1 Significance level: P = <0.0001

[1] Statistical Test using Medcalc. MedCalc uses the "N-1" Chi-squared test as recommended by Campbell (2007) and Richardson (2011).
The confidence interval is calculated according to the recommended method given by Altman et al. (2000)
https://www.medcalc.org/calc/comparison_of_proportions.php

b) A summary of the varying levels of confidence in the preparedness of newly qualified teachers: teachers versus ITT providers' views (see Chapter 5 and Chapter 6 for details of the data analysis for details)

	% of teachers' levels of confidence that their ITT had fully prepared them for entering the profession.		% of ITT providers levels of confidence that their courses prepare newly qualified teachers.		Statistical Test. All of these differences in percentage responses between groups are statistically significant at the 5% level (using a chi-square test of x^2 association).	
Key assessment aspects	Not at all or not very confident	Moderately or Very confident	Not at all or Not very confident	Moderately or Very confident	Comparison between teachers and ITT providers stating; 'Not at all or not very confident'.	Comparison between teachers and ITT providers stating; 'Moderately or very confident' level.
Criterion- and norm-referencing	70.5	27.5 (n=120)	42.4	57.6 (n=33)	Difference; 28.1% 95% CI; 9.2109% to 45.0015% Chi-squared; 8.860 DF; 1 Significance level; P= 0.0029	Difference; 30.1% 95% CI; 11.2567% to 46.9043% Chi-squared; 10.383 DF; 1 Significance level; P= 0.0013
Utility	76.6	23.3 (n=120)	41.9	58.3 (n=36)	Difference; 34.7% 95% CI; 16.5549% to 50.7175% Chi-squared; 15.378 DF; 1 Significance level; P = 0.0001	Difference; 33.0% 95% CI; 14.9473% to 49.2367% Chi-squared; 14.029 DF; Significance level; P= 0.0002
The design of standardised tests	72.8	27.2 (n=125)	41.7	58.3 (n=36)	Difference; 31.1% 95% CI; 12.9155% to 47.2438% Chi-squared; 11.946 DF; 1 Significance level; P = 0.0005	Difference; 31.1% 95% CI; 12.9155% to 47.2438% Chi-squared; 11.946 DF; 1 Significance level; P = 0.0005
Validity	68.9	31.1 (n=122)	36.6	63.4 (n=41)	Difference; 32.3% 95% CI; 14.7122% to 47.3222% Chi-squared; 13.348 DF; 1 Significance level; P = 0/0003	Difference; 32.3% 95% CI; 14.7122% to 47.3222% Chi-squared; 13.348 DF; 1 Significance level; P = 0.0003
Bias	74.2	25.8 (n=120)	37.5	62.5 (n=40)	Difference; 36.7% 95% CI; 19.0541% to 51.7055% Chi-squared; 17.656 DF; 1 Significance level; P = < 0.0001	Difference; 36.7% 95% CI; 19.0541 to 51.7055% Chi-squared; 17.656 DF; 1 Significance level; P = < 0.0001
Reliability	67.5	32.6 (n=123)	35.0	65.0 (n=40)	Difference; 32.5% 95% CI; 14.7313% to 47.4620% Chi-squared; 13.152 DF; 1	Difference; 32.4% 95% CI; 14.6297% to 47.3668% Chi-squared; 13.063 DF; 1

---

[2] Statistical Test using Medcalc. MedCalc uses the "N-1" Chi-squared test as recommended by Campbell (2007) and Richardson (2011). The confidence interval is calculated according to the recommended method given by Altman et al. (2000) https://www.medcalc.org/calc/comparison_of_proportions.php