Determination of diagnostic competence in histopathology trainees

*Dr. Daniel John Brierley*

A thesis submitted in partial fulfilment for the requirements for the degree of Doctor of Philosophy

**University of Sheffield**

The School of Clinical Dentistry

Academic Unit of Oral Medicine, Pathology and Surgery

Supervisors: Professor. P.M. Farthing and Dr. S. Zijlstra-Shaw

Submitted: August 2019

**Acknowledgements**

Firstly, I would like to thank my supervisors, Professor Paula Farthing and Dr. Sandra Zijlstra-Shaw. You provided unwavering guidance during this journey and encouraged me to pursue something that interested me. You taught me to persevere and strive for excellence. I have learnt so much from both of you.

I would also like to thank my colleagues in Oral Pathology for giving me time to do my research and allowing me to strike a balance between my NHS and research commitments.

I am very grateful for the Royal College of Pathologists for disseminating my questionnaire and Delphi invitations.

Thank you to all of those who contributed to the various elements of the PhD-your time, expertise, honesty and support helped to inform this study and make my research possible.

I would like to thank my wife, Catherine, for her love, encouragement and belief in me. Thank you also to my son, William, for putting life into perspective and giving me something to strive for.

|  |  |
| --- | --- |
| **ACGMEA** | Accreditation Council for Graduate Medical Education in America |
| **ADEE** | Association for Dental Education in Europe |
| **AfL** | Assessment for Learning |
| **AoMRC** | Academy of Medical Royal Colleges |
| **AOP** | Assessment of Performance |
| **ARCP** | Annual Review of Competency Progression |
| **CBD** | Case Based Discussion |
| **CCST** | Certificate of Completion of Specialist Training |
| **CWS** | Cochran-Weiss-Shanteau |
| **DEPS** | Dental Evaluation of Performance |
| **DOPS** | Direct Observation of Practical Skills |
| **ECE** | Evaluation of Clinical Events |
| **EPA** | Entrustable Professional Activities |
| **ES** | Educational Supervisor |
| **ESR** | Educational Supervisor’s Report |
| **EWTD** | European Working Time Directive |
| **GDC** | General Dental Council |
| **GMC** | General Medical Council |
| **IQR** | Interquartile Range |
| **ISCP** | Intercollegiate Surgical Curriculum Programme |
| **LEPT** | Learning Environment for Pathology Trainees |
| **MCQ** | Multiple Choice Question |
| **MDT** | Multi-disciplinary Team Meeting |
| **Mini-CEX** | Clinical Evaluation Exercise |
| **MSF** | Multi-source Feedback |
| **NAS** | Next Accreditation System |
| **NGT** | Nominal Group Technique |
| **NHS** | National Health Service |
| **OMFP** | Oral and Maxillofacial Pathology |
| **OSCE** | Objective Structured Clinical Examination |
| **OSPE** | Objective Structured Practical Examination |
| **PAQ** | Patient’s Assessment Questionnaire |
| **PhD** | Doctor of Philosophy |
| **PMETB** | Post-graduate Medical Education and Training Board |
| **RCPath** | Royal College of Pathologists |
| **Research team** | Daniel Brierley, Paula Farthing and Sandra Ziljlstra-Shaw |
| **ROC** | Receiver-Operating Curve |
| **SD** | Standard Deviation |
| **SHO** | Senior House Officer |
| **SJT** | Social Judgement Theory |
| **SLE** | Supervised Learning Event |
| **SLT** | Situated Learning Theory |
| **TPD** | Training Programme Director |
| **UCLA** | University of California Los Angeles |
| **UK** | United Kingdom |
| **WBA** | Workplace-based Assessment |

**List of Abbreviations**

**List of Tables**

Table 1.1. The different types of WBA used in histopathology training

Table 3.1 Questionnaire

Table 3.2. Summary of responses to questionnaire (excluding question 7)

Table 3.3. Summary of responses to question 7

Table 4.1. Summary of participants who took part in the study

Table 5.1. List of 27 items under their associated headings and alongside rating scale

Table 5.2. Summary of Delphi panel participants

Table 5.3. Summary of results from round 1 of Delphi

Table 5.4. Summary of which items reached “consensus in” (green) or no consensus (red) after round 1

Table 5.5. Summary of which items did not reach consensus after round 1

Table 5.6. Summary of results from round 2 of Delphi

Table 5.7. Summary of which items did reach “consensus in” (green) or “no consensus” (red) after round 2.

Table 5.8. Summary of which items did not reach consensus after round 2

Table 5.9. Wilcoxon matched pairs signed rank test comparing median scores for each item between round 1 and round 2.

Table 5.10. Comparison of round 1 and round 2 results

Table 5.11. List of items which moved from “no consensus” to “consensus in” after round 2

Table 5.12. Items which had the highest mean score or were ranked as being most important in round 1

Table 5.13. Items which had the lowest mean score or were ranked as being least important in round 1

Table 5.14. Items which had the highest mean score or were ranked as being most important in round 2

Table 5.15. Items which had the lowest mean score or were ranked as being least important in round 2

Table 5.16. Comparison of highest ranked items in rounds 1 and 2

Table 5.17. Comparison of lowest ranked items in rounds 1 and 2

Table 5.18. Comparison of highest rated items by mean score in rounds 1 and 2

Table 5.19. Comparison of lowest rated items by mean score in rounds 1 and 2

**List of Figures**

Figure 1.1. Miller’s pyramid

Figure 2.1. Brunswik’s lens model

Figure 3.1. Formula to calculate number of respondents required for a given level of reliability

Figure 3.2. Trainee and assessor responses to the statement: WBAs are useful in training

Figure 3.3. Trainee and assessor responses to the statement: The number of WBAs per stage is appropriate

Figure 3.4. Trainee and assessor responses to the statement: There is sufficient time to complete WBAs

Figure 3.5. Trainee and assessor responses to the statement: WBAs are able to identify diagnostic competence in histopathology

Figure 3.6. Trainee and assessor responses to the statement: WBAs are able to identify trainees in difficulty

Figure 3.7. Trainee and assessor responses to the statement: WBAs accurately reflect trainee progress

Figure 3.8. Trainee and assessor responses to the statement: The current WBAs are acceptable to you as a trainee/assessor

Figure 3.9. Trainee response to the statement: Your assessors are accessible to complete WBAs

Figure 3.10 Assessor response to the statement: You have been given sufficient guidance on how to complete WBAs

Figure 3.11. Trainee response to the statement: Your assessors are engaged and helpful when completing WBAs

Figure 3.12. Assessor response to the statement: Trainees select appropriate cases for WBAs

Figure 3.13. Trainee and assessor responses to the question: WBAs are part of the curriculum to: 1. Help monitor trainee’s development/progress, 2. Provide structured feedback to trainees, 3. Satisfy the requirements of the RCPath, 4. Provide evidence for the ARCP

Figure 3.14. Diagram illustrating the main areas of concern regarding the use of WBAs in histopathology generated from free text comments.

Figure 4.1. Categories and sub-headings after six interviews

Figure 4.2. Categories and sub-headings after discussion and reflection

Figure 4.3. The emergence of a framework

Figure 4.4. Conceptual framework to explain how consultants judge diagnostic competence in trainees.

Figure 6.1. Conceptual model of diagnostic competence

**Abstract**

**Background**

To be competent at forming diagnoses is an essential component of being a sound consultant pathologist as these diagnoses can directly affect patient care. Currently, trainees in histopathology undertake five years of specialist training before becoming consultants. During this time, they must complete eighteen workplace-based assessments (WBAs) per year and pass two major exams. However, there is no research which clearly describes the knowledge, skills and attitudes which are relevant to the competence of forming diagnoses. Further, there is little evidence to suggest WBAs are able to determine if trainees are diagnostically competent, even though they are used in part to evidence this. Understanding how consultants determine diagnostic competence in trainees may help to ensure the current assessment methods measure what is most important and ensure trainees are fit to practice at the end of training. It may also reveal the judgmental process so that expectations underpinning competence are made explicit to trainees. This should improve the quality of training.

**Aim**

This research aimed to understand how consultant pathologists determine diagnostic competence in trainees.

**Materials and Methods**

The work was conducted in three stages:

1. A questionnaire study to determine the current perception of workplace-based assessments

2. A qualitative interview study to explore and understand how consultants judge diagnostic competence in trainees

3. A Delphi study to triangulate findings from the interviews

**Results**

Several perceived issues were identified with the current WBAs, particularly their inability to determine diagnostic competence, struggling trainees and accurately reflect trainee progress. Diagnostic competence was conceptualized as a complex construct which should be assessed longitudinally. Both process and person qualities are required, with increasing expectations as trainees progress through training. Diagnostic competence is manifested in trust developing between trainee and consultant, but must be maintained following qualification. A Delphi study triangulated and validated findings from the qualitative interviews with the majority of items reaching “consensus in”.

**Conclusions**

There are problems with the current WBAs as they do not test those qualities which are most relevant to diagnostic competence nor reflect the true judgement ecology. This is problematic given the current emphasis placed on WBAs in training to evidence competence. This research has demystified how consultants determine diagnostic competence and should help to inform future assessment strategies in histopathology. In particular, a longitudinal approach to assessing competence is recommended rather than the current approach of assessing competence on individual cases.

Contents

Volume 1

[**Chapter 1** 1](#_Toc23243729)

[**Literature review** 1](#_Toc23243730)

[**1.1 Assessment** 2](#_Toc23243731)

[1.1.1 Introduction 2](#_Toc23243732)

[1.1.2 Categories of assessment 4](#_Toc23243733)

[1.1.2.1 Summative assessment 4](#_Toc23243734)

[1.2.2.2 Formative assessment 5](#_Toc23243735)

[1.1.3 Assessment for learning 7](#_Toc23243736)

[1.1.4 Standards in assessment 9](#_Toc23243737)

[**1.1.4.1 Angoff method** 11](#_Toc23243738)

[**1.1.4.2. Borderline regression method** 11](#_Toc23243739)

[1.1.5 Utility of assessment 12](#_Toc23243740)

[1.1.5.1 Reliability 13](#_Toc23243741)

[1.1.5.2 Validity 16](#_Toc23243742)

[1.1.5.3 Educational value 21](#_Toc23243743)

[1.1.5.4 Acceptability 22](#_Toc23243744)

[**1.1.5.5 Cost** 22](#_Toc23243745)

[**1.2 Competence and assessment of competence** 23](#_Toc23243746)

[1.2.1 Introduction 23](#_Toc23243747)

[1.2.2 The theory of competence 25](#_Toc23243748)

[1.2.2.1 Millers Pyramid and the Cambridge model: Competence versus performance 25](#_Toc23243749)

[**1.2.2.2 Dreyfus and Dreyfus’ Skill Acquisition theory** 29](#_Toc23243750)

[1.2.2.3 Bloom’s taxonomy 30](#_Toc23243751)

[1.2.2.4 Carter’s taxonomy 31](#_Toc23243752)

[1.2.3 Reflection and competence 33](#_Toc23243753)

[1.2.4 Professionalism and competence 34](#_Toc23243754)

[1.2.5 Competence and Capability 36](#_Toc23243755)

[1.2.6 Competency-based education 36](#_Toc23243756)

[1.2.7 Assessing competence and performance and a move towards Entrustable Professional Activities 38](#_Toc23243757)

[1.2.8 Subjectivity in the assessment of performance 41](#_Toc23243758)

[1.2.9 How is competence determined in histopathology? 44](#_Toc23243759)

[1.2.10 Diagnostic competence in histopathology 48](#_Toc23243760)

[1.2.10.1 Background 48](#_Toc23243761)

[1.2.10.2 Diagnostic competence and accuracy 49](#_Toc23243762)

[1.2.10.3 Diagnostic competence and the process of making a diagnosis 54](#_Toc23243763)

[1.2.10.4 Diagnostic competence and independent reporting 55](#_Toc23243764)

[1.2.10.5 Conclusion 56](#_Toc23243765)

[**1.3 Workplace-based assessments** 57](#_Toc23243766)

[1.3.1 Introduction 57](#_Toc23243767)

[1.3.2 Types and format of workplace-based assessment in histopathology 58](#_Toc23243768)

[1.3.3 Regulation of workplace-based assessments 60](#_Toc23243769)

[1.3.4 The need for workplace-based assessment in healthcare training 61](#_Toc23243770)

[1.3.5 Perception of workplace-based assessments in medical education in general 62](#_Toc23243771)

[1.3.5.1 Introduction 62](#_Toc23243772)

[1.3.1.2 Feedback 63](#_Toc23243773)

[1.3.1.3 Acceptability 65](#_Toc23243774)

[1.3.1.4 Utilisation 66](#_Toc23243775)

[1.3.1.5 Validity 67](#_Toc23243776)

[1.3.1.6 Reliability 68](#_Toc23243777)

[1.3.1.7 Guidance and training 69](#_Toc23243778)

[1.3.1.8 Changes to WBAs 70](#_Toc23243779)

[1.3.1.9 Conclusion 71](#_Toc23243780)

[1.3.6 Perception of workplace-based assessment in histopathology 71](#_Toc23243781)

[1.3.7 Summary 74](#_Toc23243782)

[**Chapter 2** 77](#_Toc23243783)

[**Aims and Objectives** 77](#_Toc23243784)

[2.1 Introduction 78](#_Toc23243785)

[2.2 Aims and Objectives 78](#_Toc23243786)

[2.3 Emergence of a framework 79](#_Toc23243787)

[2.4 Theoretical framework 81](#_Toc23243788)

[**Chapter 3** 87](#_Toc23243789)

[**A questionnaire exploring the perception of WBAs in histopathology** 87](#_Toc23243790)

[3. 1 Background 88](#_Toc23243791)

[3.2 Aim and objectives 88](#_Toc23243792)

[3.3 Method 89](#_Toc23243793)

[3.3.1 Introduction 89](#_Toc23243794)

[3.3.2 Design 90](#_Toc23243795)

[3.3.4 Distribution 95](#_Toc23243796)

[3.3.5 Data analysis 95](#_Toc23243797)

[3.4 Results 97](#_Toc23243798)

[3.6 Discussion 117](#_Toc23243799)

[3.7 Limitations of study 126](#_Toc23243800)

[3.8 Conclusion 127](#_Toc23243801)

[**Chapter 4** 129](#_Toc23243802)

[**A qualitative interview study using Framework Method to explore diagnostic competence in histopathology** 129](#_Toc23243803)

[4.1 Background 130](#_Toc23243804)

[4.2 Aim and objectives 131](#_Toc23243805)

[4.3 Method 132](#_Toc23243806)

[4.3.1 Participants 133](#_Toc23243807)

[4.3.2 Interview guide 136](#_Toc23243808)

[4.3.3 Procedure 136](#_Toc23243809)

[4.3.4 Transcription 137](#_Toc23243810)

[4.3.5 Analysis 137](#_Toc23243811)

[4.3.6 Reflexivity 138](#_Toc23243812)

[4.4 Results 139](#_Toc23243813)

[4.4.1 The Person 149](#_Toc23243814)

[4.4.2 The process 157](#_Toc23243815)

[4.4.3 Stage of training 169](#_Toc23243816)

[4.4.4 Trust 173](#_Toc23243817)

[4.4.5 Time 178](#_Toc23243818)

[4.5 Discussion 183](#_Toc23243819)

[4.5.1 The person 185](#_Toc23243820)

[4.5.2 The process 190](#_Toc23243821)

[4.5.3 Stage of training 197](#_Toc23243822)

[4.5.4 Trust 199](#_Toc23243823)

[4.5.5 Time 200](#_Toc23243824)

[4.6 Limitations and considerations 202](#_Toc23243825)

[4.7 Conclusion 203](#_Toc23243826)

[**Chapter 5** 204](#_Toc23243827)

[**A Delphi study to triangulate the findings from the qualitative interviews for assessing diagnostic competence** 204](#_Toc23243828)

[5.1 Background 205](#_Toc23243829)

[5.2 Aim and Objectives 208](#_Toc23243830)

[5.3 Method 208](#_Toc23243831)

[5.3.1 Participants 208](#_Toc23243832)

[5.3.2 Determining items, rating, ranking consensus, stability and number of rounds 210](#_Toc23243833)

[5.3.3 Procedure 217](#_Toc23243834)

[5.3.4 Analysis 220](#_Toc23243835)

[5.4 Results 221](#_Toc23243836)

[5.4.1 Round 1 rating of items 221](#_Toc23243837)

[5.4.2 Round 2 rating of items 227](#_Toc23243838)

[5.4.3 Round 1 ranking of items 235](#_Toc23243839)

[5.4.4 Round 1 comments relating to ranking items 237](#_Toc23243840)

[5.4.5 Round 2 ranking of items 239](#_Toc23243841)

[5.4.6 Round 2 comments relating to ranking items 241](#_Toc23243842)

[5.4.7 Comparison of round 1 and round 2 ranking of items 242](#_Toc23243843)

[5.4.8 Comparison of round 1 and round 2 highest rated by mean score 244](#_Toc23243844)

[5.5 Discussion 246](#_Toc23243845)

[5.6 Limitations 256](#_Toc23243846)

[5.7 Conclusion 257](#_Toc23243847)

[**Chapter 6** 259](#_Toc23243848)

[**Final discussion** 259](#_Toc23243849)

[6.1 Discussion 260](#_Toc23243850)

[**Chapter 7** 269](#_Toc23243851)

[**Conclusions and recommendations** 269](#_Toc23243852)

[7.1 Summary 270](#_Toc23243853)

[7.2 Conclusions for practice 271](#_Toc23243854)

[7.3 Conclusions for research 271](#_Toc23243855)

[7.4 Recommendations for practice 271](#_Toc23243856)

[7.5 Recommendations for research 272](#_Toc23243857)

[7.6 Further work 272](#_Toc23243858)

[**Appendices** 273](#_Toc23243859)

[Appendix 1: Questionnaire invitation emails 274](#_Toc23243860)

[Appendix 2: Participant Information Sheet 275](#_Toc23243861)

[**Appendix 3: Interview invitation email** 279](#_Toc23243862)

[Appendix 4: Information sheet for qualitative interviews 280](#_Toc23243863)

[Appendix 5: Interview guide 284](#_Toc23243864)

[Appendix 6: Consent form for interview 287](#_Toc23243865)

[Appendix 7: Invitation email for Delphi study 289](#_Toc23243866)

[**Appendix 8: Participant Information Sheet for Delphi study** 290](#_Toc23243867)

[Appendices 9 and 10: Instructions for Delphi and consent form 294](#_Toc23243868)

[**References** 301](#_Toc23243869)

# **Chapter 1**

# **Literature review**

**Chapter 1: Literature review**

# **1.1 Assessment**

## 1.1.1 Introduction

Assessment can take many forms and be used for varying purposes. It has been defined as “*any systematic method of obtaining information from tests and other sources used to draw inferences about characteristics of people objects or programmes”*(National Council On Measurement In Education 1999, p.72) *or* “*any purported and formal action to obtain information about the competence and performance of a candidate*” (Schuwirth and van der Vleuten 2013, p.277). Put simply, assessment involves collecting key information via the means of a relevant method to form a judgement of an individual, object or programme.

Assessment in medical education may be required for many reasons but it might be argued that its main purpose in healthcare is to ensure medical professionals are safe to practice and have the relevant knowledge, skills and attitudes to work effectively in the real-world.

Undergraduate and post-graduate assessment in medical education measures the understanding of the curricula on which training is based. Various stakeholders, including patients, the employer, the trainee and the professional regulatory bodies which oversee training, such as the General Medical Council (GMC) require that the competency of trainees is assessed. There is also a general assumption, especially in professional organisations, that members of those professions must be assessed to maintain the credibility of the organisation (Crossley and Jolly 2012).

In broad terms, assessment serves many functions within education, including (Mislevy 1996):

* maintaining credibility of individuals being certified to enter a profession
* monitoring progress of trainees and levels of attainment
* providing feedback
* assessing learning and teaching strategies
* helping to direct future learning goals.

Clearly, assessment is not just regulatory as it can have other purposes including its ability to stimulate learning, its potential to provide feedback for learning and teaching, for academic league tables, monitoring trainee progression and to identify struggling trainees. It could be argued that the various purposes of assessment can sometimes be at odds with one another and cause confusion. As described later (see section 1.1.5.2), the outcomes from assessments must only be used for their intended purpose and not extrapolated to inappropriate contexts.

Assessment tools must also be valid and reliable if they are used to certify successful completion of training or attainment of a specific qualification or degree. Assessment must be adaptable if it is to successfully measure the range of qualities a trainee is meant to possess. A person’s knowledge may not reflect their ability to complete a task or convey their understanding of a concept, so assessment tools must be varied and robust in their design and selected appropriately. Similarly, the trainee may have a range of perceptions or previous experiences which act to influence their learning and good assessment tools may facilitate the assessor to explore these factors (Masters, Adams and Wilson 1999).

## 1.1.2 Categories of assessment

### 1.1.2.1 Summative assessment

Summative assessment entails measuring trainee attainment and relates to the regulatory nature of assessment. Essentially, it aims to assess whether a trainee has reached a desired “standard” but may act to evaluate the effectiveness of the program being studied (Miller, Imrie and Cox 1998). Summative assessment usually takes the form of an examination; a distinct point in time when a trainee’s knowledge (or relevant attribute) is tested, scored and documented. For example, in histopathology, trainees sit a written exam called FRCPath Part 1 whose main purpose is to assess factual knowledge, while FRCPath Part 2 predominantly tests the ability of the trainee to apply their knowledge and experience in making histological diagnoses on previously unseen cases.

Summative assessment can create a barrier to learning if trainees only learn what they think they will be assessed on (Boud 2000). There is a danger that summative assessment creates a society obsessed with grades and performance at the detriment of learning (Boud 2000). A superficial knowledge base built on a desire simply to pass exams will not create trainees who engage with deeper learning by analysing what they are told, critiquing the literature and asking questions about the academic environment they are part of (MacLellan 2001). It may be argued that a trainee may not only be guided by the education they have received but other factors such as their family life, hobbies, professional and personal relationships and even the constraints of the hospital they work in, can also alter perceptions regarding learning, whether positive or negative. However, education often forms the foundations for learning and may be responsible for instilling values which could be more resilient to external factors.

### 1.2.2.2 Formative assessment

Formative assessment is more continuous and holistic in its approach, with opportunity for feedback and reflection. To this end, formative assessment can help trainees to set goals, monitor their progress towards these goals and facilitate an ethos of self-monitoring and self-motivation, rather than being prompted to do so by an upcoming exam or instruction from an assessor (Brookhart 2001). It aims to guide trainee learning, improve performance and is very much trainee focussed (Boud 2000; Brookhart 2001). Due to its more open and fluid philosophy, it should also provide feedback for the teacher such that learning, teaching and assessment become a more dynamic process (Robertson, Rosenthal and Dawson 1997). In histopathology, formative assessment may take the shape of an informal discussion between the trainee and the consultant regarding a particular case they have been working on together. It may embody feedback provided to a trainee following presentation at a meeting or teaching event.

It is well-known that many medical specialities, including histopathology, are continually evolving and require staff to update their knowledge and skills accordingly yet a trainee who has not experienced an educational pathway which has instilled life-long learning, may struggle to develop effectively (Sandars 2009).That is why formative assessment provides a rich environment to guide and nurture trainee development, through a process of continual feedback and, in turn, the possibility of subsequent adjustment by the trainee. Cyclical feedback encourages trainee professionalism and can encourage confidence in themselves as professionals (Robertson,Rosenthal and Dawson 1997).

**1.2.2.3 Combined approach**

A balance of both summative sand formative assessment is often implemented in education, but the boundaries between the two may be somewhat blurred. Sometimes this is inevitable but it doesn’t always necessitate change as assessment should be flexible. Nevertheless, in certain circumstances, trainees do need to know if an assessment is going to be scored or documented so they can prepare appropriately and learn from the event. From the assessor’s point of view, he/she may also experience personal conflict, as in some instances, they want to assist their trainee’s learning (formative) but in other circumstances, they may be asked to make a final decision regarding their competence (summative) (Duke 1996; Ilott and Murphy 1997; Cross, Hicks and Barwell 2001). As such, assessment should not be a source of reward or punishment but be a source of insight and facilitate trainee progress (Shepard 2000). However, it could be argued that those assessments which are meant to foster learning and development should not be used to measure attainment as the purpose is not to pass judgement. This could undermine the purpose of formative assessment and potentially put strain on the assessor-trainee relationship.

Summative and formative assessments can also be categorised as either objective or subjective. Objective assessment only has one or a limited number of correct answers and may take the form of a multiple-choice question (MCQ) examination. Objective assessment may be perceived as being more reliable but this is not always the case (van Der Vleuten and Schuwirth 2005). Objective assessments can be stifling and limit responses, which can over-simplify the constructs being examined. In subjective assessment the question posed may have several interpretations and more than one correct answer. Essay questions typically allow for more subjectivity in the answer. In reality, nearly all assessment methods have an element of subjectivity (Leach, Neutze and Zepke 2001) , and whenever assessments are designed, they require an exercise in judgement regarding their content and difficulty (Jones 1999). It can be argued that no assessment is completely objective, as most assessments have inherent biases which may stem from examiner preferences or wider cultural influences. In practical terms, it is necessary to ensure that biases or subjectivity is spread evenly across examiners so that each trainee receives a similar experience. However, this is difficult to manage as there are multiple factors to control such as how the test was designed, administered, by whom and under what conditions (Romagnano 2001). This is where examiner briefings, standard setting and post-exam feedback can help to inform future assessment rounds.

## 1.1.3 Assessment for learning

In contrast to “assessment **of** learning”, which aims to determine a certain level of competence, usually by means of an examination, “assessment **fo**r learning” aims to improve learning through constructive feedback. Assessment for learning (AfL) should be embedded within every activity undertaken by trainees, teachers and peers that seeks, reflects upon and responds to information from dialogue, demonstration and observation in ways that enhance ongoing learning (Gardiner 2006). Assessments have been proven to drive learning and have been shown to be the most significant influence on the quality of trainee learning (Wass et al.2001).AfL aims to instil learning that has depth, rather than just enabling trainees with the ability to score well on tests. Although AfL has similarities with formative assessment, some important differences have been described (Swaffield 2011):

* AfL concerns the immediate future. Formative assessments may have a long time span
* The two sides of AfL are the teachers and trainees in a specific class. Formative assessments may involve and be of use to other teachers
* In AfL, trainees exercise autonomy. Formative assessments can render trainees passive recipients of teachers’ decisions and actions
* AfL is a learning process. Formative assessment provides information to direct future learning
* AfL is concerned with learning how to learn as well as specific learning intentions. Formative assessments concentrate on curriculum objectives

Despite these supposed differences, it could be argued that there still appears to be a marked similarity with formative assessment. Indeed, both aim to improve learning and it must be remembered that assessment and learning go hand-in hand, as demonstrated by the many trainees who seek to find out what they can afford not to learn in order to focus only on the areas of the course which they have discovered they will be tested on (Wass et al*.* 2001). Effective or “good” assessments will encourage trainees to learn “what matters”, and the concept of “what matters” may reflect that particular institution’s beliefs and academic standing (Robertson, Rosenthal and Dawson 1997). Trainees will use their institution and its teachers as a benchmark of quality and aim to learn “what matters” to successfully complete the course. As assessment drives learning (Wass et al. 2001), what is not assessed, may not be learnt and remembered. In reality, assessments must focus on the most important and relevant aspects of the curriculum but must be mindful in suggesting, or sending subliminal messages, that other aspects of the course are not also important (Masters, Adams, Wilson 1999).

## 1.1.4 Standards in assessment

Standard setting involves setting benchmarks which categorise important levels of performance. This may be the level which distinguishes a pass from a fail, but often the categories are more complex and reflect a range of performances. Standard-setting is needed to ensure only those individuals who have reached the required “standard” are able to progress, be certified or achieve a particular qualification.

Assessment often requires an established criterion against which individual performance can be measured and is called criterion-referenced assessment. In this model there is a set (or absolute) standard to reach, and attainment of that standard suggests the person is competent e.g. a practical driving test. Norm-referenced assessment is a way of comparing individuals, whereby only a specified proportion of individuals will pass the exam e.g. school entrance exams. The quality of the cohort each year therefore dictates the standard required each year and results are all relative to one another.

Most standards have been formed by educational judgements and can be considered arbitrary (Norcini 2003). Nevertheless, standards must be credible and defensible (Cusimano 1996; Norcini 2003) and embody various elements to achieve this, including:

* be simple to implement and describe
* utilise informed expert judgement
* utilise good research evidence
* produce standards which are consistent to the purpose of the test
* demonstrate due diligence

It is inevitable that not all assessments use the same criteria when standard setting and flexibility is needed at times (van der Vleuten 1996; van der Vleuten and Schuwirth 2005). Hambleton and Powell have previously suggested five context-setting variables which can be considered when standard setting (1983).

* Laws. Do any pre-existing standards exist which define the minimum level of competence?
* Importance of the decision. This refers to the consequences of the decision which can be inferred from the standards set. For instance, tests which assess minor stages of progression within a course are less important than examinations which are used to certify trainees.
* History. It is important that new assessment methods are examined to see if they would be better suited to a different standard setting procedure than current assessments
* Availability of resources. Standard setting requires numerous resources such as time, availability of expertise and materials, each capable of impacting on the standard-setting process.
* Test formats, contents and length. This reflects the fact that there should be flexibility in the methods employed e.g. certain standard setting methods may be better suited to written tests than practical tests.

The two most commonly used standard setting methods used in medical education are the Angoff and borderline regression methods (MacDougall 2015) which are briefly described below (sections 1.1.4.1 and 1.1.4.2). However, it is important to stress that standard-setting does not necessarily mean the assessment is fit for purpose, as the quality assurance mechanisms should be multiple. For example, even if examiners agree that a particular question should have a pass rate of 30% because it is considered “difficult”, one might argue that this question is actually worth removing from the paper if it represents a particularly obscure fact that few trainees will be aware of. This is where examiner briefings can help review the appropriateness of questions. Indeed, there is a balance between those questions which act as discriminators to identify poor, average and excellent trainees and those questions which are just too difficult!

**1.1.4.1 Angoff method**

One form of standard setting is the Angoff process which involves a panel of experts deciding what percentage of minimally competent trainees will pass a particular question. This score is averaged, the results shared with the experts and a second round of ratings completed to create a final cut-off point for each question. The downside to this method is that the experts may struggle assigning scores to each question as the process is difficult. It depends on the expert’s own perceptions of what is difficult or easy for a trainee or which aspects of the curriculum they favour, making it a rather subjective process. However, the method is legally defensible (Tormey 2014). This method means that, potentially, a whole year-group could fail the exam, unlike norm referencing where a percentage of the year group will nearly always pass.

**1.1.4.2. Borderline regression method**

The Angoff method can be categorized as exam-centred because expert judges review the content of the test. In contrast, an examinee-centred method relies on expert decisions based on actual examinee performance (Norcini 2003). The borderline regression method is one such example and is often used in objective structured clinical examinations (OSCEs). Here, at each station, the examiner rates the trainee’s performance via a checklist and gives a global rating score. The checklist marks from all examinees at each station are then regressed on the attributed global rating scores, providing a linear equation (Hejri et al*.* 2013). This number is then split by the global judgements given by each examiner, such as “pass”, “fail” or “borderline”. The point where the best fit regression line crosses the borderline mark equates to the pass mark.

The advantages of this technique include:

* relatively non-time consuming
* can generate metrics to gauge the quality of the OSCE
* it is based on the performance of all examinees
* it incorporates expert judgement

This method of standard setting is not without fault, however. Assessments, such as OSCEs, which rely on subjective judgements of performance can undermine standard setting if the judgement is not founded on any training, guidance, marking rubric or suffers from significant bias.

Other methods of standard setting exist and include the Ebel, Nedelsky and Bookmark methods (Kaufman et al. 2000). They will not be discussed further as they are not as commonly used as Angoff and Borderline regression.

## 1.1.5 Utility of assessment

The “utility of assessment” describes the “usefulness” of any assessment and how it is derived in a multiplicative manner from its reliability, validity, educational impact, acceptability and cost (van der Vleuten 1996). In later years, “feasibility” was also added to the definition but this could be considered implicit in cost effectiveness and acceptability.

It is important that assessment can stand up to the scrutiny of regulatory bodies, especially in “high stakes” assessments where the graduates may enter a profession with significant impact on society (Murray et al*.* 2000). On a local level, the test results and the inferences drawn from them must be defensible to the candidates, examiners and educational establishment (Downing 2002; van der Vleuten and Schuwirth 2005). The utility of an assessment is a useful assessment framework to employ to achieve this ambition. It is described by the utility equation:

Utility = R x V x EI x A x C

(where R=reliability, V=validity, EI=educational impact, A=acceptability, C=Cost).

It can be seen that each element is equally important. Of note, this formula multiplies the different elements, which is why if any elements scores a hypothetical “zero”, the total utility score will always be zero. However it could be argued that certain elements of “utility” are more important than others and should have greater “weight”. For instance, validity is extremely important to an assessment as there is no point having a test that is incapable of testing what it is intended to test (see section 1.1.5.2 for definition of validity). However, despite its simplicity, the formula can act as an aide memoir to those designing assessment tools and ensure the various aspects have been considered.

### 1.1.5.1 Reliability

Reliability is the property of an assessment tool to give the same result if it were carried out on another similar group of individuals, in a similar context at different points in time. There is often confusion regarding the difference between reliability, repeatability and reproducibility (Brennan 2001). Repeatability of measurements refers to the variation in repeat measurements made on the same subject under identical conditions. Reproducibility takes into account how these measurements may vary under changing conditions (Brennan 2001).

Deficiencies in sampling can be correlated with unreliability (van der Vleuten and Schuwirth 2005) so increasing the number of test items, the number of assessments, range of examiners or even combining results of varying examinations, can help to increase reliability (Knight 2000; Wass et al. 2001). Providing training and marking proformas can also improve examiner objectivity (Knight 2000; Yorke, Bridges and Woolf 2000; van der Vleuten and Schuwirth, 2005). However, over-prescription of these can reduce validity so assessments must have a degree of flexibility and yet a minimum level of structure and standardisation (van der Vleuten 1996).

Reliability involves measuring and quantifying errors that occur during assessment, errors which reflect that human performance is variable and affected by various factors (Feldt and Brennan 1989). The reliability of the test may be affected by the perceived importance of the test, the temperature of the room, the time of day it was administered, or any countless number of factors. Threats to reliability include (Feldt and Brennan 1989):

* individuals e.g. their attitude, health, motivation
* instrument variables e.g. scoring errors, sampling errors, poor questions
* situational factors e.g. the context of the assessment
* test marker factors e.g. subjectivity and idiosyncrasy

More specifically, threats to reliability in assessments can include (Crooks, Kane and Cohen 1996):

* the halo effect (a cognitive bias which allows positive traits to positively affect the outcome of a decision)
* the impact of the relationship between the trainee and assessor
* errors in marking
* the Hawthorne effect (when people modify their behaviour when being observed)
* inter-rater reliability

Reliability of assessments may be improved by (Harlen 2007):

* group moderation meetings
* agreement panels
* statistical reference
* external inspection of samples
* defining marking criteria
* *post hoc* adjustment of marks
* accreditation of institutions

It could be argued that while trying to achieve increased reliability, assessments become increasingly objective but also decontextualized. Indeed, assessments might become contrived and artificial, and more emphasis should be placed on the importance of setting, individual perceptions and “authentic” testing. However, this is difficult to implement in reality as sometimes the most authentic testing might not be cost-efficient, feasible or acceptable to the teachers and trainees. This relates back to the utility of an assessment and that no assessment is perfect!

Sometimes a more qualitative approach, as used in qualitative research is recommended because in some instances it offers a more authentic approach. As such, reliability in qualitative research is referred to as “credibility” and “dependability”(Lincoln and Guba 1985). This is because the quantitative or positivist approach to reliability is harder to apply to qualitative research where methods and phenomena may not be as strictly controlled. Reliability in qualitative research is best regarded as a fit between what researcher’s record as data and what actually occurs in the natural setting that is being researched. That may mean that two independent researchers can draw different meaning from exactly the same setting, yet both are reliable interpretations because there may be multiple versions of the same reality. This qualitative approach could therefore offer more authentic testing in some instances, but there are times when the quantitative approach of having a “correct answer” is needed as there is not a place for differences in interpretation. For example, if a patient has a cancer of the oral cavity, the trainee needs to recognise this in their Objective Structured Clinical Examination (OSCE). That is why the quantitative stance is often adopted in medicine, where risks to patients are not acceptable.

### 1.1.5.2 Validity

As reliability is easier to measure, it has often received more attention from medical educators, possibly at the detriment of also determining whether assessment tools are valid. Epstein and Hundert share this opinion, suggesting that many assessment tools may form reliable assessments of knowledge and skills, but the core qualities which constitute a sound clinician have been neglected (2002). Clearly, validity is more important than reliability; there is no point assessing a trainee if the assessment is not testing what it should test, even if it can do it reliably. It is important that validity is achieved, especially in “high-stakes” assessment such as the certification of trainees, or otherwise certain individuals who do not meet the required standard of competency may qualify erroneously. In healthcare, incompetence can lead to serious consequences to patients. For example, if a pathologist was incompetent to make accurate diagnoses, their erroneous report could result in a patient receiving the wrong treatment. Therefore, the procedures for ensuring validity should be transparent so that external regulatory bodies can be satisfied that trainees have been assessed appropriately.

The traditional definition of validity refers to the ability an assessment to test what it is intended to test and represents what it was meant to describe, theorise or explain (Winter 2000). If an assessment is invalid, then it is worthless. An assessment tool may only remain valid so long as it is used on the population for which it was intended, so that it is useful and appropriate. Validity also relates to the interpretation of the results of an assessment and not just the test itself (Downing 2003).

Validity is usually divided into content-related, criterion-related, and construct related but there is often considerable overlap (Brualdi 1999). Validity may also encompass many other types of validity including internal and external validity, face validity, predictive validity, educational validity and consequential validity. Each of these different types of validity will now be discussed.

Content-related: Extent to which the assessment tests the domains being tested within a specified subject area. Curriculum blueprints are often used to ensure all relevant areas are tested.

Criterion-related (correlative): Test scores from different exams are compared to one another for the same individuals e.g. course work versus examination results.

Construct-related: Extent to which the test measures the underlying concepts or psychological constructs being tested. Constructs may include intelligence, self-esteem or critical thinking. These traits are often integral to the profession and may underpin the knowledge or skills that are required by the trainees (Messick 1994).

Internal and external validity: In assessment, internal validity aims to demonstrate that the conclusions drawn from examination results can be sustained by the data. External validity asks if the examination results satisfy external regulators of the profession e.g. the General Dental Council (GDC) quality assures curricula in dental schools to ensure standards are upheld across all dental schools).

Face validity: Put simply, if a test has face validity then it appears to be a valid test to the people who use it. Although not an objective measure, it is important because learners and teachers need to think a test is credible if it is to work. It is closely related to “acceptance” of the tool and the assessment being authentic. It can affect both assessor and trainee engagement in the assessment process (Neary 2000) and is akin to the concept of meaningfulness (Linn, Baker and Dunbar 1991). Meaningfulness is important because when examiner and trainee don’t engage with an assessment process, they may not want to complete assessments or only focus on aspects they perceive as being meaningful. Similarly, if rating scales are felt to be too rigid, supervisors may prefer to evaluate performance informally (based on their own perceptions of a sound performance) with forms filled in to reflect this rather than against the scale descriptors provided (Gomez-Mejia 1998).

Predictive validity: This type of validity aims to determine if current results can predict future performance and can be exemplified by entrance examinations for certain school or courses, where the results from such examinations may infer whether the candidate has sufficient potential to do well at the school.

Educational validity: Explores the relationship of the assessment to its educational function, such as the ability of the assessment to influence trainee learning, how to provide or increase feedback and striking the right balance of formative and summative assessments.

Consequential validity: To what extent can consequences be drawn from the assessment and assessment results. To this end, there can be many and varied consequences, some of which may be beneficial e.g. the assessment identifies that the teaching provided was appropriate to the candidates; while others may have a negative impact e.g. increasing workload for examiners. Consequences may affect the individuals who sat the exam or even the wider society (Downing 2003).

As many different types of validity have been described, this has led to the concept of validity becoming overcomplicated and increasingly difficult for assessments to satisfy the varying definitions. This led Messick and Kane to evolve the meaning of validity to try and unify certain concepts (Messick 1989; Kane 2006). Their more focused approach is to consider construct validity as being the only real type of validity which should be underpinned by evidence from multiple sources. Messick suggested six distinct aspects of construct validity that are interdependent and complementary forms of validity evidence which include 1) content validity 2) substantive validity 3) structural validity 4) generalisability 5) external validity and 6) consequences. However, Messick’s framework does not suggest which sources of evidence are most important or how they might change for different assessments. Kane’s theory addresses this issue of prioritisation as he advises what evidence can be collected to support (or refute) each of the assumptions linked to the assessment and, by extension, the inferences made from it.

In Kane’s theory, if the initial assessment involves assigning a score to a trainee (scoring), the next phase would be generating an overall test score which represents their performance (generalisation). The next stage is how this generalisation is used to make an assumption about how the trainee might perform in real-life situations (extrapolation). The final stage is how the information is used to make a decision about the trainee (implications). Each stage represents an inference with its own assumptions. For each step, evidence must be gathered to support the assumptions. The most complex attributes require the most evidence to substantiate the inferences made from them and vice versa. Just because a test has been validated, this does not mean the results of that test are going to be applied and interpreted properly and as such validation should not be viewed as an endpoint but rather a process.

In healthcare, the ultimate goal of assessment tools is to have publicly accountable measures of the competence of trainees prior to them entering the profession. Here, Messick’s and Kane’s theories are particularly relevant, as the assessments used to determine competence must 1) be sufficiently rigorous and supported by appropriate validity evidence and 2) only used in an appropriate context so the inferences made from the assessment are substantiated.

### 1.1.5.3 Educational value

The educational value of an assessment is an important outcome to examine. Formative exercises rely on quality discussion between the assessor and trainee following certain events or tasks. These episodes of feedback can set new learning objectives, identify strengths and weaknesses and try to address more appropriate learning strategies for the trainee. Summative assessments can also drive learning as trainees seek to learn relevant material to their examination. Van der Vleuten (1996, p.51) notes:

The tremendous impact that the assessment program has on the learner… and there is some hauntingly truth in that students do whatever they are tested on and are not likely to do what they are not tested on.

To this end, constructive alignment can be used so that teaching and assessments are aligned to the desired learning objectives (Walsh 2007). Constructive alignment therefore teaches what really matters, so trainees can see exactly how their teaching feeds into the assessments they are given. This method underpins the current outcome-based approach to education, where learning outcomes are created first, with the teaching and assessment practices devised afterwards to achieve the former. For example, in dentistry, The Association for Dental Education in Europe (ADEE) helped to devise learning outcomes for undergraduate dental education on which curricula and assessment strategies can be based (Field, Cowpe and Walmsley 2017). Assessment tools which can foster reflective practice are also important so that trainees engage more with a deeper level of learning from each experience they have.

### 1.1.5.4 Acceptability

Acceptability can vary between the examiner and the trainee and within each respective group (Lowry 1993). Acceptability may stem from an individual’s experiences, beliefs and values (van der Vleuten 1996). Understanding these items is complex and a pragmatic approach is to consult with the relevant individuals when designing assessments and providing necessary information to ensure the final assessment is acceptable (van der Vleuten 1996). If the examiner and trainee don’t engage with the assessment, then the educational value will be lost. Acceptability of assessments can be linked to whether trainees feel the assessment is valid because if a trainee can see no value in the assessment tool, they might choose to ignore it or disengage with the assessment process where possible.

**1.1.5.5 Cost**

Cost is a useful parameter within assessment because it aims to ensure that time, resources and finances are used appropriately when assessments are designed, administered and marked. MCQs can be very quick to mark because answers are either right or wrong (and can even be scanned by computers), while essays can take a considerable amount of time to mark because they require judgement and interpretation. Efficiency is very much dependent on having appropriate infrastructure, planning and communication amongst colleagues, as exemplified by OSCE’s (van der Vleuten and Schuwirth 2005).

In reality, no assessment tool is perfect and trade-offs have to be made depending on the purpose and nature of the assessment. For instance, workplace-based assessments (when used appropriately) can offer high educational impact, but they might not be performed as reliably as summative assessments such as MCQs. Often, this problem is circumvented by using a range of assessment tools to triangulate assessment outcomes.

# **1.2 Competence and assessment of competence**

## 1.2.1 Introduction

In order to measure competence, one must first define what it is. It has been described for health care professions by Southgate (1999) as the ability of trainees to perform clinical tasks consistently to resolve health issues of individuals while being aware of the social context. Tasks should be completed efficiently, effectively, economically and in a humane manner (Southgate 1999).

However, competence is used widely and in many varying contexts resulting in many different meanings (Eraut 1994). For example, sometimes competence is related to a positive outcome (e.g. completing a task successfully), but may also be considered the baseline level (e.g. it was a satisfactory but not excellent). Competence is often viewed as a binary system, where one can either be competent or incompetent at completing a task. However, this can sometimes lead to a superficial view of competence, and a linear scale of competency may provide more feedback to learners and better reflect the range of competencies observed. Competence may be difficult to define because it cannot be directly observed; only inferred from demonstrated behaviours. It is made of several components or characteristics, some of which may be recognisable while others may be imperceptible.

The behaviourist would probably define competence as the ability to perform a specific set of tasks (Eraut 1994), but it could be argued that the approach of separately assessing individual skills may not reflect the more complicated and interchangeable environment of the workplace and be criticised as having a cursory stance (Grant 1999). Some people may prefer this approach as it easier to separate out individual tasks and build a framework on which assessment of these competencies can be built. However, the disadvantages have included lowering of standards in certain professions and its ignorance of more abstract or tacit behaviours, skills and attitudes (Grant 1999). The overall perception of an individual’s performance may then be overlooked. Focusing on specific skills may also neglect an appreciation of the need to adapt and be flexible in the work environment and the transfer of skills to new or unfamiliar environments.

Trainers and consultants often use the term “competence” when discussing the relative merits of their trainees. In their opinion, they may know what they are looking for when making a judgment regarding whether their trainee is competent or incompetent at a particular task. Defining competence and objectively measuring it may pose much more difficulty, however. Therefore, it is important that the trainer and trainee both understand what constitutes “competence” if the trainee is to have a fair chance at becoming “competent”. Competence may be related to a simple or discrete task (e.g. tying a suture) or used to determine if a trainee completed a complicated procedure requiring many different skills (e.g. the entire surgical operation).

Competence may encompass what is “acceptable” to a particular medical speciality, as judged by its peers or patients; it may simply be whether the trainee got the “right” answer; if they were “safe” or implemented the “correct” treatment plan; maybe a competent performance is an impression someone has of you; is it how you cope with difficulties and conduct yourself or the ability to undertake procedures and tasks without constant supervision or support? It may be viewed simply as the ability to “get the job done” or incorporate a combination of practical and theoretical knowledge, along with certain behaviours and values.

A judgement of competency could potentially vary depending on who is judging, the task being undertaken and even the day of the week. For example, some examiners, by their very own nature, examine more harshly than others (Hawk-dove effect) and on certain tasks (Hawthorne effect). One task may require excellent communication and empathy (e.g. informing a patient of a terminal diagnosis) whereas another calls for acute concentration and excellent manual dexterity (e.g. performing a complex operation). Clearly, the concept of competence appears less tangible than it first seems. Competence can be elusive (Ilott and Murphy 1997).

## 1.2.2 The theory of competence

### 1.2.2.1 Millers Pyramid and the Cambridge model: Competence versus performance

In Miller’s conceptual model (Figure 1.1), there are traditionally four stages of development. The first two stages at the bottom of the pyramid, “knows” and “knows how”, can be assessed using assessment tools such as written and oral tests. “Shows” and “does” are the higher levels of the pyramid, with the former usually assessed in exams such as OSCEs and the latter normally demonstrated from observations of the trainee in the “real world” and Workplace-based Assessments (WBAs). Successful progression to the highest levels of the pyramid requires the trainee to gain the necessary skills, knowledge and behaviours expected of a professional in their relevant specialist field (Miller 1990).

Expert

Novice

**Figure 1**.**1.** Miller’s pyramid (Miller 1990).

Thus practical experience increases and is augmented with existing skills and knowledge, eventually reaching “knowledge-based practice” theorised by Ramussen over 20 years ago (Rasmussen 1983).Rasmussen described situations where decisions have to be made even when not all the necessary information is available-a situation often encountered in medicine (Rasmussen 1983). He argued that in practical tasks, skills are acquired first before the full knowledge relating to the skill has been learnt. Over time, experience is gained from using these skills and augmented with knowledge, such that the individual would eventually reach “knowledge based practice” which is the highest level. Long has quoted Rasmussen’s theory as forming the basis for competency based training in medicine (Long 2000).  However, it could be argued that Rasmussen’s theory lends itself to more practical fields such as surgery, rather than the work of a pathologist. For instance, the skill of making a diagnosis cannot be acquired before all the knowledge has been acquired, as one would be ignorant to the range and depth of possible pathological diagnoses and come to an erroneous conclusion.

Performance and competence have been conceptualised in the literature as being different, whereby competence is what a person is capable of in ideal circumstances whereas performance is how that competence is affected by the environment or motivation of the individual to succeed at the task at hand. However, many examinations assess this underlying competence (in the ideal environment), as otherwise it could be argued that the environment was not consistent between individuals and the test considered unfair (Bargagliotti 1999). But what about when the trainee enters the real workplace-can they still function to the required level? This underlying competence is akin to “knows” in Miller’s pyramid whereas performance relates to the highest level of the pyramid, “does” (Miller 1990); and being able to “do” something under less than perfect conditions. Employers and the general public are more interested in an individual’s performance rather than their ability to just pass exams and this view is in alignment with the two main reasons for assessing competency: certification and fitness to practice.

Competence is a reflection of what one does when one is being watched, under ideal conditions, where one might feel challenged to demonstrate their exemplary knowledge, skills and attitudes. Performance, however, deals with how a trainee performs in the workplace when the degree of observation and control over external factors is less rigorous. A trainee may be competent but incapable of a competent performance when conditions are not ideal (Eraut 1998; Rethans et al. 2002; Schuwirth et al. 2002). Furthermore, factors such as anxiety stemming from knowing you might be assessed, or altering your behaviour for the simple reason that you know “at some point” you might be assessed, is not likely to be representative of how a trainee would perform in the real world. Assessments should actually focus on equipping trainees with the skills which will encourage them to continue to learn, adapt and respond to the real working environment (Hargreaves 2016). A competent performance would encompass the necessary knowledge, skills and personal qualities one might expect as well as the ability to adapt, continue to learn and remain safe. Professional competence is not just a set of unrelated competencies joined together, it is the ability to adapt and respond to environmental factors (Epstein and Hundert 2002). Trainees must pass through an educational system that is “future-proofed” such that the trainee can give a performance which would be acceptable upon completion of training and within real work environments and continue to be competent and deliver a competent performance, years after qualifying (Harris and Simpson 2005).

Akin to the idea that exam-focused trainees who fail to appreciate deeper learning have not become equipped to self-regulate and self-monitor their future learning needs, the ability of a trainee to adapt to a change in environment rather than having a “one size, fits all” approach, is clearly advantageous. The context of practice and the way that a professional’s work interacts with this context is often reciprocal (Hunt, Adamson and Harris 1999). This means that a trainee’s actions and behaviours can affect their working environment as the trainee is an integral part of that environment.

It could be argued that one fault of Miller’s pyramid is that it suggests that competence predicts performance, which is not necessarily true. Clinical practice is infinitely complex and many factors can affect an individual’s performance. Miller’s pyramid has now been modified by Rethans and colleagues to form the “Cambridge model” which reflects factors which influence performance other than basic competence (Rethans et al 2002). These influences include system related influences such as guidelines and individual related influences such as mental health and relationships with colleagues. This model acknowledges that competence should be assessed in realistic environments and we should not just assess knowledge and practical skills in isolation. Performance therefore includes the ability to manage ambiguous problems, tolerate uncertainty and make decisions with limited information (Schon 1983).

**1.2.2.2 Dreyfus and Dreyfus’ Skill Acquisition theory**

This theory describes how individual’s acquire competence through formal instruction and personal experience, gradually evolving from novice to expert (Dreyfus and Dreyfus 1980). It was originally developed for the training of pilots and comprised five stages:

1. Novice: Rule-based behaviour, strongly limited and inflexible
2. Advanced Beginner: Incorporates aspects of the situation and organises key aspects of situation presented
3. Competent: Acting consciously with definitive planning and action plan
4. Proficient: Sees the situation holistically as a whole and acts from personal conviction
5. Expert: Has an intuitive understanding of the situation and transcends rules

This linear model of competence acquisition could be criticised for being too structured. Expertise does not necessarily develop in a step-wise manner either. In the healthcare setting, intuitive decision-making can be dangerous, and certain situations require evaluation of a number of tests to formulate the correct treatment plan for a patient, with little place for intuition. Benner has utilised the Skill Acquisition Theory to describe nursing practices, but suggests that perception of ethical issues and being emotionally involved with patient care are also involved in developing expertise and intuition (Benner, Tanner and Chesla 1996). Sometimes theory does not sufficiently reflect the complex phenomena that occur in the workplace (Benner, Tanner and Chesla 1996).

### 1.2.2.3 Bloom’s taxonomy

Bloom's taxonomy is a set of three hierarchical models covering learning objectives in cognitive (thinking, knowledge), affective (emotions, feelings) and psychomotor domains (physical) (Bloom et al. 1956). In particular, the cognitive domain is used in education to help devise curricula, assessment tools and learning objectives. The original cognitive domain had six levels: recall; comprehension; application; analysis; synthesis; and evaluation. However, this was later revised to replace nouns with verbs and a slight reordering of the items: Remembering, understanding, applying, analysing, evaluating and creating (Anderson and Krathwohl 2001). At the bottom of the pyramid, remembering and understanding are associated with” lower order thinking” whereas analysing, evaluating and creating” are forms of “higher order thinking”.

Although originally designed to aid learning, the three domains described in the taxonomy (cognitive, psychomotor and affective) can also be applied to the workplace but there are some important differences. For instance, theoretical knowledge is considered different to professional knowledge, with the latter being more dynamic and acquired independently of theoretical knowledge (Benner 1982). Professional knowledge may be composed of propositional, craft and experiential knowledge (Beeston and Higgs 2001). Propositional knowledge has foundations in theory and research. Craft knowledge relates to developing skills to practice one’s own profession, similar to intuition and other tacit behaviours. Experiential knowledge is often personal to the individual and comes from their experiences in life and work, in keeping with Bloom’s affective domain. The interaction of the various components of professional knowledge reflect the ambiguous nature of the “real world”(Harris 1993).

Regarding the nature of “skills”, these are obviously integral to professional practice but by trying to list each individual skill, one may lose the complexity and dynamic nature of the overall “whole”(Schuwirth et al. 2002). Skills in professional practice have been described as being technologies, craft and art (Harris 1993). Technologies are specific and prescribed procedures which are not open to interpretation; technologies are therefore relatively easy to assess (Chapman 1998). Craft skills allow for interpretation by the individual and may be adapted to particular situations. Artistic skills relates to having “artistry” in your profession, almost like having flair or an ability to work at a level that colleagues may strive for, but may never achieve.

Different professions place differing degrees of emphasis on particular “attitudes”, which probably stems from what is important and expected of that profession. Sometimes skills are confused with attitudes. For instance, professionalism may be described as a skill, but includes aspects of accountability, tolerance, reflection and integrity (Higgs and Hunt 1999).

### 1.2.2.4 Carter’s taxonomy

Bloom’s taxonomy identifies important attributes which help to decipher the idea of competency, but perhaps it does not satisfactorily describe the learning outcomes expected of a trainee entering their profession. These shortcomings were explored by Carter (Carter 1985). For instance, knowing how to do something is not the same as actually being able to do it. Other criticisms of Bloom’s taxonomy are related to the categories being too broad and not allowing examination of specific qualities. Carter’s taxonomy contains three principle divisions, each with a cognitive and an affective component. The divisions are knowledge, skills and personal qualities. Taking the knowledge division as an example, the cognitive component would cover factual knowledge whereas the affective component covers experiential knowledge. To this end, knowledge is more than just knowing facts, it is about understanding the principles that underpin the facts. Carter’s approach therefore emphasises the qualities required for professional practice and the importance of each category can be adapted to fit each individual profession.

For instance, a teacher may need to place emphasis on the value of the individual but a manager would place more emphasis on groups of people (Carter 1985), allowing a more tailored approach. With regard to the “skill” category, it would be simplistic to focus only on a particular skill, but Carter’s approach includes aspects of judgement and reflection, offering a more holistic taxonomy which is aligned to professional practice.

Higgs and Hunt also developed a model of competency for health professionals, where competence is viewed in the context that the professional works in and covers four main domains of professional competence (Higgs and Hunt 1999). They focus on skills and qualities which are more integral to the workplace and the broader context of the profession, including technical competence, interpersonal competence, interacting with others, coping with change, professional responsibilities, accountability and even interactions with society.

Clearly, knowledge, skills and attitudes are integral to competency and do not exist in isolation. Dreyfus and Dreyfus highlight that good practice requires sound knowledge, and similarly, sound knowledge without evidence of practice will not produce the correct behaviours, especially in complex scenarios (Dreyfus and Dreyfus 1996). It is therefore important not to view the integral aspects of competency as being separate and unrelated, but one should appreciate the interplay between knowledge, skills and attitudes in achieving competency. For instance, one cannot manage patients on a hospital ward without sufficient medical knowledge, but even with sufficient knowledge, the individual must be aware of the ethical and cultural implications of their proposed treatment (Boshuizen and Schmidt 2000).

## 1.2.3 Reflection and competence

Competency is not a static quality, somehow maintained throughout one’s career by the virtue of passing an exam 20 years ago! Reflection is strongly linked to improving and maintaining clinical competence (Hargreaves 2016). Reflection is a metacognitive skill involving a process of deconstruction and reconstruction where individuals make sense and give meaning to their practice (Schon 1983). Reflection can be thought of as using thoughtful practice into a potential learning situation (Kember et al. 2000). It has become an integral part of training and assessment in nursing, teacher education, social care and the training of doctors (Boud 2000). Regular feedback and in-depth discussions between a trainer and trainee may help the trainee to examine their performance in detail and act upon aspects of their performance which require development (Prescott-Clements et al. 2011). Reflection improves deep and lifelong learning, improving professional performance for the benefit of patients (Hargreaves 2016). For example, reflective practice was found to promote clinical competency in medical imaging undergraduates, which helped them to prepare for professional practice (Chelliah and Arumugam 2012). Medical schools are being encouraged to design curricula that align with reflective practice or motivate trainees to want to learn (Kusurkar et al*.* 2012). Medical education should promote learners’ reflection and self-directed assessment seeking (Eva and Regehr 2008).

However, it has also been said that there is little evidence that reflection correlates with competence (Mann, Gordon and MacLeod 2009). This may be because it is a difficult area to assess as there are not defined “levels of reflection” such that one can correlate “excellent reflectors” with those who are most competent. Indeed, reflection is also a very personal process with much of it being unobservable. Even when it is encouraged and individuals are asked to document their reflection, does this really indicates the level of reflection that has actually occurred?

## 1.2.4 Professionalism and competence

Personal qualities are often valued in the assessment of competence, incorporating many aspects of professionalism. For example, a recent study showed that doctors take into account how individuals interact with others, their communication skills, teamwork abilities and ethical principles when assessing competence (Hojat et al. 2007). Indeed, professionalism differs from knowledge and skill capabilities as it has been described as personal qualities beyond the obligatory medical knowledge and procedural skills that health care professionals must possess (Veloski 2006). Professionalism is now considered an integral part of training curricula (Zijlstra-Shaw, Robinson and Roberts 2012). Medical professionalism has been defined as a set of values, behaviours and relationships (The Royal College of Physicians 2005) which especially includes integrity, compassion, altruism, continuous improvement, excellence and working partnership with members of the wider healthcare team (The Royal College of Physicians 2005). Here, the definition does not just describe behaviours such as “compassion” and “altruism” but also includes aspects of reflection such as “continuous improvement” and vaguer terms such as “excellence” which presumably relates to aspects of knowledge, skills and behaviour which deliver excellent patient care. In contrast, The Accreditation Council for Graduate Medical Education in America (ACGMEA) stated six core competencies for resident education which are being used to design curricula (medical knowledge, patient care, professionalism, interpersonal and communication skills, practice-based learning, and systems-based practice) (Merlin et al. 2014). Here, professionalism is considered a separate competency, distinct from the other competencies such as “communication” and “medical knowledge”. Confusion is compounded when the term “professional competence” is used, as this phrasing tends to incorporate aspect of competency and professionalism, such as:

Professional competence is the habitual and judicious use of communication, knowledge, technical skills, clinical reasoning, emotions, values, and reflection in daily practice for the benefit of the individual and community being served (Epstein and Hundert 2002, p. 226)*.*

Given the complex nature of professionalism it can be difficult to separate it from competency. Indeed, professionalism has been described as a second-order competence and is manifested in all that we do and say (Zijlstra-Shaw, Roberts and Robinson 2013). It could be argued that professionalism therefore acts to unify and harmonise our competencies. Maybe professionalism and competency have become confused by some because when they seek to measure “professionalism”, they can oversimplify what it is by viewing it as a standalone competency when it really shouldn’t be examined in isolation?

## 1.2.5 Competence and Capability

Competence may be underpinned by certain personal characteristics of the individual, such as their attitude. But do these personal attributes translate into a competent performance (Eraut 1998). This can be explained by the term “capability” which is more a property of the individual and their ability to adapt to new environments, learn from others and transfer their competencies to others. With this in mind, capability empowers and facilitates an individual to learn new competencies (Eraut 1994). Successful work performance is more than performing discrete tasks successfully but also involves effective holistic integration and coordination of these tasks.

## 1.2.6 Competency-based education

Training in medical and dental education often aims at determining when a trainee is competent to move to the next stage of training. In general terms, organizations may stipulate their desired competencies so people know what they need to do to produce the desired outcomes of the organization and this forms a framework for appraisals and development of staff. There is an increasing emphasis towards competency based training in the United Kingdom (UK) and other countries worldwide, where there is pressure to increase accountability and set minimum standards for entry into practice (Brightwell and Grant 2013).

Competency-based education is usually assessed longitudinally as it takes time for trainees to acquire the necessary knowledge, skills and attitudes to be signed off or be “competent” at particular tasks. This echoes Carr who comments that:

It is generally recognised that it is no longer sufficient for doctors to consider themselves competent but...demonstrate they are clinically competent to perform certain roles (2004, p. 63 ).

Therefore, it could be argued that a trainee is not fully competent until they are signed off for all the roles or tasks. Individuals must therefore recognise the limits of their abilities and not partake in work beyond their scope of competency. Interestingly, despite the emphasis placed on competency-based education, there is little consensus among trainees, trainers, and committees on what this approach entails (Leung 2002).

There has therefore been an effort to make competency based education more explicit, with, for example, the ACGMEA listing six competencies expected of medical graduates: medical knowledge, patient care, interpersonal and communication skills, professionalism, practice-based learning and improvement, and systems-based practice (Rose et al. 2009). These competencies reinforce the fact that having the necessary knowledge and skills are not enough and individuals must develop a wider range of skills and attitude so that they can develop throughout their career (Merlin et al. 2014). Indeed, one criticism of some competency-based education systems is that its reductive nature can sometimes lead trainees into thinking they understand and have met all that is expected of them when they finish training. However these expectations do not always align to the context of real work responsibilities or the likely scope of practice (Graham et al*.* 2009; Carraccio and Burke 2010). Sometimes the subcomponents of a competency do not always add up to the “whole” or reflect the connections between them (Rasmussen 1983).  Competency based education has been criticized as being politically motivated so that the government could assign resources based on the outcomes of performance (Leung 2002). The system has also been accused of being impersonal and rigid, which can demotivate some individuals (Carr 2004). Competencies not represented in traditional curricula might not be explicit to trainees and therefore undervalued, especially when trainees have limited access to role models. For example, in the Canadian accreditation system (CanMEDS), advocacy (defined as physicians responsibly using “their expertise and influence to advance the health and well-being of individual patients, communities, and populations)” constitutes an explicit expectation for every member in the profession (Frank 2005). While this ethos might help trainees to see what qualities are valued, it still relies on having enough staff to act as role models.

However, competency based education does also have several advantages, including:

* it improves public accountability of medicine
* standards are transparent to trainees, the profession and the public
* by defining and listing core competencies, it is sometimes easier to then identify gaps in training which may stem from varied clinical exposure
* as progression through training is based on acquiring competencies, trainees can progress at their own pace (although this theoretical flexibility may not be a reality in practice)

## 1.2.7 Assessing competence and performance and a move towards Entrustable Professional Activities

The assessment of competence is done for a number of reasons (Kak, Burkhalter and Cooper 2001) including:

* healthcare reform
* organisational performance
* planning for new services
* measuring training outcomes
* liability and ethics
* risk management
* certification of providers
* selection of new staff

For the purposes of this discussion, measuring training outcomes is most relevant. This can be assessed in many ways including OSCEs, written examinations, direct observation, peer ratings and patient satisfaction surveys (Hojat et al. 2007). However, the most commonly utilised form of assessment in graduate medical education involves senior members of the team (e.g. consultants) rating performance using pre-designed forms, such as WBAs (Gray 1996; Silber et al*.* 2004). In other words, it is the work trainees do in the real working environment, or their performance, which is most important rather than separate competencies which tend to be recorded at undergraduate level. Performance assessments focus on a trainee’s ability to understand underlying concepts of a task rather than being just a test of knowledge or specific skill (Masters, Adams and Wilson 1999). These assessments are undertaken in the workplace where direct observations of the trainee or individual can be made (Boursicot, Roberts and Burdick 2010). Performance assessments are often fairly dynamic, exploiting changes in the environment, assessor and format to document a well-rounded perception of performance in the workplace (Arnold 2002; Cruess et al. 2006; Van Mook et al. 2007). Clinical context must also be integrated into curricula but without detracting from the need to develop a sound knowledge base to begin with (Harris and Simpson 2005). A criticism of this approach is its time consuming nature, the need to provide adequate training for observers and completing multiple assessment over time in different circumstances (Luijk et al. 2000).

Peer assessment is also utilised in many fields as peers often foster close relationships with their colleagues, offering unique insights into their values and attitudes (Arnold et al. 2007). These moral and affective domains are also more accurately evaluated by peers than examiners or regulatory bodies (Asch and Saltzberg 1998). It is important that peer assessment insight does not interfere with relationships between trainees, as these are important to their psychosocial development.

The problem with self-assessment can be that it relies on the individual to report on their own limitations, a process which heavily relies on the ability to reflect and self-regulate. However, it is still an important exercise to teach trainees because it is conducive to instilling life-long learning and professionalism. If organised properly, such as when the trainees are given prior warning to collect evidence on their performance and to discuss it with their colleagues, it has been shown to be acceptable form of assessment to trainees and assessors (Rees and Shepherd 2005).

More recently, Entrustable Professional Activities (EPAs) have been introduced. An EPA can be defined as a unit of professional practice that can be fully entrusted to a trainee, as soon as he or she has demonstrated the necessary competence to execute this activity unsupervised (ten Cate 2015). The concept was developed by ten Cate (2005) as a means to facilitate competency based postgraduate medical education. While competencies are descriptors of the qualities of individual persons, EPAs describe the work that is being done or must be done in the workplace (ten Cate 2015). They essentially ask the assessor or supervisor what really matters e.g. Can I trust this trainee to get on with the task? In histopathology, this might translate to a consultant trusting a trainee to report pathology cases independently. Pressure on patient safety and stricter supervision over trainees is an increasing trend and a major challenge for medical educators is to let trainees assume responsibility for patient care (Kennedy et al. 2005; Teman et al. 2014). An EPA-based competency curriculum aims to establish this gradual increase of responsibility and autonomy in a safe and justifiable way (ten Cate, Snell and Carraccio 2010). Even though EPAs are not widely in use within the UK evidence from other countries so far seems positive (Hauer et al. 2013; ten Cate 2013; ten Cate 2005; Chang et al. 2013; McCloskey, Dormen and Conran 2017; Pittenger et al. 2016; Jones et al. 2011). For example, some authors have argued that what physcians do in practice is far greater than the sum of competencies within a curriculum, making EPAs a highly desirable alternative (Shaughnessy et al. 2013). An article in the British Medical Journal said EPAs would help curriculum writers identify the most important tasks to be mastered during training rather than focusing on trainees ticking multiple boxes (Black, 2016). These tasks can then be linked to the most crucial competencies, such that trainees can see what competencies are preventing tasks from being entrusted to them (Black 2016). It could be argued that there are situations when trainers must know that their trainee can be trusted to perform in a particular clinical situation and EPAs address this real need. In the past, the reductionist approach to assessment has risked measuring what is measurable but not important (Snadden 1999).

## 1.2.8 Subjectivity in the assessment of performance

The measurement of clinical competence is a difficult task as many of the assessment tools rely on human judgement (Albanese 2000; Govaerts et al. 2007; Schuwirth and Ash 2013; Wood 2014). The task is complex as the perception of an individual’s knowledge, skills and behaviours is based on another individual’s observations of that person. These subjective judgements underpin assessment tools such as WBAs. Assessor subjectivity has been explored by Govaerts and van der Vleuten and in particular what constitutes a “good” performance and how this may depend on who is making that judgement (Govaerts and van der Vleuten 2013). Within the fields of medicine and psychology, it has been shown that decision-makers often have poor insight and consistency into how they make and apply judgements to scenarios (Unsworth, Harries and Davies 2015). This has prompted numerous research studies which have tried to understand examiner bias despite observing similar trainee performance (Govaerts et al. 2007; Gigerenzer and Gaissmaier 2011). It appears that examiner judgement is influenced by their own experiences, biases and ideas about what constitutes a certain level of performance (Gigerenzer and Gaissmaier 2011).

Almost any judgement is influenced by a long list of facets that includes, at the least (Crossley 2013):

* the baseline stringency or leniency of the judges (the hawk-dove effect)
* the case-related biases of individual judges (the hobby horse effect)
* the subject-related biases of individual judges (the halo effect)
* the ability of individual subjects across all cases of challenge
* the case-related ability of individual subjects (aptitude)

The way an assessor critiques their trainee provides rich data which may reflect the assessor’s own biases and values more much more than the assessment provides information on how the trainee coped with the particular problem being posed (Crossley 2013). Would this assessor’s judgement be consistent on a different day of the week and how would it compare in a year’s time? (Crossley, 2013)

Tacit behaviours such as communication skills are harder to assess than knowledge as they require a much more personal interpretation of the individual and may be more at risk from subjectivity. Interestingly, when researchers have tried to break down tacit behaviours such as professionalism into distinct elements which constitute that behaviour, it doesn’t make the assessment of that behaviour any easier, suggesting that the overall judgement of someone’s professionalism, for example, may be more useful than the sum of the individual parts (Regehr et al. 1998).

This makes the planning of assessment difficult and unpredictable (Kogan et al. 2011) which is a largely unresolved issue (Cook et al. 2009). For instance, studies that have tried to incorporate behavioural anchors found only small improvements (Donato et al. 2008). Two randomised controlled trials examining if rater training could improve rater scoring showed either inconsistent or insignificant effects (Holmboe, Hawkins and Huot 2004; Cook et al. 2009) Determining and understanding what cues examiner’s use when making their judgements would therefore be a useful exercise if their biases cannot be eliminated.

In reality, most assessments, in any given field, have an element of subjectivity and designing assessment tools which are entirely objective may be impossible. For example, assessors have been found to nearly always value academic knowledge over tacit behaviours (Jones, 1999). As trainees are usually assessed by experienced individuals, it would seem reasonable to assume that these assessors could make an appropriate judgement when called to do so. For instance, when a trainee is awarded a Doctor in Philosophy (PhD), this is usually at the discretion of the assessors present in their viva. Similarly, in the design of an assessment tool, the content to be examined by the assessment tool is often determined by a panel of experts (Messick 1989). It could be argued that subjectivity is sometimes useful as there are instances when there is more than one way of saying or doing something and being subjected to differing opinions on this can help develop a broader view in trainees. It is therefore difficult to entirely remove subjectivity from assessments, but the key is to focus on those assessments which have the biggest inferences, such as high-takes exams which determine completion of training. These must be supported by good evidence, which would include how the assessment accounted for subjectivity. This is likely to consist of providing appropriate calibration of judges and clear marking rubrics to help guide marking.

## 1.2.9 How is competence determined in histopathology?

The Royal College of Pathologists (RCPath) has adopted a competency-based framework to guide its curricula and assessments, such that competency is judged in stages (A-D) according to the trainee’s level of experience (The Royal College of Pathologists 2015). The curriculum was updated in 2015 and provides an extensive account of training, the entire content of which is beyond the scope of this chapter. However, to summarise, the curriculum includes:

* the rationale, learning theories and content of the curriculum
* what is expected of trainees at each stage of training
* the knowledge, skills and attitudes relevant to histopathology
* the assessment tools utilised during training
* generic skills required in histopathology, such as professionalism

The training in histopathology is approximately 5 years with the two major examinations being the FRCPath Part1 (usually taken between years 2-3) and the FRCPath Part 2 (usually taken between years 3 and 5). When trainees pass their FRCPath Part 2 exam, they enter Stage D of training which lasts a minimum of 12 months. The guidance provided by the RCPath states that it is “an opportunity for the trainee to start to develop areas of specialist interest and also attain confidence with independent reporting”. This stage of training is also present so the trainee can gain all the competencies required to be signed off and acquire their Certificate of Completion of Specialist Training (CCST). To this end, the curriculum recommends minimum numbers of cases that a trainee must report and these cases should range from straightforward to complex. In addition, specific stage D competencies include:

* demonstrate a level of knowledge and skill consistent with practise as a general or specialist consultant in histopathology in the National Health Service (NHS) to include Multi-disciplinary Team meeting (MDT) presentation.
* demonstrate the ability to report independently
* explore specialist interest if applicable or maintain in depth general reporting
* develop experience of teaching histopathology trainees and other professional groups

The RCPath has also utilised a competency-based framework for graded responsibility, such that trainees may be “signed off” for certain diagnoses during training (The Royal College of Pathologists 2009). Here, the different body systems (e.g. head and neck pathology) have an associated list of lesions which a trainee should be able to report at each stage of training. Once a trainee is signed off for each lesion they would then, in theory, be able to report this lesion independently. At level 1 a trainee can be signed off for reporting fibro-epithelial polyps in the head and neck, yet it is only at level 4 (the highest level) that a trainee can be signed off for a granular cell tumour. While it is important that trainees are given responsibilities so that they can have experience of acting like a consultant, it is not clear how helpful it is to ascribe certain lesions to particular stages of training. For example, while a fibro-epithelial polyp is straight-forward to diagnose, not every lesion which is polyp-shaped will be a fibro-epithelial polyp. To explain further, a spindle cell carcinoma can also present as a polyp and the differences can sometimes be quite subtle. Could a trainee potentially miss the spindle cells and sign this case out as a fibro-epithelial polyp if they have never seen a spindle cell carcinoma before? There may be some truth in the phrase “ignorance is bliss”!

Currently, it is not clear what evidence is used to support graded competency decisions but the outcomes of WBAs could offer some insight. However, independent reporting is not utilised in all training centres and often, trainers only allow independent reporting once trainees have passed their FRCPath Part 2 exam. In some centres, these reports would also be audited by the overseeing consultant to ensure correct diagnoses are being signed out by the trainee. Eventually, once the consultant can see that correct diagnoses are being consistently given, the trainee would no longer need auditing. This step-wise approach can help to protect patient safety as well as providing evidence that the trainee’s work has been quality assured.

Within the histopathology curriculum document, it states the competencies required for each stage of training and the assessment tools used to assess these competencies (Royal College of Pathologists 2015). More detail is given in the “Blueprints for the pathology assessment systems” document which maps assessment tools to each of the learning outcomes (Royal College of Pathologists 2010). Exams (FRCPath Part 1 and 2) are sometimes listed, but the educational supervisors report (ESR) and WBAs are utilised most widely. The ESR is an annual statement on the trainee’s progress, which to some extent, may use outcomes from WBAs to inform its content. Based on the curriculum and assessment blueprints, WBAs are probably the most frequently used assessment tool to assess competency in histopathology. Of note, the curriculum actually states that WBAs are meant to be formative in nature but one might argue that they are being used to assess competency which is more in line with summative assessment. The confusion is compounded further by the curriculum stating that a minimum number of WBAs are required per year. Setting minimum numbers for a formative assessment risks making the assessment sound summative as formative assessments should be geared towards individual learners and not a set prescription.

Identifying relevant competencies and grouping or assigning them to overall outcomes can be challenging (Jones et al. 2011) but in the United States, the Next Accreditation System (NAS) has utilised educational milestones within pathology education (Vitek et al. 2014; Naritoku et al. 2014). Educational milestones are created for a particular speciality and document, in a sequential way, a trainees’ progression through a defined set of competencies. Trainee progression is assessed at regular, established intervals throughout training, using tools analogous to WBAs. Similar to milestones, rubrics are used widely in education to assess performance (Moskal and Leydens 2000; Jonsson and Svingby 2007). A rubric consists of a set of statements or criteria which explicitly define what is expected of a trainee at different stages of training. When designing rubrics, five areas are often examined: elements which are being assessed, levels of achievement possible, descriptions of performance and their consequences at each stage and a rating scheme. Advantages of using analytic rubrics include reduced subjectivity, clearer expectations of trainees and trainers and the ability to give more structured feedback (Moskal and Leydens 2000; Jonsson and Svingby 2007).

## 1.2.10 Diagnostic competence in histopathology

### 1.2.10.1 Background

Patients might suggest that the single most important skill of a pathologist is the interpretation and diagnosis of disease. The histological diagnosis is, after all, the gold standard on which patient treatment is based. A pathologist who cannot make an accurate diagnosis is unable to perform in his or her role. It is generally expected that there should be zero error rate or patients may potentially receive the wrong treatment if their diagnosis is incorrect. In reality, a zero error rate is rarely achieved (Safrin and Bark 1993; Lind et al. 1995; Ramsay 1999). Despite the thoroughness of the curriculum provided by the RCPath, summarised in the previous section (1.2.9), it is not clear how consultants determine if their trainees are competent to make diagnoses. For instance, when a consultant signs off a trainee to report certain cases, or completes a WBA suggesting a trainee is “meeting expectations”, how did they reach that decision? How did that consultant know that the trainee was diagnostically competent for that case? Is a trainee diagnostically competent just because they get the diagnosis right or is the indictor for this much less tangible?

Although summative assessments like FRCPath Parts 1 and 2 appear to contribute to the assessment of trainee diagnostic competence, these tests cannot test every facet of a trainee’s knowledge or diagnostic capability. Regular, similar tests having little presence during training. WBAs are utilised in histopathology to assess competence, but (as it will be seen), there are multiple issues related to WBAs and little evidence to support their use in histopathology.

As the review of the literature on competence suggests one may rightly argue that competence is far more complex than getting diagnosis correct. For example, it has been discussed that competence encompasses aspects of professionalism, communication, adaptability and consistency, to name but a few. Many of these qualities are also found in the RCPath training curricula, but how they relate to making a diagnosis is not explicit. The curriculum is more an overview of desirable pieces of knowledge, skill or attitude required of trainees. It is also not clear how complex these individual attributes are in histopathology or how they relate to one another. Indeed, competency-based frameworks have been criticised for over-simplifying competency (Rasmussen 1983). Furthermore, curricula do not always translate well into the real working environment where other factors may be more important (Merlin et al. 2014).

### 1.2.10.2 Diagnostic competence and accuracy

There appears to be some truth in the assumption that diagnostic competence is about making accurate diagnoses. For example, there are a number of studies in the literature which have examined diagnostic accuracy in histopathology, suggesting it is important to the profession (Shanteau et al. 2002; Parham et al. 2010). Often, these studies have compared “expert” diagnoses to trainee diagnoses, in order to assess competence.

There is no agreed method for identifying experts or an agreed standard against which experts can be assessed, but in reality, all trainees are trying to become experts in histopathology by the end of their training. Experts may have vast amounts of experience in a given field, numerous peer reviewed papers, excellent communication and self-confidence or be judged on their professional status, yet these characteristics do not necessarily equate to diagnostic expertise (Parham et al. 2010).

An objective measure of expert performance has been proposed by Shanteau et al which is based on the ratio of an individual’s discriminatory ability to their level of inconsistency (2002). This measures two parameters which are considered to reflect an expert’s ability: an ability to differentiate between similar but different entities along a continuum, and secondly, consistency of the individual at making those choices (Einhorn 1974). The measure of discrimination is divided by the degree of inconsistency, with this ratio giving a descriptive statistic as to the degree of expertise (Parham et al. 2010). A high degree of discrimination and little inconsistency gives a large ratio that is associated with a greater degree of expertise. To this end, one would expect that trainee levels of expertise would increase as they progress through training; their scores gradually becoming closer to the “expert” or specifically, the consultant score.

A small study examining whether it is possible to identify expertise in breast pathology using the CWS (Cochran-Weiss-Shanteau) ratio found that it was a useful tool in differentiating between trainees and consultants (Parham et al. 2010). In this study, 15 breast lesions were selected which reflected the spectrum of breast disease in routine practice. Each diagnosis was broadly categorised into one of four groups (benign, atypical, in-situ disease and invasive carcinoma). The CWS scores showed a statistical difference between average trainee score and consultant score. It also suggested that it is often difficult to rely on a “correct diagnosis” or “gold standard” when making comparisons. In some cases, experts may not agree on a diagnosis, or alternatively all the experts may agree but agree the wrong diagnosis. However, the CWS ratio does not require a correct diagnosis. The two parameters utilised examine the ability of a pathologist to discriminate between possible diagnostic entities and their consistency in undertaking this process. This CWS ratio (discrimination/consistency) obtained in this way is a measure of expertise (Parham et al. 2010). Discrimination can be shown by generating a wide variety of responses to individual cases; one can exhibit consistency by repeating the same response to all stimuli (Shanteau et al. 2002).

Aptitude testing has also been explored in histopathology although aptitude testing has yet to be adopted into specialist training as currently there in insufficient evidence to support its use (Cross 2005). In general pathology, the Objective Structured Practical Examination (OSPE) does help to assess aptitude for pathology as the medical and dental curricula contain little exposure to histology and most trainees enter histopathology training with no previous experience. There are several qualities or skills which have been recognised as important attributes to possess two of which are: visuospatial analysis and visuospatial memory (Cross 2005). Histopathology diagnoses are made by a mixture of pattern recognition and heuristic logic and this ability to recognise and recall visual patterns varies between individuals. Vigilance is also an important skill, as there are several tasks that require a methodical and thorough approach with a low error rate. These include checking that laboratory numbers on slides match those on request forms; ensuring one has viewed the entirety of the histological section and double-checking histology reports before they are sent out to clinicians. The assessment of risk is also relevant to histopathology. For example, making a decision on how many blocks to take from a cancer excision specimen is a routine task. Too few blocks may result in an important feature being missed; yet over-zealous sampling may be wasteful and time-consuming. A careful and sensitive programme of assessment of trainee progress is probably needed, so that trainees who lack these skills may be given extra support or advised to pursue an alternative career path. One of the most important end results of implementing these skills is making a diagnosis by light-microscopic examination. Several studies have tried to measure this in histopathology, which are outlined below (Cross et al. 2000; Cross, Dennis and Start 2002).

In one study, observers were asked to examine 100 slides of colorectal polyps, 50 of which were adenomatous and 50 hyperplastic and then asked to classify them into one of these two diagnostic categories (Cross et al. 2000). Results from experienced histopathologists showed a very high level of inter-observer agreement with kappa statistics very close to 1 (perfect agreement). A trainee early on in training also completed this test. Strong learning behaviour took place because initially her performance had a kappa statistic of -0.3 on the first 20 cases but rose to 0.8 on the last 20 cases even though no tuition or support was given. The authors suggested this improvement was through self-discovery and indicated a strong aptitude for histopathology.

In another study, participants were asked to diagnose breast needle biopsies on a continuous linear scale from 100% certainty of a benign diagnosis through to 100% certainty of a malignant diagnosis (Cross, Dennis and Start 2002). Values were then plotted as a receiver operating characteristic (ROC) curve, and the area underneath this curve was taken as measure of overall performance. In this study, trainees at the start of their training were given the slides (in randomised order) at two points separated by a number of months, and two ROC curves were plotted. All the trainees in the study showed a sharp increase in performance which was most marked over the earliest part of their training (Cross, Dennis and Start 2002). These findings suggested that the trainees learnt in the intervening months and became more accurate at diagnosis as a result.

Gaining a better understanding of what cognitive processes are involved in diagnosis, how these vary from one individual to another, and which ones are the most effective in classification could provide the pathologist with more robust tools for improving routine diagnostic decision-making and clear benefits for patients who fall foul of variable diagnostic classification methods (Hamilton et al. 2009). Poor reproducibility in diagnosis is well documented in the literature, for instance in breast and in prostate grading, which may be partly attributable to the fact that different experienced pathologists favour different features when arriving at a given diagnosis (Dalton, Page and Dupont 1994; Allsbrook et al. 2001). For example, the study by Fandel and colleagues found that interpretative bias played a significant role in diagnosis, where architectural features in prostate neoplasia impacted greatly on final diagnosis, regardless of the nuclear morphology (2008). An approach to try and disentangle the understanding of diagnostic processing has been explored using Bayesian belief networks. Here, a pathologist has to break down the decision sequence into a set of definitive steps, each of which is defined by its contribution to the final decision, and where a final diagnostic probability is based on the accumulation of these diagnostic steps (Diamond et al. 2002; Morrison et al. 2002). Some pathologists argue that this stepwise approach is limiting and pathologists should be taught to recognise patterns in a Gestaltian fashion and match these to diagnostic categories (Hamilton et al. 2009). However, these steps can be very useful to trainees as it provides a sound, logical sequence to follow and learn from.

The problem with measuring expertise, aptitude or accuracy is that none of them necessarily equate to a “diagnostic competence” as this construct likely encompasses many more elements, as previously discussed (in sections 1.2.3 and 1.2.4). The work of Shanteau and colleagues (2002) in measuring expertise gives scores which are merely a comparison with the “expert” opinion and don’t offer an insight into a trainee’s overall competence. Even though these tests have utility as an objective measure of light-microscopic diagnostic skills, it is worth noting that other important skills such a communication, professionalism and overall ability to manage a case are not assessed using these methods.

If we look at the study examining breast needle biopsies (Parham et al. 2010), this test is related to a fairly specific area in histopathology. For instance, breast needle biopsies involve looking only at cells, whereas interpretation of a breast tumour resection involves examining the cells, but also its overall morphology, its borders, whether it shows perineural or lymphovascular invasion, if it is mucous producing and so on. The ability to test diagnostic competence over a much more variable selection of cases is clearly far more challenging. Also, being able to determine if a lesion is benign or malignant, although extremely important, is not the only task for a histopathologist –a specific diagnosis if often required e.g. is it a ductal carcinoma, lobular carcinoma or adenoid cystic carcinoma? If this idea is extrapolated to the huge number of entities a trainee must eventually be able to diagnose, designing tests that are adaptable and robust enough to assess this, poses a significant obstacle. Although the final diagnosis is important, the steps taken (or interpretation) in reaching the diagnosis could be viewed as being just as important, which relates back to these tests missing the “bigger picture”. Further issues which may arise if these tools were used in training relates to their applicability and practicality as the financial and time costs of designing these tools, updating them and analysing the results would be an onerous task.

### 1.2.10.3 Diagnostic competence and the process of making a diagnosis

A few authors have tried to explain the steps involved when making diagnoses (Bussolati 2006; Pena and de Andrade-Filho 2009). For instance, Pena and de Andrade Filho suggest the process of diagnosis can be split into four domains: cognitive, communicative, normative and medical conduct (2009). Essentially, the cognitive domain relates to mentally processing what is on a slide; the communicative domain involves conveying the diagnosis with due regard for clinical information; the normative domain involves being aware of what is expected when diagnosing lesions, such as having recognised ways to grade and stage tumours; and the medical conduct domain largely involves aspects of professionalism and considering the appropriateness of ordering further tests or consulting other pathologists.

However, these studies are not based on any evidence but are merely suggestions and commentaries. Therefore, it is not possible to draw solid conclusion from them. Secondly, the steps involved in making a diagnosis do not necessarily equate to competence, as it is probably just one aspect of competency, similar to the previous studies which just examined accuracy of diagnoses. Thirdly, even if the approach to diagnosis was known, it is not apparent which steps are most important to consultants when assessing competence in their trainees. Nonetheless, these studies do appear to describe quite a complex mental process and acknowledge that there are no current studies which clarify the matter any further (Bussolati 2006; Pena and de Andrade-Filho 2009).

### 1.2.10.4 Diagnostic competence and independent reporting

A single study was identified which used a focus group with trainees and qualitative interviews with consultants to understand the perceived learning needs for graded responsibility in histopathology (Finall and Allery 2016). In order for independent reporting to occur, the data suggested six things were important: “accumulation of experience” through seeing around 100 cases per week; “attending multi-disciplinary team meetings” to review complex cases and develop communication skills; close “consultant supervision” to help trainees prepare for independent reporting; “diagnostic acumen” in terms of trainees knowing the implications of certain diagnoses on patient care; “recognising normal” which involved trainees needing to show they could distinguish normal from abnormal and also the range of appearances for certain tumours; “recognising pitfalls” in terms of trainees being worried about certain entities they had not yet encountered and might miss. A criticism of this paper would be that it was a small study without data saturation for the qualitative interviews. Furthermore, the data from the focus groups and interviews stemmed from a single pathology centre and may not reflect wider opinion on the topic. However, in terms of diagnostic competence, one might argue that a consultant agreeing to a trainee independently reporting is a valid indicator of that trainee’s diagnostic competence. There may be aspects detailed in this study which relate to competence, but the study sought to understand the learning needs for independent reporting to occur, not the qualities required for it to be handed to a trainee. Hence, the six areas described do not describe the knowledge, skills and attitudes relevant to being diagnostically competent but actually reflect the need for trainees to generate significant experience before it is appropriate for them to act independently. The study also suggests that it is not just the ability to formulate diagnoses which leads to independent reporting, but there are wider issues such as communicating diagnoses and understanding the implications of certain diagnoses.

### 1.2.10.5 Conclusion

Currently, although it appears aspects of accuracy and sound process are part of being diagnostically competent, the literature does not provide a clear description of what diagnostic competence is, how it is judged in trainees and the contexts relevant to the judgement. Despite this, histopathology trainees are required to complete WBAs during training as a marker of their competency. The next section will discuss how WBAs came to be part of training curricula and the current issues with their use. However, what should already be apparent is that it unlikely they would be able to identify diagnostic competence if the literature is unable to define and describe the knowledge, skills and behaviours relevant to this construct.

# **1.3 Workplace-based assessments**

## 1.3.1 Introduction

WBAs are used in a variety of workplaces, but are used widely within medicine and dentistry as an assessment of trainee progress, as recommended by the GDC and GMC. They have been defined as the “assessment of day-to-day practices undertaken in the working environment” (Miller and Archer, 2010). Before this, specialist trainees were assessed using local assessment tools, requiring review panels to be aware of regional differences in assessment of trainees and their possible subjectivity (Wragg et al. 2003). Alongside the outcomes from professional exams, input from Educational Supervisors (ES), Training Programme Directors (TPD) and representatives from the local deanery, this information is collated and reviewed to inform the outcome for each trainee at the Annual Review of Competency Progression (ARCP). This meeting determines whether a trainee has reached a prescribed standard to progress to the next year of training and the ARCP panel may issue objectives to those trainees who have failed to reach this standard. The ARCP panel is ultimately responsible for termination of the trainee contract if improvements are not made, but should also consider and resolve any issues which the trainee may raise.

In general histopathology and oral and maxillofacial pathology (OMFP), a minimum of 18 WBAs with satisfactory outcome are expected to have been completed by the trainee at the start of training. Thereafter, the number of WBAs required at each stage of training and the proportion which should be self-directed varies between histopathology specialities.

## 1.3.2 Types and format of workplace-based assessment in histopathology

WBAs in histopathology include Direct Observation of Practical Skills (DOPS), Case-based Discussions (CbDs) and Evaluation of Clinical Events (ECEs). In addition, three Multi-Source Feedbacks (MSFs) are required of histopathology trainees over the course of their training. A description of these different WBAs is given in Table 1.1

|  |  |
| --- | --- |
| **Tool** | **Description** |
| Direct observation of practical skills (DOPS) | A trainee is observed performing a practical task in their clinical environment e.g. macroscopic assessment of pathology specimen. |
| Case-based discussion (CbD) | A trainee presents and discusses a case they have worked to assess depth understanding, clinic-pathological correlation and ethical implications e.g. discussion around histopathology and prognostic factors of melanoma |
| Evaluation of Clinical Events (ECEs) | A trainee is assessed on their performance dealing with a complex task or event, often including aspects of teamwork and communication e.g. presentation of a case at a multidisciplinary team meeting. |
| 360o multi-source feedback (MSF) | A trainee selects several co-workers who are tasked with offering a judgement on the trainee’s performance. |

**Table 1.1. The different types of WBA used in histopathology training**

For DOPS, CbDs and ECEs, trainees are graded on a scale of 1-6 against numerous parameters. A score of 1 or 2 indicates the trainee is “below expectations”, 3 “borderline”, 4 “meets expectations” and a score of 5 or 6 indicates a trainee is “above expectations” for the stage of training. The assessor should then discuss the outcome with the trainee and give feedback on the event. The result of an MSF is fed back to the trainee, usually as a form of mean score for a particular characteristic and an overall group mean score (the mean score for all trainees in the sample group for that circulation of the MSF).

WBAs are meant to document trainee progress by assessing a range of personal and professional qualities and suggest areas for improvement (Wilkinson et al. 2008). Each WBA should map to the domains of the curriculum, as approved by the relevant regulatory body, which appears to be the case in histopathology. WBAs are completed by the overseeing consultant histopathologist or an integral member of the pathology team, including laboratory technicians and biomedical scientists. The majority of cases to be assessed are selected prospectively by the trainee but some are chosen by the assessor with input from the educational supervisor.

WBAs may be completed in paper format and stored in a personal development portfolio in readiness for the ARCP. In many specialties, however, WBAs can be completed online, accessed by the trainee and sent to the assessor to complete. In general histopathology, the Learning Environment for Pathology Trainees (LEPT) and the RCPath electronic training portfolio serve this purpose. For the majority of medical and dental specialties, the online environment for utilizing WBAs is called the Intercollegiate Surgical Curriculum Programme (ISCP). This website houses the curricula for these specialties and a secure area which contains trainees’ web-based portfolios, WBAs and guidance on completing WBAs.

## 1.3.3 Regulation of workplace-based assessments

On its website, the RCPath stipulate the required standards of WBAs, their uses and definitions within pathology (The Royal College of Pathologists 2015; The Royal College of Pathologists 2018).

WBAs have been mandatory for all trainees in the UK appointed since August 2007 (Department of Health 2007) . The Postgraduate Medical Education and Training Board (PMETB) was originally responsible for setting the standards of post-graduate medical education and training which were developed in conjunction with the Academy of Medical Royal Colleges (AoMRC) in 2009. In 2010 PMETB merged with the GMC and the standards were updated and are still in use today (General Medical Council 2010).

The RCPath is responsible for organising the FRCPath Part 1 and Part 2 examinations-the two main summative assessments which every trainee has to pass to reach consultant level. However, a trainee must also complete a UK training programme with satisfactory ARCP outcomes, be awarded a CCST and be entered on to the GMC/GDC Specialist Register. Alongside passing these exams, a trainee must be able to have documentary evidence to show they have done the minimum number of WBAs to the required standard each year. The Part 1 examination is a knowledge-based test either written or multiple choice format, depending on whether the candidate is a general or OMFP trainee. The Part 2 exam is a test of practical and diagnostic skills, examining the candidate’s ability to apply theoretical knowledge in a clinical context. Year 1 trainees in general pathology also sit the OSPE, which determines at an early stage if the trainee has sufficient aptitude for the specialty. The OSPE must be passed to progress with training.

## 1.3.4 The need for workplace-based assessment in healthcare training

While summative assessments focus more on trainee aptitude and ability to become a sound clinician, there is also expectation that trainees are responsible, ethical and patient-focused individuals with excellent communication and interpersonal skills (General Medical Council 2009).

This has stemmed from a number of high-profile cases of malpractice in the medical profession which have resulted in more pressure on medical and dental schools and the Royal Colleges to produce graduates and specialists who are well-rounded individuals (Teasdale 2002). WBAs are one of the primary tools used to assess this, as researchers have found that artificial exam situations do not always predict how a trainee may act in real-life situations (Rethans et al. 2002). The use of OSCEs has been proven to reflect poorly with how trainees perform in the workplace (Rethans et al. 2002). If WBAs can be used to ascertain if a trainee has reached the desired level of competence, one might argue that it should be proven that they can identify struggling trainees, address learning needs and foster future development. However, designing WBAs which are thorough, feasible and valued by the trainees and trainers is challenging (Prescott et al. 2002).

Ideally, WBAs should encourage trainees to reflect on their skills and behaviours and create an ethos for life-long learning and development. During training, a balance of formative and summative assessments should be utilised, ensuring trainees are competent, safe and able to work independently. An expert advisory panel for the UK’s Independent Inquiry into “Modernising Medical Careers” considered assessment methods used during medical training, acknowledging that it is sometimes unclear when assessment is “aimed at supporting self-improvement and remediation and when it is regulatory and summative”, highlighting the fact that it can be ambiguous if WBAs should be used in a formative or summative manner (Tooke 2008).The GMC has stated that WBAs are a summative tool used to assess “competency”, but the literature would suggest that their effectiveness, educational value and numbers required to demonstrate competency, requires further evaluation (Bindal, Wall and Goodyear 2011; General Medical Council 2011).

Defining underperformance is a concept which trainers may feel they can judge but is difficult to define in words. It has been described as a trainee “requiring intervention beyond the normal level of supervisor-trainee interaction” and may be at risk from subjectivity if current assessment tools are not working effectively (Mitchell et al. 2013). One study found that 16.6% of struggling trainees were identified using WBAs while the remainder were merely identified by trainer observations (Black and Welch 2009). Therefore, it may be that WBAs serve little function in identifying these individuals, but they may help to formalise the process of identification and address learning needs early on. Another study revealed an association between training difficulties and lower mean scores on WBAs (Mitchell et al. 2011). However, it was highlighted that WBAs shouldn’t be used to certify trainees and their real value is their ability to provide feedback. They should also be used in conjunction with other assessment methods to be representative of trainee competence (Mitchell et al. 2011).

## 1.3.5 Perception of workplace-based assessments in medical education in general

### 1.3.5.1 Introduction

There are mixed opinions in the literature regarding the perception of WBAs, which probably reflects that there are different assessors, WBA and clinical contexts to take into account within each individual speciality. This means that it is important for each speciality to ensure its own WBAs are fit for purpose. It is also important to note that some of the evidence regarding WBAs stems from qualitative research and certain research outcomes may be limited by their lack of generalizability. Furthermore, some studies recruited participants by voluntary means, which may select a motivated and biased sample, affecting the internal and external validity of the results (Sargeant et al. 2003; Pelgrim et al. 2012). However, selecting the correct methodology in medical and dental education is difficult sometimes, and there are different benefits of using the positivist (deductive reasoning) and interpretivist (inductive reasoning) approaches (Bullock 2010).

### 1.3.1.2 Feedback

A key component of WBAs is meant to be for the assessor to give feedback to the trainee about areas which they may do well in and other areas which may need improvement. This information can be used to guide the trainee and nurture future development (Wood 2010). Research on WBAs in dental training have found WBAs to be clear, comprehensive and able to identify trainee difficulties, increasing trainee confidence and highlighting areas of weakness. (Grieveson et al. 2011). However, a larger systematic review of WBAs in medicine has shown there is little evidence to suggest WBAs promote feedback, encourage learning or improve trainee performance (Miller and Archer 2010). A trainee who is unprofessional but has superior knowledge is not the same as another trainee who is professional but has an inferior knowledge base to his peers. This is why WBAs may be perceived as giving a superficial overall judgement of a trainee, when profile differences should actually be explored and interpreted (Crossley 2013). In orthodontics, a recent questionnaire-based study concluded that WBAs were acceptable to trainers and trainees (Sia and Harrison 2015). However, doubts were raised regarding whether they are used correctly and if they accurately reflect trainee progress (Sia and Harrison 2015) . Again, the aforementioned studies reflect the ambiguity in the literature regarding the perceived usefulness of WBAs. The lack of a clear consensus may stem from the nature of these studies. For instance, Miller and Archer (2010) included 16 papers in their systematic review, but half of their papers related to MSFs while the remainder were a mixture of DOPS, Clinical Evaluation Exercises (Mini-CEX) and other combinations of tools. Furthermore, this review examined WBAs across the whole of medicine in which there are numerous specialties to consider. In contrast, Grieveson’s conclusions stemmed from WBAs used in dental foundation training and not specialty training (2011). Here, the WBAs were also completely different, comprising Dental Evaluation of Performance (D-EPs), CBDs and Patients' Assessment Questionnaires (PAQs) rendering the results non-comparable. Finally, the study from Sia and Harrison looked at a specific group of individuals; orthodontic trainees and trainers (2015). The authors determined agreement with statements using a 5-point Likert type scale, while Grieveson’s study used a dichotomous scale of “agree” or “disagree”. These differences in study design mean that other specialties must be careful when interpreting results or extrapolating findings to different medical contexts.

Feedback should be tailored to the individual and given immediately after the encounter to be most valuable (Nicol and Macfarlane‐Dick 2006). However, trainees may be reluctant to change their behaviour and trainees have to feel that there are sufficient mechanisms in place to nurture their development (Miller and Archer 2010). Studies exploring why feedback may have limited impact on a trainee and how the trainee processes the feedback is a growing area of interest (Bindal, Wall and Goodyear 2011; Kogan et al. 2012). To this end, trainers may feel they have insufficient time and expertise to foster development in their trainees and trainees sometimes view WBAs as a chore with little respect for their importance (Bindal, Wall and Goodyear 2011; Kogan et al. 2012). One systematic review concluded that:

Considering the emphasis placed on workplace-based assessment as a method of formative assessment, there are few published articles exploring its impact on doctors’ education and performance (Miller and Archer 2010, p. 1).

Clearly, user acceptability and engagement is also essential to the success of any education innovation and the lack of engagement has been suggested to stem from a perceived lack of purpose for WBAs, time constraints and addressing the quality of feedback (Massie and Ali 2015).

### 1.3.1.3 Acceptability

One study suggested that WBAs can also become tiresome for trainees (Pereira and Dean 2009) which may result in trainees selecting cases retrospectively when a suitable period of “free time” arises. This study utilised surgical trainees and consultants to garner opinion, with 49% of respondents rating the assessments tools as “poor” or “very poor”. This negative response may have stemmed from reduced training hours and the bureaucratic burden of having to update an online portfolio (Pereira and Dean 2009). However, a relatively recent study from 2017 examined the perception of procedure-based assessments with surgical trainees who did acknowledge that these tools can hold significant value when used correctly (Shalhoub, Marshall and Ippolito 2017). The study also identified current barriers to this including a lack of support from seniors and inappropriate selection of cases (Shalhoub Marshall and Ippolito 2017). This sentiment is mirrored in a study conducted with foundation programme doctors, where 60% disagreed with the statement that generating an e-portfolio of WBAs “created a positive learning experience” (McKavanagh, Smyth and Carragher 2012). One of the largest studies examining the impact of foundation training on trainees surveyed 1065 foundation doctors in the UK (Dean and Duggleby 2013). With respect to WBAs, the study identified that 76% of respondents felt the ability of these tools to reflect trainee ability was “poor” or “very poor”. Although the number of trainees surveyed was large, it is important to note that the response rate was only 5% and the results may therefore suffer from responder bias. However, the negative perceptions noted in the aforementioned studies might stem from the pressures of working long hours, which is particularly relevant to foundation doctors and surgical trainees. For example, both doctors in training and trainers feel the European Working Time Directive (EWTD) is continuing to have a significant negative impact upon training (Wakeling et al. 2011). This might be another reason why WBAs tend to be more accepted in dentistry (Grieveson et al. 2011; Sia andHarrison 2015) where the working hours are more sociable.

### 1.3.1.4 Utilisation

The perceived time-consuming nature of WBAs can reduce enthusiasm for completing them and a reduction in quality feedback. Trainees are meant to have WBAs completed regularly through the year, yet foundation trainees submit 40% of their WBAs in the last six weeks of their first year, perhaps because they feel they will attain higher scores towards the latter end of the year (Davies et al. 2009). DOPS can be particularly stressful for trainees as it can be intimidating to have your clinical skills closely observed (Mitchell et al. 2011; Pelgrim et al. 2012). It is also possible for trainees to “select” cases for discussion that they prefer and avoid difficult cases (Mitchell et al. 2011), so it important that educational supervisors ensure that each trainee has been thoroughly assessed, regularly, over a range of cases and by different assessors (Norcini et al. 2003). Trainees who know they might be underperforming are more likely to approach a nursing colleague or non-clinical assessor to complete a WBA, perhaps trying to seek a more favourable outcome for their assessment which more senior colleagues may not provide (Mitchell et al. 2013; Barrett et al. 2016). It is imperative to monitor WBAs and to ensure trainees do not manipulate who and on what they are assessed on (Murphy et al. 2009). On the other extreme, trainees who always perform well on WBAs may feel that their feedback is of limited use-compounding the idea that WBAs can become a “tick box” exercise for these trainees.

Recent work has discussed the importance of self-regulation theory, which includes “goal-directed behaviour, use of specific strategies to attain goals, and the adaptation and modification to behaviours or strategies to optimise learning and performance” (Sandars and Cleary 2011). This research has shown that high- and low-achieving learners use self-regulation differently, and if it were possible to help struggling trainees to self-regulate more effectively, then their performance may improve (Sandars and Cleary 2011).

### 1.3.1.5 Validity

Although there is some support for the utility of WBAs, problems surrounding their validity due to assessor scoring has been problematic (Hawkins et al. 2010; Pelgrim et al. 2011) . Clearly, if only one trainer assesses the trainee for every task they do, the assessment could be vulnerable to bias and suffer from a lack of validity and generalisability (Dauphinee 1995; Epstein and Hundert 2002). WBAs may require the assessor to judge whether a trainee “meets expectations”, which is inevitably at risk from subjective influences (Alexander 1996; Chapman 1998). In the literature however, there is debate regarding whether this perceived subjectivity unduly affects assessments of trainee performance. While Alexander (1996) found that assessor judgements may subjectively influence assessment grades other researchers who employed generalisability theory found that trainee knowledge and tasks sampled had a much greater impact on assessment score (Friedman and Mennin 1991; Govaerts, van der Vleuten and Schuwirth 2002; Keen, Klein and Alexander 2003). MSF have been shown to lead to performance improvement and are particularly useful tools for assessing interpersonal skills, communication, professionalism and team work (Lockyer 2003). However, there is little evidence that alternative WBAs, such as DOPS, lead to improvement, although subjective reports on their educational impact are positive (Miller and Archer 2010). Also, a large study in 2009 found that “consultants and sisters (senior nurses) were more likely to raise concern in MSFs about junior doctors compared to managers, fellow Senior House Officers (SHOs) and junior colleagues” (Bullock et al. 2009) indicating that MSFs also have their limitations, especially if the trainee can select the participants for their MSF. Other studies addressing subjectivity suggest some trainers feel guilty about giving trainees low scores, while others compensate with overly high scores for difficult cases (Norcini et al. 2003; Crossley et al. 2011). Staff can be reluctant to document unprofessional behaviour for various reasons, sometimes because there is internal conflict between their mentor and assessor roles (Boenink et al. 2005; Van Mook et al. 2007). Certain interaction effects, such as gender, have been reported, where a particular assessor consistently gives better scores to either males or females (Schmittdiel et al. 2000). All these aforementioned factors will undermine the assessment’s validity. Furthermore, if trainees become aware that their assessments may be subjective and influenced by bias, resentment is likely to manifest and undermine their perception of using WBAs.

### 1.3.1.6 Reliability

Trainers often feel burdened by WBAs, compounded by the number of WBAs required of each trainee and the multiple trainees each trainer is meant to assess. The quotas for WBAs are meant to increase reliability as outcomes can be triangulated, but given the increasing pressures on the NHS, it is understandable that certain trainers may feel that patient care should be prioritised before WBAs. Research in surgical specialties has already questioned the prescribed number of WBAs for each trainee, suggesting lower quotas may be just as reliable for determining trainee aptitude in a given area (Beard 2008; Marriott et al. 2011). Recent work by Crossley has identified that conventional scales in WBAs are felt to be invalid and construct alignment reduces assessor disagreement and increases assessor discrimination (Crossley et al. 2011). Reliability modelling showed that by using these new scales, the number of assessors required to achieve a generalisability coefficient ≥ 0.7 fell from eight to three for CBDs (Crossley et al. 2011). Generalisability theory may be applied to measure the contributions that all relevant factors make to the result (observer, situation, case, assessor and their interactions) (Crossley et al. 2002).

### 1.3.1.7 Guidance and training

Studies have shown that appropriate guidance for assessors was essential for successful WBA implementation and comprehensive training is required for both trainees and trainers (Ryland et al. 2006; Kirton et al. 2013). This may be particularly important when assessors have had limited training on completing WBAs and are unsure about how to provide useful feedback. Assessors could be calibrated to improve inter-rater reliability scoring to differentiate between various performance levels (Davies et al. 2009; Kogan, Holmboe and Hauer 2009; Crossley et al. 2011). For example, the Mini-CEX is often quoted as being subject to significant assessor error, in particular both “hawk/dove” error and error caused by assessors’ differing views of what constitutes an “average” performance (Norcini et al.1997; Holmboe et al. 2003; Wilkinson et al. 2008). Nevertheless, a randomised controlled trial concluded that there was insufficient evidence that ‘calibration training’ reduced the variability in judges’ interpretations of performance, but there is evidence that ‘social-cultural’ interventions such as discussion of the merits of certain examples of performance lead to more shared interpretations of performance (Holmboe, Hawkins and Huot 2004).

### 1.3.1.8 Changes to WBAs

A possible solution to increase enthusiasm has included introducing WBAs into undergraduate education to instill their educational benefit and to foster a positive perception that may continue into post-graduate training. However, the perception of WBAs in medical trainees is also reported to be negative, suggesting that even before graduation, trainees are sceptical about the value of WBAs (Ali and Goh 2017). Furthermore, WBAs have been rebranded in some medical fields as “assessment of performance” (AOPs) and “supervised learning events” (SLEs) to try and overcome the conflict between their summative and formative functions (Burr, Wade and Watts 2018). For SLEs, there are text boxes for written feedback and all the tick-boxing elements removed to encourage formative assessment and feedback. AOPs are intended to be summative and measure competence in trainees and as such may still be used as evidence at the ARCP unlike SLEs. The long-term impact of SLEs requires research but recent publications suggest there may be hesitation in embracing them as educational tools (Rees et al. 2014).

Financial incentives and providing dedicated time for trainers to complete WBAs have also been suggested as possible ways to improve the feasibility of WBA (Hauer, Holmboe and Kogan 2011; Massie and Ali 2015). However, if life-long learning, collaboration and self-regulation are going to flourish in medical and dental education, it may be argued that trainers must embrace a culture which supports trainees in achieving these objectives. Financial incentives may give short-lived gratification to trainers and not translate into improved educational benefits to trainees.

### 1.3.1.9 Conclusion

The current drive for outcome-based medical educational can lean too heavily on the ability to produce reliable and measurable outcomes. There is a danger that the complexities of clinical practice are not adequately captured by WBAs as the individual elements which form sound clinical performance are not best viewed in a piecemeal fashion (Frank et al. 2010a; ten Cate and Billett 2014). A recent study suggests that observing clinical performance over time is more valuable than assessments that focus on single encounters (Oerlemans et al. 2017)

As it stands, WBAs are the main tool used in healthcare to assess performance in the workplace. There is patchy evidence to support whether the current WBAs provide useful feedback, when are the best times to use them and how they could be used most effectively, emphasising a need for more research to explore WBAs in medical and dental training (van der Vleuten and Schuwirth 2005; Kogan, Holmboe and Hauer 2009; Miller and Archer 2010). In particular the evidence stems mostly from questionnaire data where response rates and responder bias may not accurately reflect the opinion of everyone involved and also from qualitative data which is not intended to be generalised. The variable contexts in which WBAs are used and the fact that different tools are used in different specialties confounds the ability to draw clear conclusions from the literature.

## 1.3.6 Perception of workplace-based assessment in histopathology

The evidence base regarding the perception and use of WBAs in histopathology is even more limited than the literature on WBAs in medical education in general. However, the outcomes of WBAs are still held in the same high regard during training and may be used at the ARCP to identify struggling trainees.

A single study was found which explored the experiences and perceptions of trainers in using the DOPS tool with histopathology trainees using a questionnaire-based approach (Finall 2012). The study supported the use of DOPS in histopathology but raised concerns about the impact on trainer time, whether DOPS is used in a formative manner and concerns about the amount of guidance provided to trainers (Finall 2012). This study suffered several limitations including a small sample of only seven histopathology consultants (a response rate of 46%), it did not explore trainee perceptions and it only used questionnaires to gather information. It missed an opportunity to use qualitative methods to explore the themes it raised. However, these initial studies are still valuable and can help to inform future studies and research directions (Silverman 2006).

In most histopathology departments, cases must be checked by a consultant after a trainee has written a provisional report. It could be argued that histopathology trainees may be exposed to a lot of one-on-one teaching when these routine cases are examined on the multi-headed microscope. Often, trainees can learn and be guided by their consultant’s superior experience during these regular episodes of discussion, even if they are not formally recognised by a WBA. This one-on-one time is an ideal environment for low-stakes assessment and constructive feedback (Rekman et al. 2016). These regular episodes of interaction also have the potential to identify struggling trainees.

Those who feel disheartened by the number of WBAs they have to complete may feel this regular supervision renders a discussion regarding how many WBAs are required of histopathology trainees. If reduced numbers of WBAs are needed, this would reduce the administrative burden of WBAs and as a result, possibly increase levels of enthusiasm.

With increasing pressures on academic and NHS consultants in histopathology, it is vital that WBAs are not viewed as a tick-box exercise, becoming tedious and undervalued. If trainers and trainees place too much emphasis on major exams, such as FRCPath parts 1 and 2, there may be fewer opportunities to address trainee learning needs and encourage life-long learning, perhaps resulting in individuals who become exam-focussed and show little regard for developing their skills following qualification. Due to the academic nature of histopathology and the ability to report cases alone in one’s office, this could lead to professional isolation. Histopathology is a complex specialty and helping trainees to identify their strengths and weaknesses is vital so that individuals recognise when they have reached the limit of their diagnostic capability and have the confidence to refer a case to another expert. Second opinion reporting in head and neck pathology, for instance, has been shown to be particularly important (Mullin, Brierley and Speight 2015).

As such, significantly more research is needed to assess WBAs in histopathology. For example, CbDs, ECEs and MSF have yet to be examined in a histopathology setting and there is not a single study which explores in detail the uses, merits, limitations and attitudes towards any WBA in histopathology. This is particularly important when one considers that histopathology training is quite different to the majority of other specialities in medicine and dentistry, specialties which the majority of literature regarding WBAs relates to. Histopathology trainees do not generally see patients and their day-to-day activities include practical cut-up of surgical specimens and examining histology sections under a microscope, the latter, in particular, requiring high levels of concentration and a unique set of skills. A consultant histopathologist carries a large burden of accountability, being responsible for diagnosing if a lesion is benign or malignant and thus determining the type of treatment a patient may receive. Clearly, it is paramount that any assessment tools used in histopathology training are fit for purpose, tools which may ultimately minimise risk to patients by ensuring newly qualified consultants have progressed through a robust, thorough and holistic educational process.

## 1.3.7 Summary

Assessment is frequently used in the training of medical trainees to assess competence. Most importantly, an assessment must be valid otherwise it could potentially lead to erroneous decisions on the development of competency in that individual. Currently, WBAs are used widely in healthcare to assess competency. They are recommended throughout the histopathology curriculum and are the most readily utilised tool in training to assess competence. Given the tensions that exist between service delivery and training, high-quality learning opportunities and assessment are essential to ensure successful completion of training and to maintain high standard of care in the NHS (The Royal College of Physicians 2018).

WBAs have faced a number of criticisms from many studies in the wider healthcare literature. In histopathology, there is little literature to support the use of these tools and it is not known if they are capable of measuring competence. This is compounded by the fact that there is not a clear description in the training curricula regarding what constitutes diagnostic competence and what factors should be taken into account. Despite these uncertainties, it is clear that WBAs are mandatory during histopathology training and ask consultants to make competency judgements of their trainees. This is especially important given that histopathology is a unique speciality within medicine. Trainees do not directly interact with patients and much of their work involves unobservable mentation when they make diagnoses. This is in contrast to the practical elements more readily observed in a surgeon’s work or through the interaction between a physician and his patient. It is therefore imperative to ensure the current WBAs in histopathology are fit for purpose, as the general literature on WBAs may be less relevant to histopathology.

Previous studies have utilised aptitude tests and comparisons with experts to gauge competence in diagnostic histopathology (Cross et al. 2000; Cross, Dennis and Start 2002; Shanteau et al. 2002; Parham et al. 2010). Although these studies offer some insight into diagnostic competence, the wider literature on competence suggests that looking at diagnosis alone is unlikely to reflect the full judgement ecology. Furthermore, studies have examined diagnosis in isolation and not how consultants judge diagnostic competency in their trainees. This is important to understand as it is these judgements that could determine whether a trainee is progressing appropriately through the training programme or if trainees are able to qualify without actually being diagnostically competent. Indeed, it has been suggested that the most important priority in dental education is the role of assessment in identifying competence (Ajjawi et al. 2017).

It is therefore imperative to ask histopathology trainees and consultant about their perception of the current WBAs and particularly if they are capable of measuring diagnostic competence. This may identify the possible strengths and weakness of WBAs in histopathology, while also addressing a large void in knowledge in the healthcare literature. Following this, it is important to understand how consultants determine diagnostic competence in trainees. This will help to understand the knowledge, skills and behaviours that are most relevant to histopathology as well as providing further insight into the possible issues with the current WBAs. To this end, a systematic review from 2017 examining the Mini-CEX tool, concluded that despite current theories and research exploring judgement in assessment, issues of bias and rater idiosyncrasy are not fully resolved (Lee, Brain and Martin 2017) and the solution lies in judges providing specific and in-depth narrative to justify their ratings to provide insight into what is valued by the profession. The underlying qualities and the nature in which the information is used to inform judgements may help to guide the design of future assessment tools and ensure diagnostic competence is being properly assessed.

# **Chapter 2**

# **Aims and Objectives**

**Chapter 2: Aims and Objectives**

## 2.1 Introduction

The literature review demonstrated that the current assessment systems in histopathology may have limitations and there is little evidence to support the use of WBAs. In particular, it is not clear what diagnostic competence is, how consultants judge this quality in their trainees and whether WBAs are a useful tool to measure it. As described at the end of the previous chapter, it is important to first understand the current perception of WBAs to see if problems with their use currently exist. Next, the cues (knowledge, skills, behaviours) relevant to diagnostic competence and how consultants judge this in their trainees should be understood. Finally, the qualities identified should be triangulated with the stakeholders in pathology training to ensure they are valid and acceptable indicators of diagnostic competence.

## 2.2 Aims and Objectives

To understand how diagnostic competence in histopathology trainees is determined

1. To explore the perceptions of WBAs in histopathology by assessors and trainees using a questionnaire to understand:
2. how WBAs are perceived in histopathology including a comparison to the themes identified in the healthcare literature: their perceived usefulness, validity and acceptance.
3. if WBAs in histopathology are perceived to measure diagnostic competence
4. To explore diagnostic competence in histopathology by means of a qualitative interview study using Framework Method to understand:
5. what diagnostic competence encompasses in histopathology
6. the cues consultants look for in their trainees when determining diagnostic competence
7. the judgement ecology which pertains to determining diagnostic competency in trainees
8. To use a Delphi study to triangulate the findings from the qualitative interviews in assessing diagnostic competence to:
9. determine if the cues identified from the qualitative interviews are valued by experts.
10. determine the relative importance of individual cues
11. identify any additional cues

## 2.3 Emergence of a framework

There are numerous theories which may be used to frame our research aims and objectives, such as experiential learning (Kolb 1984) and self-regulation theory (Sandars and Cleary 2011). However, three appear particularly relevant to this project which are Critical Theory, Situated Learning Theory and Social Judgement Theory. These are discussed further hereafter.

Initially, Critical Theory was examined which is concerned with power distribution and how groups of individuals or subsets of society can affect, control or oppress other groups (Tyson 2006).

It could be argued that WBAs were born out of a necessity to satisfy governing bodies such as the GMC/GDC and Royal Colleges, without due regard for the implications and educational merit of their use. To this end, educators may question why WBAs are mandatory for trainees and why there has been a relentless push towards competency based assessment and increasing numbers of assessments. Is it for the educational benefit of the trainee or a result of a dominant force that has influenced the education of trainees out of a desire to control and regulate? Critical theory would ask that the source of power is identified to try to understand what has influenced the decisions that have been made. Healthcare is impacted greatly by political policy and socioeconomic reforms, and as a result, the training of trainees and the “end product” of their education is strongly affected by what the regulatory bodies deem important at that time.

While Critical theory may have helped to explore the issues governing assessment strategies in histopathology, in reality, the outcomes of the research might have had little effect on the status quo as power would still ultimately lie with the regulatory bodies. The research may then have lacked impact.

Another theory which was explored was Situated Learning Theory (SLT) first proposed by Jean Lave and Etienne Wenger which describes how learning occurs in a “community of practice” (Lave and Wenger 1991) whereby learning is a social process. Clinical activity usually occurs in teams and these teams are referred to as “communities of practice”. Within these groups, knowledge and skills are disseminated to the members through lectures, seminars or day-to-day contact between members. It is suggested that learners (or trainees in this context) are “legitimate peripheral participants” of the group (Lave and Wenger 1991). By taking part in the activities of the group, they are able to learn the necessary knowledge, skills and attitudes of full-members and gradually be accepted as full-members.

SLT has strong ties with specialist training in healthcare. In histopathology, trainees learn “on the job” and gradually acquire the necessary knowledge, skills and attitudes from their seniors. This can be an active or passive process, but in histopathology, this forms the majority of training as formal lectures or seminars either occur infrequently or can’t cover the depth or breadth of the curriculum. WBAs currently have elements of SLT because trainees are rated on a scale which asks the consultant to determine if the trainee falls below, meets or exceeds their expectation. This expectation is likely to stem from the consultant’s perception of what his peers would also expect, since each consultant has also had to pass through training and meet the expectations of the group before them. SLT could then be used to explore how trainees learn from their consultants and how competencies are developed.

SLT may have helped explore some of the important mechanisms by which competencies are passed down from consultant to trainee, but without knowing what the underlying qualities were that underpin diagnostic competence, it seemed that the focus of the study would be misplaced. A theory which would help determine the knowledge, skills and behaviours relevant to how consultants determine diagnostic competence in trainees, and the associated context, seemed much more appropriate. This would also have the possibility of providing an evidence base to inform future assessment strategies in histopathology. Social Judgement Theory is therefore discussed in the following section (2.4).

## 2.4 Theoretical framework

Judgement is a process of taking all factors into account, deciding what is most relevant and then acting accordingly. Clinical judgement can be difficult to grasp in the literature as numerous authors use the term interchangeably with terms such as critical thinking, clinical decision making and problem-solving (Benner, Tanner and Chesla 1996; Dowding and Thompson 2003; Lee, Chan and Phillips 2006). An individual’s perception is made up of many elements, such as their knowledge and personal experiences, which can affect how they interpret the cues from the environment which they are viewing (Hofer and Pintrich 1997).

Many assessments exercise judgement to make decisions about individual performance, which makes it very relevant to this study. Indeed, perceptions of WBAs in the healthcare literature will be linked to how assessors view competency and the factors they take into account.

Social Judgement Theory (SJT) is a theoretical framework that has been developed from the original work of Brunswik (1952), particularly the “lens model” that was later adopted by Hammond and colleagues (1964). SJT creates a way to understand clinical judgements, how individual cues (or facets of information) are used in making judgements and whether these judgements are consistent between individuals. Brunswik’s idea of probabilistic functionalism explores the interaction of the individual and his environment, how few things are certain, and outcomes are probabilistically related based on environmental variables (Cooksey 1996a). It is important to understand how consultants determine diagnostic competence in trainees for several reasons.

* It is imperative that the assessment tools in histopathology are aligned with the knowledge, skills and attitudes important to the profession. Those that relate to diagnosis are particularly important.
* Aligning assessment tools to these qualities and rendering the qualities explicit to trainers and trainees delivers guidance on how to assess and achieve diagnostic competence
* As well as the qualities relevant to diagnostic competence, it is important to understand the full judgement ecology (such as context) so that assessments are administered in line with how the competency is viewed (e.g. timing and application)
* If WBAs are replaced by other assessment tools in the coming years (such as EPAs), these assessments will still require consultants to make judgements regarding trainee competency. Clearly, the underlying “judgement” is still a key issue worth exploring.

SJT lends itself to understanding diagnostic competency in histopathology because there are no “gold standards” which can be used to determine if a correct decision regarding a trainee’s competency has been made. SJT has been utilised in several clinical fields, including the use of clinical information by nurses and the diagnostic ability of clinicians (Wigton, Hoellerich and Patil 1986; Thompson et al. 2005). Here, participants are presented with real-life scenarios (usually presented as vignettes) and ask to make a judgement on the information presented e.g. to refer or not refer a patient.

Examining Brunswik’s lens (Figure 2.1), the left side of the model is the “true state” or the particular situation an individual is exposed to (this is the judgement ecology). In this case, the “criterion” would be judging a trainee’s diagnostic competency. There are then various cues (e.g. X1 could be the trainee’s communication skills) which contribute to the ecology and have different weights. The right side of the model represents the “weighting” that the judge (or assessor in this case) gives to the cues when making a decision about their trainee (and these weightings may be different to the true state). These weights can be ascertained by presenting relevant individuals with a large number of scenarios and by altering the values of the individual cues. In the classical application of SJT, regression analysis is often used to determine what cues participants use or value in the decision making process (Cooksey 1996b).

**Cues**

Criterion

Judgement

**Cue utilisation**

**Ecological validity**

**X1**

**X2**

**X3**

**X4**

**Figure 2.1.** Brunswik’s lens model (Brunswik 1952). X1-X4 indicate different cues.

Judgement analysis (the methodology of SJT) does not rely on subjective interpretations but merely asks the individual to make judgements as they normally would. Clearly, the validity of such studies requires the cases presented to be realistic descriptions of the chosen environment, so called “representative design”. A good SJT study would utilize highly experienced highly experienced individuals about tasks that accurately resemble their reality (Brehmer and Brehmer 1988).

When trying to understand the judgement task, Cooksey recommends not only studying cue utilisation but the full judgement ecology, including the nature of the judgement and its context (Cooksey 1996b). According to Cooksey, the SJT framework requires the investigation of judgement to follow the following steps.

* Conceptualise the judgemental problem
* Understand the judgement ecology
* Identify relevant cues
* Sample cue profiles
* Sample participating judges
* Obtain judgements
* Capture individual judgement policy
* Compare policies

To this end, the current perception of WBAs in histopathology will help to form a basis to conceptualise the judgmental problem given that they are frequently used to assess competency during training. The judgement of diagnostic competency can then be explored to elicit the cues relevant to the judgement task and how these are utilised and valued by individuals. The last step “compare policies” is only possible when there is a recognised “optimal” way of doing something, such that one can compare optimal judgements with participant judgements. As little is known about the concept of diagnostic competence, this step is not possible.

This study will use a mixed-methods approach to achieve these objectives with an explanation for the method chosen found at the start of each relevant chapter. Qualitative research is being used increasingly in healthcare to elicit people’s understanding and the unique meaning they attach to their experiences (Greenhalgh and Taylor 1997). Qualitative research is exploratory in nature and may elicit questions which can be answered using quantitative studies. Mixed methods research utilises both research paradigms and combines the richness of qualitative methods with the rigour of quantitative techniques (Sale, Lohfeld and Brazil 2002; Cresswell et al. 2003). Triangulating data sources like this helps to give validity to research conclusions (Creswell 2014).

# **Chapter 3**

# **A questionnaire exploring the perception of WBAs in histopathology**

**Chapter 3: A questionnaire exploring the perception of WBAs in histopathology**

## 3. 1 Background

WBAs are used widely within healthcare training and aim to examine real-life, professional practice rather than the somewhat artificial situations created for examination purposes. However, the evidence base for their use in healthcare, and particularly in histopathology, is limited (see Chapter 1, sections 1.3.5 and 1.3.6).

As pathologists are responsible for making diagnoses, it is important to understand whether the current WBAs are capable of measuring the knowledge, skills and behaviour relevant to this.

The first step in Cooksey’s framework (see Chapter 2, section 2.4) involves understanding the context of the judgement problem. To this end, although the main focus of the overall study is to understand how consultants judge diagnostic competence in their trainees, first of all it is important to gather information on the overall perception of WBAs in histopathology as they are the main tool used to assess competency in general. This helps to frame the focus of the study and offers insight into how it fits into the overall judgment ecology.

## 3.2 Aim and objectives

**Aim**

To explore the perceptions of WBAs in histopathology by assessors and trainees.

**Objectives**

A questionnaire will be used to:

1. understand how WBAs are perceived in histopathology including a comparison to the themes identified in the healthcare literature: their perceived usefulness, validity and acceptance.
2. understand if WBAs in histopathology are perceived to measure diagnostic competence

## 3.3 Method

### 3.3.1 Introduction

Questionnaires are commonly used in medical education research for the purposes of trainee evaluation, feedback and opinion gathering (Artino et al. 2014). Indeed, questionnaires are particularly useful for gathering data about concepts that can be difficult to quantify, such as beliefs, attitudes and opinions. To this end, they have been used extensively in the evaluation of WBAs in medical and dental education (Wilkinson et al. 2008; Murphy et al. 2009; Pereira and Dean 2009; Grieveson et al. 2011; Finall 2012; Kirton et al. 2013). An online questionnaire was chosen to achieve the objectives of this first study for a number of reasons.

* The target population was geographically dispersed and best reached via online means
* The target population was large meaning the method had to be easily administered and data easily analysed
* Anonymity would allow participants to express opinions freely
* Respondents could respond at their leisure which can be important for busy NHS employees

### 3.3.2 Design

The questions used were derived from themes identified in the literature regarding the experience of using WBAs in other medical and dental specialties including their perceived usefulness, purpose and number required (Miller and Archer 2010; Bindal Wall and Goodyear 2011; The General Medical Council 2011). In addition, a question was included to ask about whether assessors and trainees felt WBAs measured diagnostic competence to tie in with the main research question. The questionnaire was sent to the RCPath for approval.

A Likert-type scale was chosen for questions 1-6 and 8-10 with 5 options ranging from “strongly agree” to “strongly disagree”. A Likert-type scale was used because they offer participants more opportunity to express their opinion compared to dichotomous response options (Likert 1932). A 5-point Likert-type scale was chosen as it has been reported that narrower scales can truncate opinion and scales larger than 7 points can increase frustration level of respondents and reduce response rate (Sachdev and Verma 2004; Carifio and Perla 2007).

Question 7 gave respondents four options to choose from as well as “other” as an option which included a free text comment box. The questionnaire was piloted with two specialist trainees and a consultant histopathologist to ensure the questions were clear and appropriate. No changes were suggested.

The final questionnaire can be seen overleaf (Table 3.1).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Strongly agree** | **Agree** | **Neither agree nor disagree** | **Disagree** | **Strongly Disagree** |
| Q1. WBAs are useful in training |  |  |  |  |  |
| Q2. The number of WBAs per stage is appropriate |  |  |  |  |  |
| Q3. There is sufficient time to complete WBAs |  |  |  |  |  |
| Q4. WBAs identify diagnostic competence in histopathology |  |  |  |  |  |
| Q.5 WBAs identify trainees in difficulty |  |  |  |  |  |
| Q.6 WBAs accurately reflect trainee progress |  |  |  |  |  |
| Q7. Which of these four options do you think are more important? WBAs are part of the curriculum to: | **Help monitor trainee development/progress** | **Provide structured feedback to trainees** | **Satisfy the requirements of the Royal College** | **Provide evidence for ARCP** | **Other** |
|  |  |  |  |  |

**Table 3.1 Questionnaire (continued on next page)**. Participants were asked to select one option from those available (full participant information sheet can be found in Appendix 2). Please note questions 9 and 10 were different for assessors (normal font) and trainees (highlighted in italics).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Strongly agree** | **Agree** | **Neither agree nor disagree** | **Disagree** | **Strongly disagree** |
| Q8. The current WBAs are acceptable to you as a trainer/trainee |  |  |  |  |  |
| Q9. You have been given sufficient guidance on how to complete WBAs for trainees |  |  |  |  |  |
| *Q9. Your assessors are accessible to complete WBAs* |  |  |  |  |  |
| Q.10 Trainees select appropriate cases for WBAs |  |  |  |  |  |
| *Q10. Your assessors are engaged and helpful when completing WBAs* |  |  |  |  |  |

**Table 3.1 Questionnaire continued**. Participants were asked to select one option from those available (full participant information sheet can be found in Appendix 2). Please note questions 9 and 10 were different for assessors (normal font) and trainees (highlighted in italics).

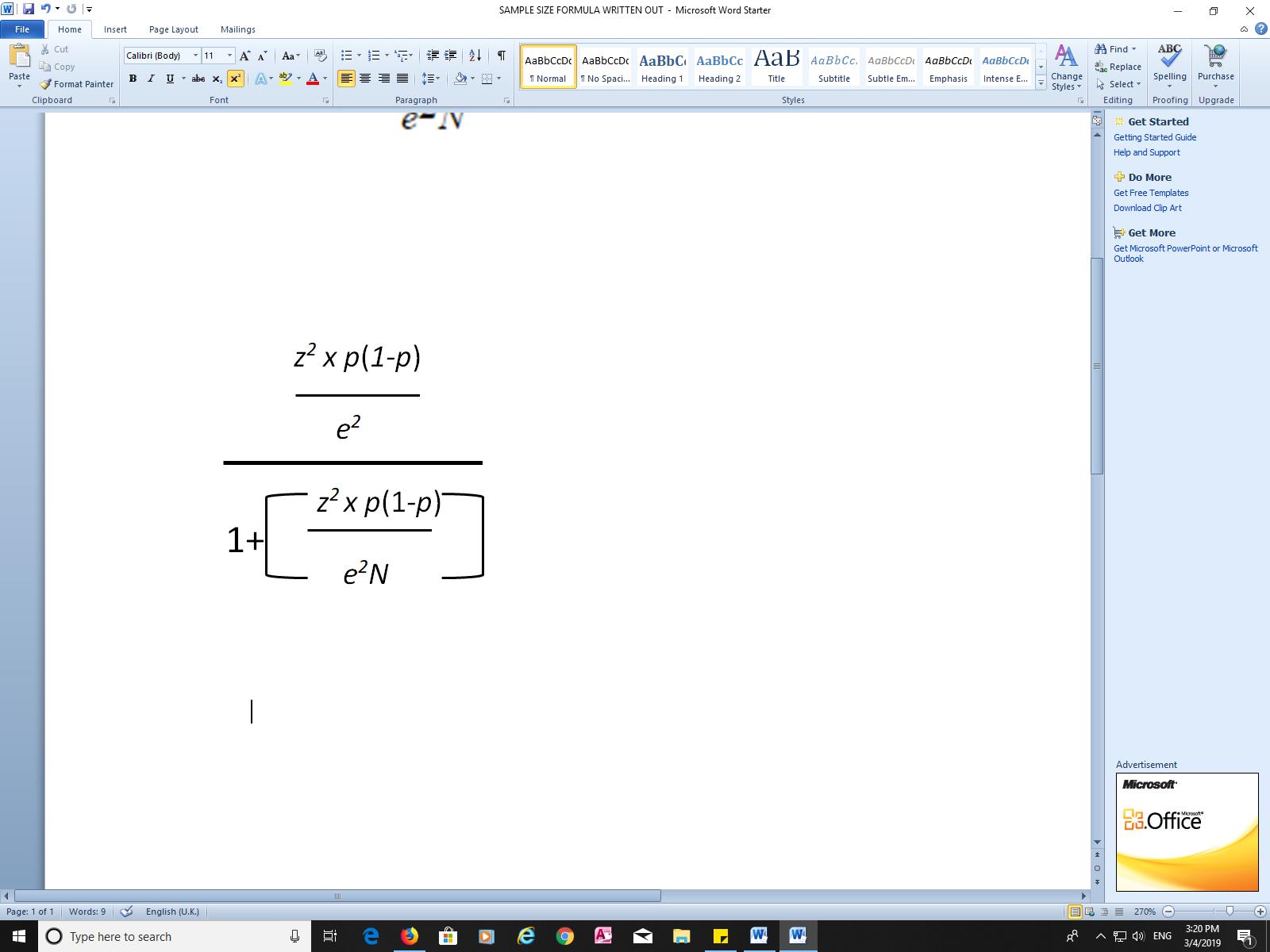
**3.3.3 Target population**

The questionnaire targeted all histopathology consultants and trainees in the United Kingdom as well as those biomedical scientists who complete WBAs for trainees. These individuals were chosen as they complete the majority of WBAs for trainees and therefore potentially offer the greatest insight.

The RCPath agreed to use their mailing list to distribute the questionnaire which included all registered assessors (2175) (consultants and biomedical scientists) and trainees (519) from the specialties of general histopathology, OMFP, paediatric pathology, neuropathology and forensic pathology. Collaboration with RCPath was important for several reasons:

* One study has shown that response rates for postal questionnaires are significantly lower from commercial organisations than universities (Iglesias, 2002). It was anticipated that by using RCPath to send the invitation email, respondents would have confidence that the questionnaire was from a reputable source and they would be more likely to respond.
* Working with the RCPath was the most efficient method for targeting the relevant individuals as they hold an updated mailing list of the aforementioned individuals
* The results of the questionnaire and/or future studies are potentially of interest to the RCPath given their role in overseeing assessment tools for trainees. Lack of consultation at this stage may have adversely affected future collaboration in the field of pathology education.

In order to ensure reliable conclusions could be drawn from the questionnaire data, it was important to know what an acceptable response rate would be. There is no agreed standard for response rates but the purpose is to achieve representativeness (Fincham 2008). This can be calculated with the formula shown in Figure 3.1 below (Daniel 1999):



**Figure 3.1. Formula to calculate number of respondents required for a given level of reliability** (Daniel 1999)**.**  Population Size = N, Margin of error = e, Z-score = z (The z-score is the number of standard deviations a given proportion is away from the mean), Percentage value = p (percentage picking a choice, expressed as decimal).

Using the above formula 221 trainees out of the 519 targeted would have to reply to achieve results that are reliable with a 95% confidence level and 5% margin of error. To explain further, if the response to a particular question was “60% strongly agree”, in actual fact the true number which actually strongly agreed would fall between 55% and 65% (5% margin of error), 19 times out of 20 (95%).

However, it is important to point out that the above formula is suited to questionnaires that have used random sampling. In this study design, the entire population is targeted and if the entire population responded to the questionnaire, then representativeness would be achieved. However, what is more likely is that a proportion of this population will respond so random sampling is not achieved as those who respond are not a random selection. They will be systematically different than those who do not respond (Cook, Heath and Thompson 2000). Upon consultation, RCPath, indicated that previous questionnaires to the same group of individuals had suffered very poor response rates so it was felt necessary to target the entire population in order to capture opinion. Therefore, the response rate formula above was used only as a guide and those methods which are reported to increase response rates were utilised (Iglesias 2002). These were

* keeping the length of the questionnaire short
* keeping the design and text clear and simple
* reminders to non-responders
* alternative platforms (such as email, post, websites)

These points were addressed by limiting the questionnaire to 10 questions and keeping questions short and explicit. A link to the online questionnaire was also included in the RCPath’s eNewsletter which targeted the same population. The link to the questionnaire was included in the LEPT (Learning environment for Pathology Trainees) system, which is the online platform that assessors and trainees use to document WBAs. Reminder emails were not permitted by RCPath.

### 3.3.4 Distribution

The invitation email (Appendix 1) and questionnaire link was sent by the RCPath to all aforementioned individuals and left open for an 8 week period. After approximately a month, the link to the questionnaire appeared in the eNewsletter and LEPT to try and increase the response rate. The questionnaire was delivered via Survey Monkey™ and included a description of the project and other relevant documentation including information sheet, contact details, a reminder that the questionnaire was voluntary and that responses would be anonymous (see Appendix 2). The study had ethical approval from the University of Sheffield (reference 008042).

### 3.3.5 Data analysis

Data were entered into SPSS V.21.0 (SPSS, Chicago, Illinois, USA) and were analysed with a p<0.05 considered a statistically significant difference. Cronbachs alpha test was used to determine the internal reliability of the questionnaires.

In order to determine whether a significant difference in opinion between assessors and trainees responses was present, statistics on the data was required. Given that Likert-type items give ordinal data, the non-parametric Mann Whitney U test was required (de Winter and Dodou 2010). The comparison included questions 1-6 and question 8. Question 7 was excluded as it did not to offer Likert-type responses and questions 9 and 10 were excluded as these were different questions for the assessors and trainees.

Question 7 of the questionnaire allowed participants to provide a free text comment of up to 9 lines of text or 90 characters. Open-ended responses can prove difficult to analyse and qualitative or quantitative approaches have been adopted when managing this issue (O’Cathain and Thomas 2004). Importantly, these comments should be analysed because they can offer further insight into the issues being explored. This can identify further research questions or help to make sense of the responses given to other questions.

In favour of a qualitative approach, one might argue that it is more appropriate because the responses are written entirely by the participant without influence from the researcher. Secondly, some responses can be quite lengthy and detailed with a quantitative approach unable to explore and describe the data appropriately. However, the data is not strictly qualitative either because some participants only provide a few words as their answer. Furthermore, there is no opportunity for opinions and experiences to be explored further by the researcher. Participants may also have been influenced by previous questions, framing their ideas and thought processes.

It was decided that the free text comments would be analysed qualitatively by thematic analysis to try and find common issues or “codes” within the data. This method was chosen as it is a relatively simple way to organize comments into common themes, thereby facilitating understanding of the main issues. Firstly, the free text comments were each analysed by three individuals to find common ideas and opinions, grouping them into “codes”. They met to discuss their findings, confirming the codes and highlighting comments from the respondents which supported these. The individual “codes” were then grouped under three main themes through an iterative process of discussion and reflection (for more detail on thematic analysis, see Chapter 4, section 4.3.5).

## 3.4 Results

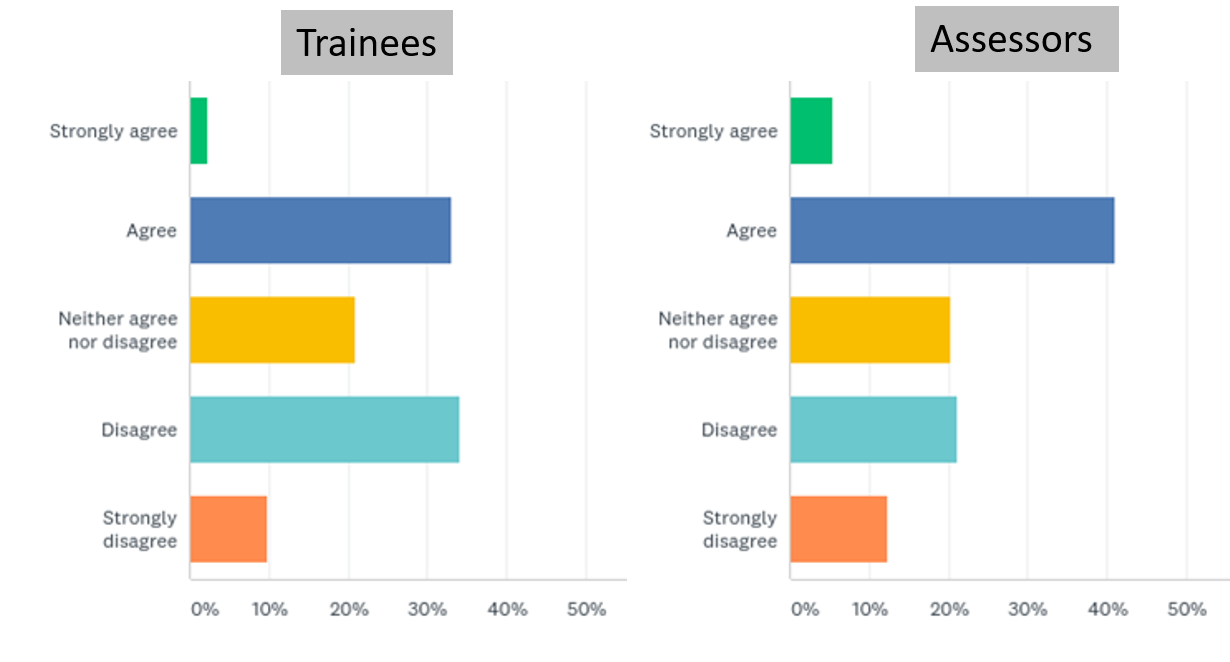
The assessor questionnaire was completed by 242/2175 assessors (response rate 11.1%) and the trainee questionnaire by 91/519 trainees (response rate 17.5%) - a sufficient response rate to ensure the results of both questionnaires were reliable at the 95% confidence level with a 10% margin of error (Daniel 1999). The internal reliability of both the questionnaires was good with overall Cronbach’s alphas of 0.87 and 0.88 for the assessor and trainee questionnaire respectively.

A Mann Whitney U test showed there was no significant difference in the perception of WBAs between the assessors and trainees overall (*p* =0.277).

The results of each question are summarised hereafter in text and graphical form with summary Tables 3.2 and 3.3 thereafter.

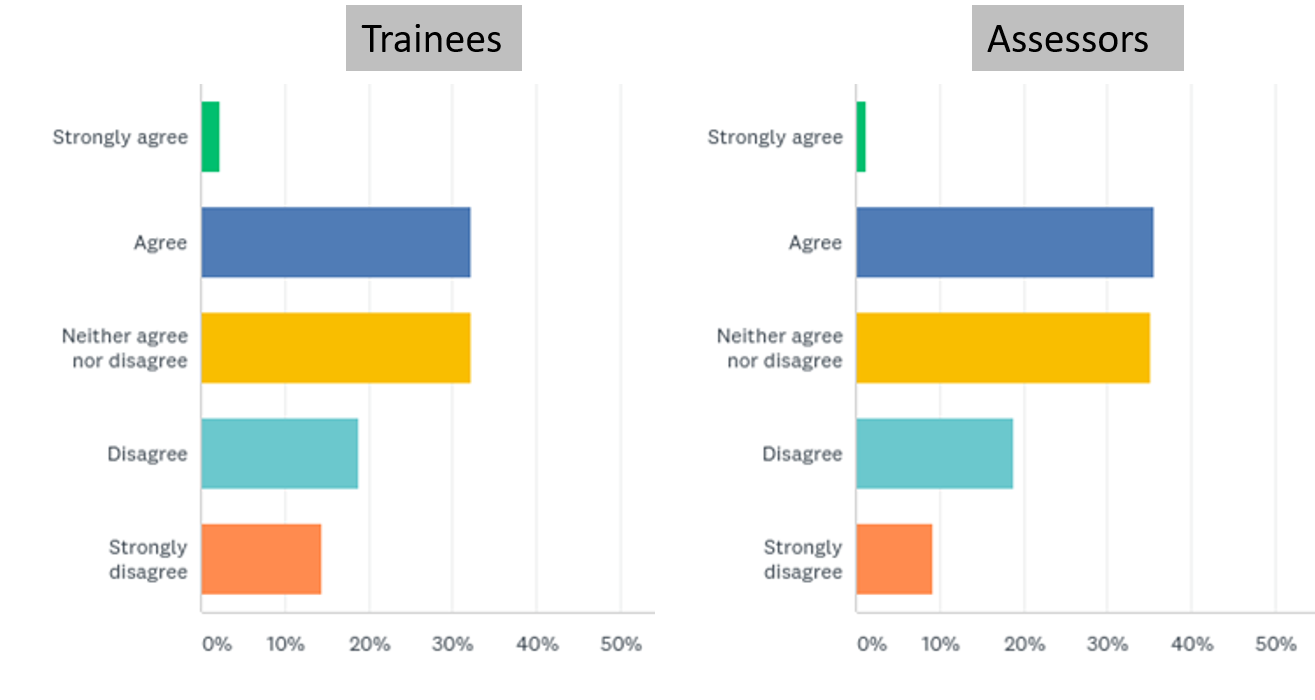
**Question 1**: **WBAs are useful in training (Figure 3.2)**

44% of trainees strongly disagreed or disagreed that WBAs are useful in training, and while a noteworthy proportion of assessors shared a similar perspective (33%), a greater number of assessors strongly agreed or agreed that WBAs are useful (46%) compared to trainees (35%). Similar proportions of trainees and assessors neither agreed nor disagreed with this statement.

**Figure 3.2. Trainee and assessor responses to the statement: ‘WBAs are useful in training’**

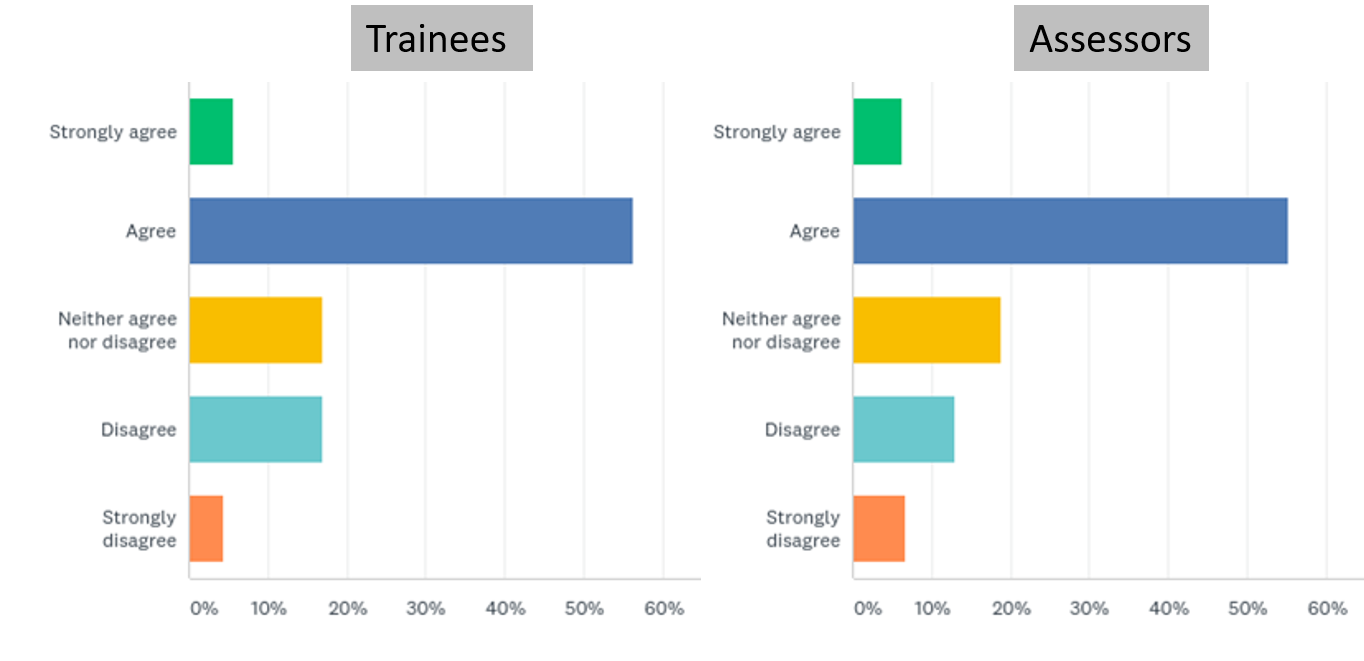
**Question 2: The number of WBAs per stage is appropriate (Figure 3.3)**

There were no clear cut results for this question from either trainees or assessors. Roughly equal numbers of trainees and assessors strongly agreed or agreed (34% and 37%), neither agreed nor disagreed (32% and 28%) or strongly disagreed or disagreed (33% and 32%) respectively that the number of WBAs per stage of training is appropriate.

**Figure 3.3. Trainee and assessor responses to the statement: ‘The number of WBAs per stage is appropriate’**

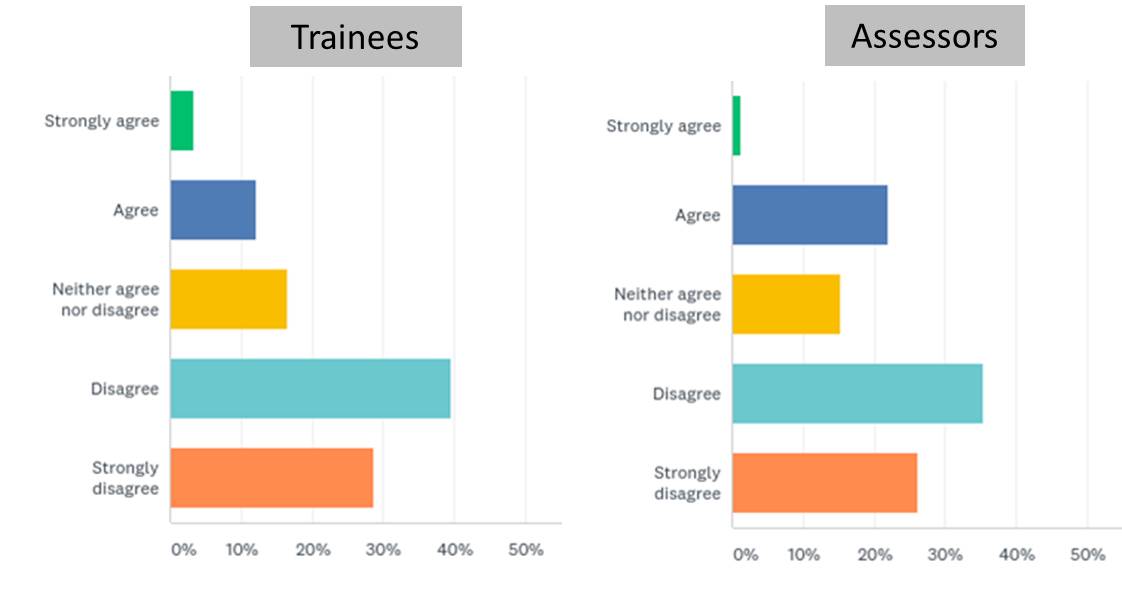
**Question 3: There is sufficient time to complete WBAs (Figure 3.4)**

The majority of both trainees and assessors felt there was sufficient time to complete WBAs. 62% of trainees and 62% of assessors strongly agreed or agreed and only 21% of trainees and 20% of assessors strongly disagreed or disagreed with this statement. 17% of trainees and 18% of assessors neither agreed nor disagreed with the statement.

**Figure 3.4. Trainee and assessor responses to the statement: ‘There is sufficient time to complete WBAs’**

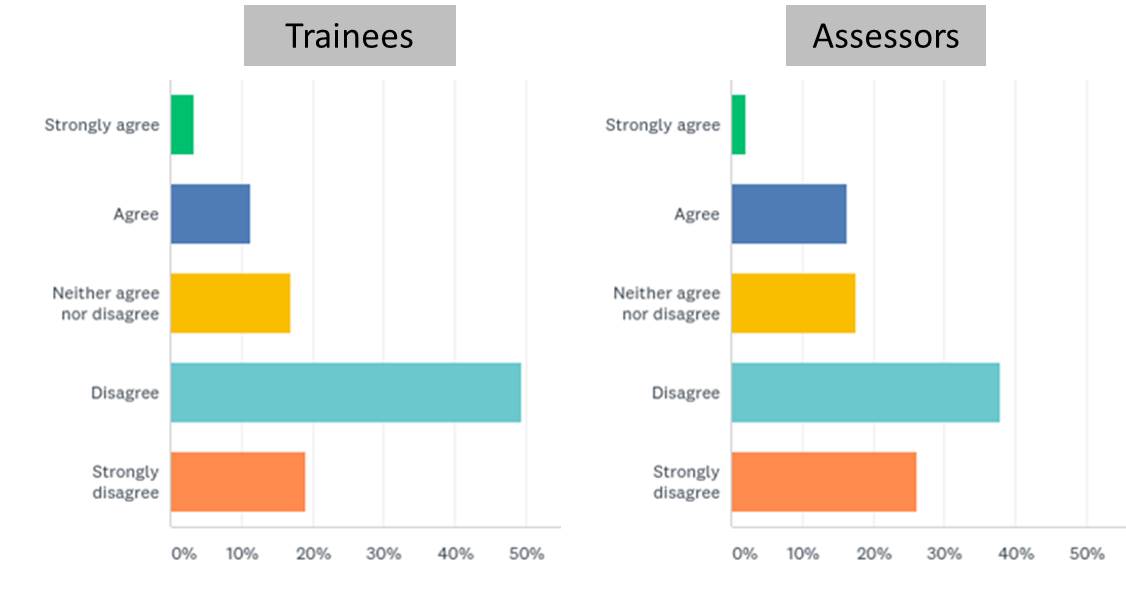
**Question 4: WBAs identify diagnostic competence in histopathology (Figure 3.5)**

Roughly two thirds of trainees and assessors strongly disagreed or disagreed (68% and 62% respectively) that WBAs identify diagnostic competence in histopathology. Conversely only 15% of trainees and 23% of assessors strongly agreed or agreed with this statement. 17% of trainees and 15% of assessors neither agreed nor disagreed with the statement.

**Figure 3.5. Trainee and assessor responses to the statement: ‘WBAs are able to identify diagnostic competence in histopathology’**

**Question 5: WBAs identify trainees in difficulty (Figure 3.6)**

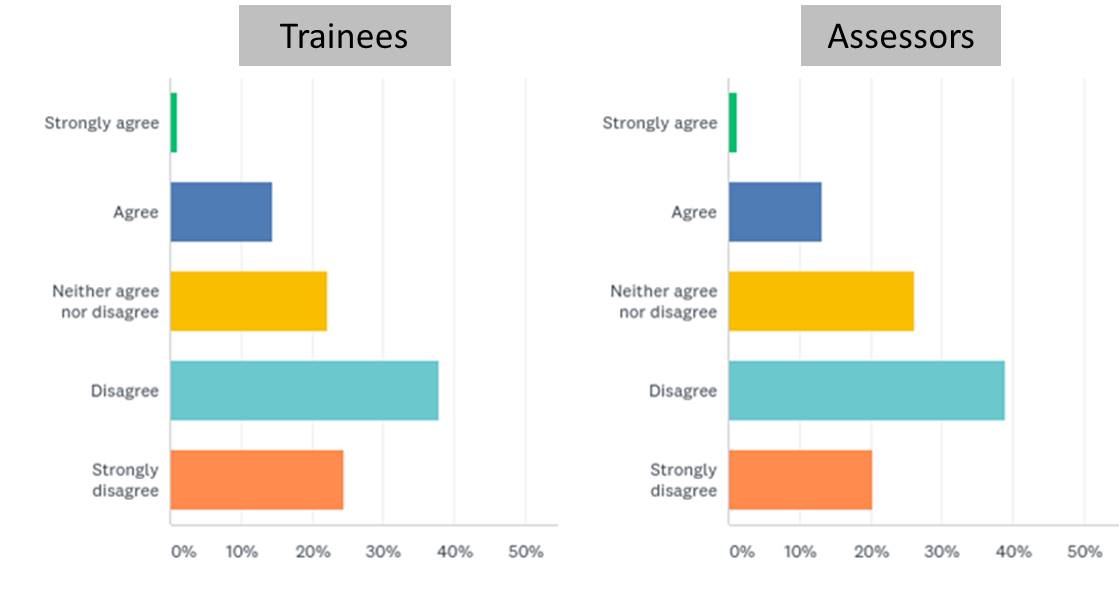
Similar to the results of question 4, two thirds of both trainees and assessors strongly disagreed or disagreed that WBAs identify trainees in difficulty (69% and 64% respectively)whilst only 15% of trainees and 18% of assessors agreed they did identify such trainees. 16% of trainees and 18% of assessors neither agreed nor disagreed with the statement.



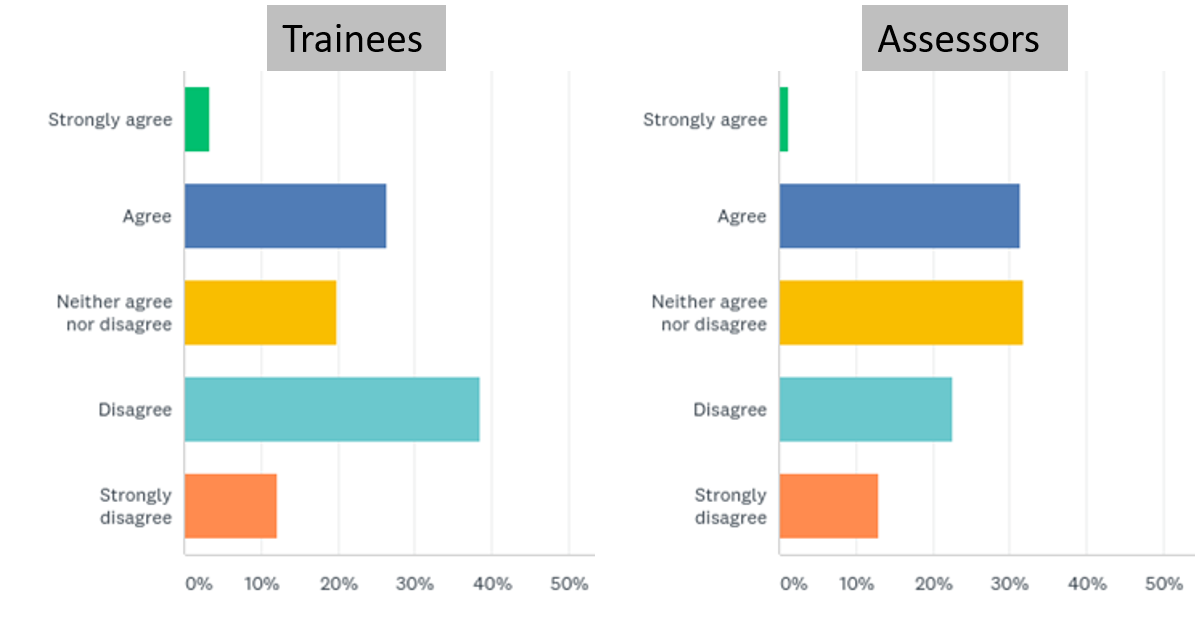
**Figure 3.6. Trainee and assessor responses to the statement: ‘WBAs are able to identify trainees in difficulty’**

**Question 6: WBAs accurately reflect trainee progress (Figure 3.7)**

The majority of both trainees and assessors strongly disagreed or disagreed that WBAs accurately reflect trainee progress (62% and 59% respectively) and conversely only 16% of trainees and 15% of assessors strongly agreed or agreed they did. 20% of trainees and 26% of assessors neither agreed nor disagreed with the statement.

**Figure 3.7. Trainee and assessor responses to the statement: ‘WBAs accurately reflect trainee progress’**

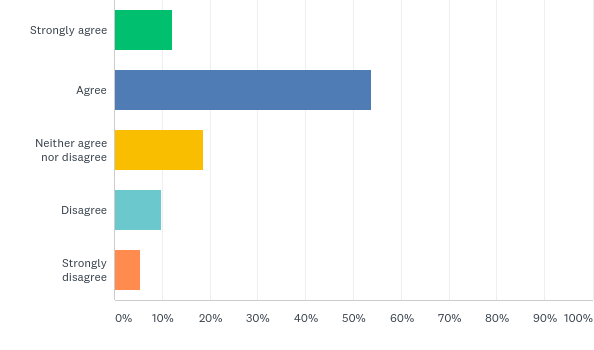
**Question 8**: **The current WBAs are acceptable to you as a trainee/assessor (Figure 3.8). \*Note, the results of question 7 will be reported below.** Approximately a third of respondents thought the current WBAs were acceptable (33% of trainees and 30% of assessors), but a larger proportion of trainees (51%) than assessors (36%) strongly disagreed or disagreed with this statement. 16% of trainees and 34% of assessors neither agreed nor disagreed with the statement.



**Figure 3.8. Trainee and assessor responses to the statement: The current WBAs are acceptable to you as a trainee/assessor**

**Question 9 for trainees:** **Your assessors are accessible to complete WBAs (Figure 3.9)**

The majority of trainees strongly agreed or agreed that their assessors were accessible (66%) and only 15% strongly disagreed or disagreed with this. 19% of trainees neither agreed nor disagreed with the statement.



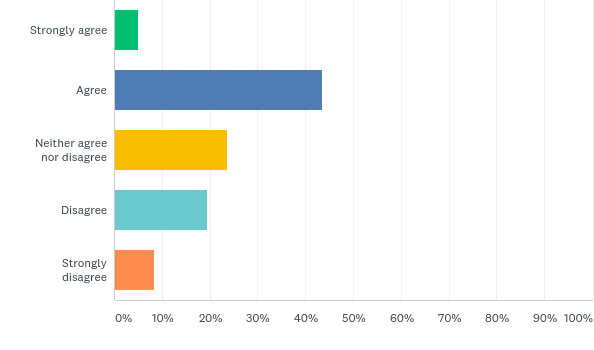
Trainees

**Figure 3.9. Trainee response to the statement: ‘Your assessors are accessible to complete WBAs’**

**Question 9 for assessors**: **You have been given sufficient guidance on how to complete WBAs (Figure 3.10)**

Almost half of assessors (49%) strongly agreed or agreed that they had been given sufficient guidance on how to complete WBAs. However, 28% strongly disagreed or disagreed with this statement. 77% of assessors neither agreed nor disagreed with the statement.

statement.



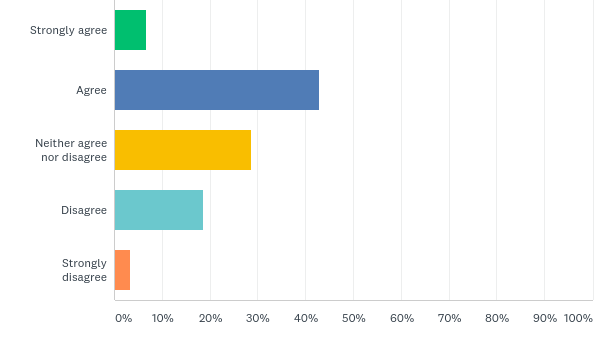
**Assessors**

**Figure 3.10. Assessor response to the statement: ‘You have been given sufficient guidance on how to complete WBAs’**

**Question 10 for trainees:** **Your assessors are engaged and helpful when completing WBAs (Figure 3.11)**

Almost half of trainees (49%) strongly agreed or agreed that their assessors were engaged and helpful when completing WBAs and 22% strongly disagreed or disagreed. 29% of trainees neither agreed nor disagreed with the statement.

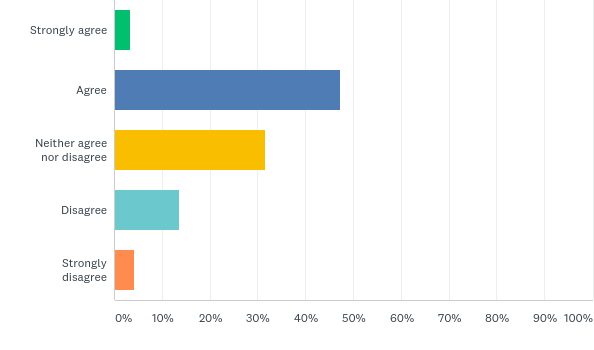
Trainees



**Figure 3.11. Trainee response to the statement: Your assessors are engaged and helpful when completing WBAs**

**Question 10 for assessors:** **Trainees select appropriate cases for WBAs (Figure 3.12)**

51% of respondents strongly agreed or agreed and only 18% strongly disagreed or disagreed that trainees selected appropriate cases for assessment. 31% of assessors neither agreed nor disagreed with the statement.



Assessors

**Figure 3.12. Assessor response to the statement: Trainees select appropriate cases for WBA**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Statement | % Strongly agree/agree | | % Neither agree nor disagree | | % Strongly disagree/ disagree | |
|  | **A** | **T** | **A** | **T** | **A** | **T** |
| 1. WBA are useful in training | 46.3 | 35.2 | 20.3 | 20.9 | 33.5 | 44.0 |
| 2. No. WBA per stage appropriate | 36.9 | 34.4 | 35.2 | 32.2 | 28.0 | 33.2 |
| 3. Sufficient time to complete WBA | 61.5 | 61.8 | 18.8 | 16.9 | 19.6 | 21.4 |
| 4. WBAs identify diagnostic competence | 22.2 | 15.4 | 15.2 | 16.5 | 61.6 | 68.2 |
| 5. WBAs identify trainees in difficulty | 18.4 | 14.6 | 17.5 | 16.9 | 64.2 | 68.5 |
| 6. WBAs accurately reflect trainee progress | 14.5 | 15.5 | 26.1 | 22.2 | 59.0 | 62.2 |
| 8. WBAs are acceptable to you: assessor/trainee | 32.7 | 29.7 | 31.8 | 19.8 | 35.6 | 50.6 |
| 9.Sufficient guidance on how to complete WBAs | 48.6 | - | 23.7 | - | 27.8 | - |
| *9. Assessors accessible to complete WBAs* | - | 66.0 | - | 18.7 |  | 15.4 |
| 10. Trainees select appropriate WBAs cases | 50.6 | - | 31.5 | - | 17.9 | - |
| *10. Assessors are engaged & helpful with WBAs* | - | 49.5 | - | 28.6 | - | 22.0 |

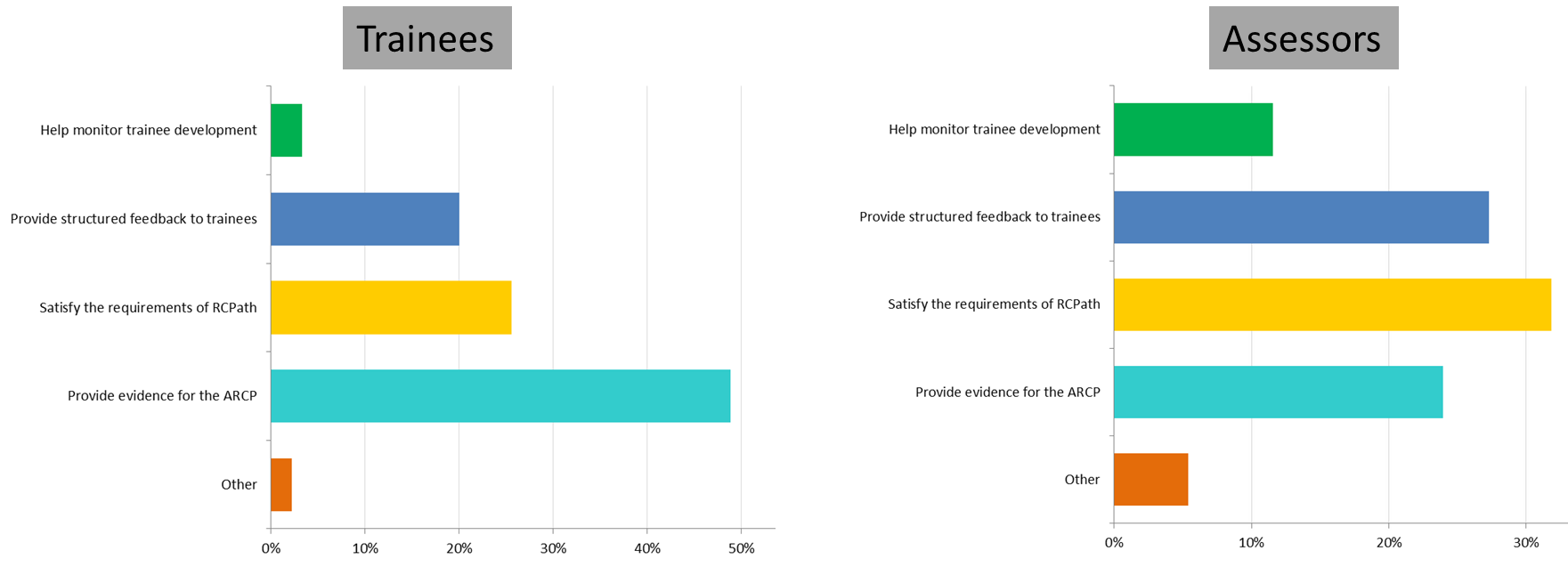
**Table 3.2. Summary of responses to questionnaire (excluding question 7) A=assessors, n = 242. T=trainees, n= 91.** Please note questions 9 and 10 were different for assessors (normal font) and trainees (highlighted in italics).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 7. Which of the four options are most important? WBAs are part of the curriculum to: | Help monitor trainee development | Provide evidence for ARCP | Satisfy requirements of RCPath | Provide structured feedback | Other |
| % assessors choosing option | 11.6 | 27.3 | 31.8 | 24.0 | 5.4 |
| % trainees choosing option | 3.3 | 25.6 | 48.9 | 20.0 | 2.2 |

**Table 3.3. Summary of responses to question 7. A=assessors, n = 242. T=trainees, n= 91.**

**Question 7**: **Which of these options are more important?** **(Figure 3.13)**

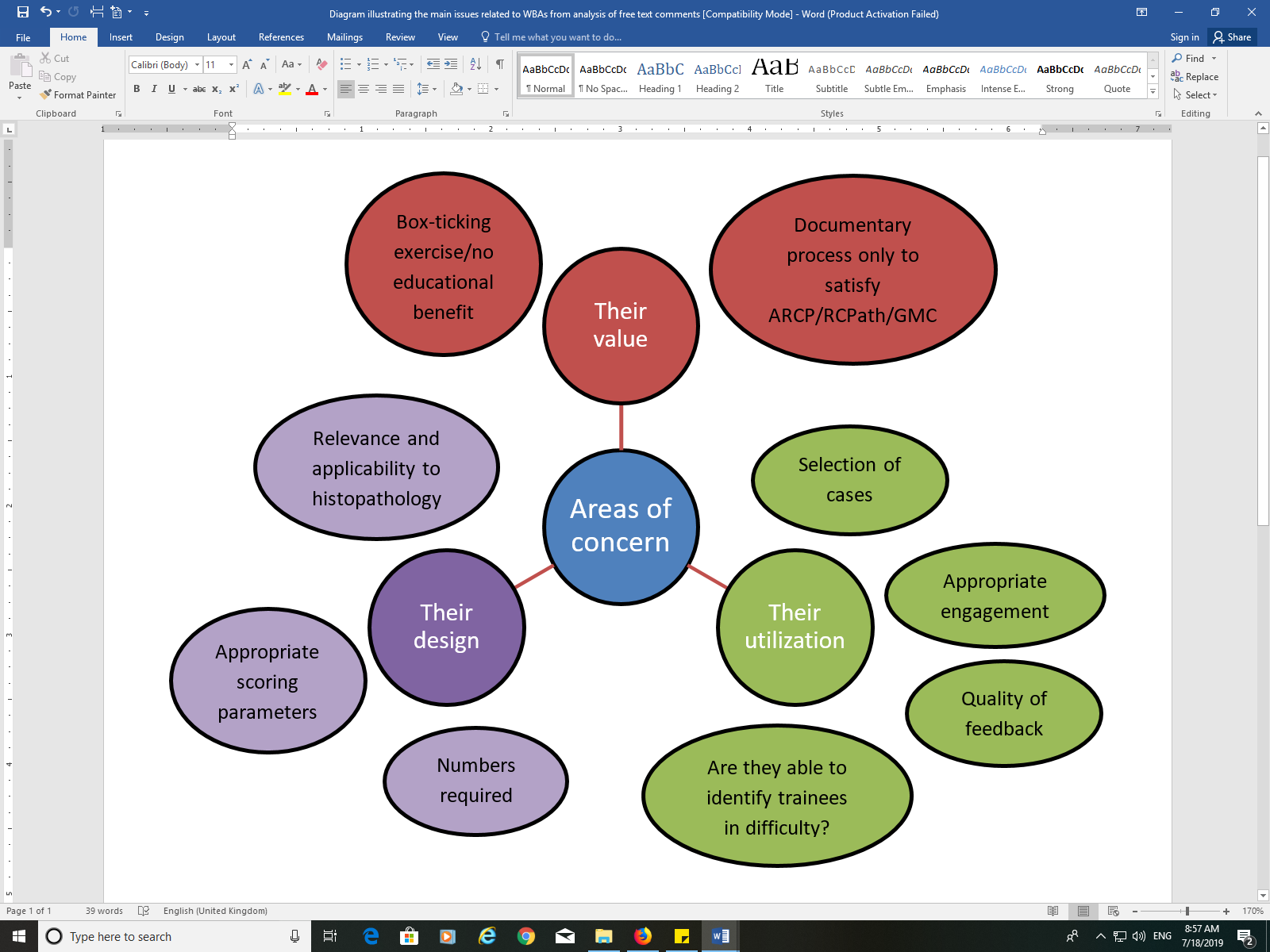
WBAs are part of the curriculum to: 1. Help monitor trainee development/progress, 2. Provide structured feedback to trainees, 3. Satisfy the requirements of the RCPath, 4. Provide evidence for the ARCP, 5. Other (see Table 3.3 and Figure 3.13). Approximately a quarter of trainees (26%) and 27% of assessors chose “Provide evidence for the ARCP” to explain why WBAs are part of the curriculum while 49%% of trainees and 32% of assessors thought it was to “Satisfy the requirements of the RCPath”. Only a small proportion of trainees (20%) and 24% of assessors said they were part of the curriculum to “Provide structured feedback to trainees” and very few (3% of trainees and 12% of assessors) chose the option “Help monitor trainee development/progress”.

****



Question 7 yielded 42 free text comments from 7/91 trainees and 35/242 assessors which as they were very similar were combined for analysis. A few respondents also emailed the project lead with further comments which were also included in the analysis. Interestingly, the majority of respondents used this opportunity to express their opinions on WBAs generally, with the comments covering a range of issues.

The main issues from the comments related to the design, value and utilization of WBAs and are described with free text comments and Figure 3.14.



**Figure 3.14. Diagram illustrating the main areas of concern regarding the use of WBAs in histopathology generated from free text comments.** ARCP, Annual Review of Competency Progression; RCPath, The Royal College of Pathologists; WBA, Workplace-based Assessment

3.5 Design

Three sub-themes related to the design of WBAs: relevance and applicability to histopathology, appropriate scoring parameters and appropriate number of WBAs to complete. Selected quotes are given below to demonstrate these three sub-themes.

**Relevance and applicability to histopathology**

* *“The current work based assessments are orientated to frontline hospital medicine and not histopathology and I think that they should be modified for each specialty to make them more relevant.”*
* *“I believe they are there because they are used in other specialties without sufficient consideration as to whether the method would carry equivalence in a very different kind of training environment”.*
* *“In my experience neither party finds them acceptable as assessments and therefore their utility is lost”.*

**Appropriate scoring parameters**

* *“I don't agree with the grading system that is used as it allows for too much intra-observer variability, I think this could be simplified for better consistency”.*
* *“An academic exercise in my opinion which doesn’t really help the trainee as you can never be negative or critical. Everyone is usually satisfactory which isn't the case in reality”.*
* *“The categories for scoring are not very appropriate”.*

**Appropriate number of WBAs to complete**

* *“I think it's more important to do fewer but more worthwhile assessments than to have to reach an arbitrary minimum number”.*
* *“…they are often somewhat forced and artificial; part of that is due to the process being driven by a quota - I would have more sympathy if there was no quota (leaving it to local training committee's and educational supervisors to explore why a trainee was WPBA-lite)”.*
* *“This is part is due to the LEPT system recording only 18 assessments (less in stage D) which means that including multiple WBAs of the same kind of activity is not an option. Requiring a far larger number of WBAs would allow them to be used appropriately to document trends in ability, and also would force them to become more integrated into the weekly/daily routine…*

**Value**

Two sub-themes related to the value of WBAs: box-ticking exercise/no educational benefit and documentary process only to satisfy ARCP/RCPath/GMC. Selected quotes are given below to demonstrate these two sub-themes:

**Box-ticking/or no educational benefit**

* *“Overall, I think the assessments are more of a box-ticking administrative exercise…”.*
* *“These assessments are box ticking, educationalists' nonsense”.*
* *“I would support WBA being scrapped and not replaced with anything”.*

**Documentary process to satisfy ARCP/RCPath/GMC**

* *“There is an obsession with documentation which does nothing to improve training or identify issues”.*
* *“The main but limited utility of WBA is to document things that a trainee can do or has done so there is some documentation that can be gained to support portfolios for ARCPs”.*
* *“They should be there to help provide and record structured feedback to trainees as a formative process. Unfortunately, they are not set up this way and have become just another set of hoops to jump through”.*

**Utilisation**

Four sub-themes related to the utilisation of WBAs: selection of cases, appropriate engagement, quality of feedback given to trainees and whether they are able to identify trainees in difficulty**.** Selected quotes are given below to demonstrate these four sub-themes:

**Selection of cases**

* *“Unfortunately the trainees only chose cases when things went well, so the forms may not be representative of true performance. It might be better for the assessor to choose the event”.*
* *“What we need is regular subjective feedback from multiple assessors based on performance in the job (like dissecting/reporting cases over the period of a month with a trainee working closely with you), not ones specially selected which prove absolutely nothing”*
* *“I often observe of trainees completing the majority during the last month prior to submission of RCPath paperwork”.*

**Appropriate engagement**

* *“Some trainers and assessors really engage with the process appropriate questions are asked and good feedback is given”.*
* *“I think that assessors are not consistent when completing WBAs and seem to have different understandings of how to complete them”.*
* *“In my opinion, WBAs are a good tool that are used poorly”.*

**Quality of feedback**

* *“…I have submitted some to some assessors who have not even changed anything I have completed and have signed it off with no useful feedback at all.”*
* *“It is extremely sad that on occasions trainees don't get structured, constructive feedback other than when dong WBAs, and even worse to see WBAs completed with no feedback from the trainer”.*
* *“It is a good opportunity to give feedback but should always be done with the trainee in attendance. Otherwise it's a pointless exercise. It's also a good opportunity to talk about other issues, positive or negative”.*

**Ability to identify trainees in difficulty**

* *“I feel that workplace-based assessments are probably particularly useful for identifying trainees-in difficulty or 'problem' trainees who are disengaged with training or unsafe (i.e. those at the extreme end of the spectrum)”.*
* *“I thought work-based assessments were intended to identify struggling trainees but I don't think that they are a very robust tool for this purpose”.*
* *“It is never possible to be truthful on these assessments as those trainees in difficulty cannot be identified and weeded out for extra help or advice.”*

## 3.6 Discussion

As described in the literature review (Chapter 1, section 1.1.5), the “utility of assessment” describes the “usefulness” of any assessment and how it is derived in a multiplicative manner from its reliability, validity, educational impact, acceptability and cost (van der Vleuten 1996). Given the involvement of PMETB in setting the standards of post-graduate education (including WBAs), and that it adopted the “utility of assessment” in its document “Developing and maintaining an assessment system-a PMETB guide to good practice” (Oliver 2007), the results will be discussed with references to these parameters.

The first question which asked assessors and trainees if WBAs are useful in training was integral to the survey and served as an overall measure of the “utility” of these tools in histopathology. Interestingly there appears to be a negative perception of their usefulness particularly among trainees. Whereas 46% of assessors thought WBAs were useful in training only 35% of trainees thought so and conversely 44% trainees disagreed or strongly disagreed they were useful compared with 33% of assessors. Although this difference between the groups is relatively small and does not reach statistical significance there is supporting evidence from the results of Q8 that it may be real. More than 50% of trainees compared with only 36% of assessors strongly disagreed/disagreed with the statement that ‘Current WBAs are acceptable to you as an assessor/trainee’. User acceptability of WBAs has been shown to be a major problem (Massie and Ali 2015). This is particularly disappointing because WBAs are meant to be designed for the benefit trainees by fostering development, yet half of trainees in this study judged them to be unacceptable assessments. It is not clear why trainees in particular appear to have a negative opinion of the usefulness or acceptability of WBAs. Indeed, it cannot be explained by the results of questions which looked at the purpose of the WBAs since the majority of trainees and a similar proportion of assessors (roughly two thirds) did not think WBAs identify diagnostic competence (Q4), identify trainees in difficulty (Q5) or accurately reflect trainee progress (Q6). One possibility may be that trainees think WBAs are part of the curriculum for regulatory rather than educational reasons, something that also been expressed by surgical trainees (Powell, DiMarco and Saeed, 2014). In Q7 trainees and assessors were asked why they thought WBAs were part of the curriculum. Almost 75% of trainees compared with 56% of assessors thought they were to provide evidence for the ARCP or the RCPath and only 23% of trainees and 39% of assessors thought they were to help monitor trainee progress or provide structured feedback to trainees. These responses may explain why a large number of trainees do not find WBAs useful-as there is little perceived educational benefit. However, it could be argued that the term “useful” has different meaning to assessors and trainees. For instance, a trainee might interpret “useful” from the perspective of a WBA being able to identify their weaknesses and guide their development. An assessor, however, might interpret this question differently, interpreting “useful” as a means to provide evidence to the ARCP panel (whether that evidence be accurate or otherwise).

Analysis of free-text comments provide further insight into this, but of note, a number of assessors also commented on the perceived regulatory nature of WBAs. Comments included *“These assessments are box ticking, educationalists' nonsense”*; *“just another set of hoops to jump through”* and an *“obsession with documentation which does nothing to improve training or identify issues”.* Until assessors and trainees are more accepting of WBAs or any future iteration, then the educational value of these tools will suffer. One way of improving user acceptability may be to change the assessments so that trainees in particular feel that they have something to gain by doing them-whether that be improved feedback or by altering the forms themselves, so the descriptors of performance are more “valid” and better geared towards histopathology. It is important to note that irrespective of the perceived purpose and acceptability of WBAs, junior doctors are reported to be under more stress of late with low morale (Rimmer 2016). This may be having a negative effect on how they currently perceive WBAs.

Looking at the issue of feedback first, 49% of assessors felt they had been given sufficient training to complete WBAs (Q9) but approximately equal numbers had no opinion or disagreed. Ideally, all assessors should feel they have had enough training so that each trainee is assessed similarly to his peers and the feedback provided to the trainee is useful. Studies exploring why feedback may have limited impact on a trainee and how the trainee processes the feedback is a growing area of interest (Bindal, Wall and Goodyear 2011; Kogan et al. 2012). Trainers may feel they have insufficient time and expertise to foster development in their trainees while trainees sometimes view WBAs as a chore with little respect for their importance (Bindal, Wall and Goodyear 2011; Kogan et al. 2012). Appropriate guidance and training is also important to ensure WBAs are utilised how they were intended and so deliver the outcomes they are meant to assess (Ryland et al. 2006; Kirton et al. 2013). This confusion in utilisation is illustrated by trainees commenting that some assessors *“signed it (WBA) off with no useful feedback at all”* while an assessor acknowledged that *“on occasions trainees don't get structured, constructive feedback other than when doing WBAs, and even worse to see WBAs completed with no feedback from the trainer”.* This threatens not only the educational value of these assessments but also their reliability and validity. For example, the RCPath guidelines on WBAs show that their purpose is both summative and formative which is reflected in the scales on the forms to rate trainee performance (summative) and also a free-text box at the bottom of the form intended for feedback (formative). This dual purpose may give rise to confusion when completing forms and help to explain a trainee comment that they *“…think assessors are not consistent when completing WBAs and seem to have different understandings of how to complete them”.* Related to this, it has been shown that staff can be reluctant to document unprofessional behaviour for various reasons, which can stem from internal conflict between their mentor and assessor roles (Boenink et al. 2005; Van Mook et al. 2007). Subjective feedback such as this will clearly affect the assessment’s validity.

An assessment is only valid so long as the way it is administered and the inferences made from that assessment are appropriate (Kane 2006). A recent review found a widespread negativity towards WBAs as learning aids with one of the main issues that needed to be resolved being a perceived lack of purpose for WBAs (Massie and Ali 2015), a point which was also highlighted in our results. If WBAs in histopathology are intended to be formative, then one might argue that there is no reason to have a certain number completed and available at an ARCP. This is supported by a respondent comment that *“it's more important to do fewer but more worthwhile assessments”.* One systematic review also questioned the impact these assessments have on performance:

Considering the emphasis placed on workplace-based assessment as a method of formative assessment, there are few published articles exploring its impact on doctors’ education and performance (Miller and Archer 2010, p. 1).

If WBAs are intended to be summative (which the curriculum also suggests), the results from this survey suggest the outcomes assessed are problematic. For example, if WBAs are not perceived to identify diagnostic competence (Q4), it is not surprising that respondents also felt that they won’t be able to identify struggling trainees or reflect trainee progress (Q5 and Q6) either. That is because competence is linked, in part, to these outcomes too. The literature shows that some trainees feel WBAs are not a true reflection of their capabilities which is echoed in a systematic review concluding that there is a paucity of evidence to show WBAs can identify competence in surgical trainees (Bindal, Wall and Goodyear 2011; Torsney, Cocker and Slesser 2015). Regarding struggling trainees, one study found that 16.6% of struggling trainees were identified using WBAs while the remainder were merely identified by trainer observations (Black and Welch 2009). It might be argued that the number of WBAs undertaken during each year of training is not enough as they act as mere snapshots of performance. That is because performance with one patient (or diagnosis) does not necessarily indicate performance on other cases (Cohen, Halvorson and Gosselink 1994) and consistency of performance is particularly important in pathology, where errors in diagnosis could harm patients. However, it is important to note that WBAs are not the only tool intended to identify diagnostic competency, struggling trainees and trainee progress and the curriculum does state that the educational supervisor’s report and exam performance are taken into account. However, given that the outcomes of WBAs will inform the educational supervisors report and that there are only three exams during training, in reality WBAs are currently the main source of evidence to document these outcomes.

It is not clear why WBAs do not identify diagnostic competence but, a possible explanation is that the forms do not reflect the knowledge, skills and attitudes that are most relevant to competency in histopathology. This is supported by comments such as WBAs are incorrectly *“orientated to frontline hospital medicine”* and *“that they should be modified for each specialty to make them more relevant.”* The fact that these questions received such negative perceptions is concerning as WBAs are taken to the ARCP (alongside other relevant information) in order to help assess these outcomes. This in turn may explain why some respondents from the questionnaire viewed WBAs as a *“tick-box exercise”* which further consolidates the perception that WBAs are not useful. Analysis of the free text comments (Figure 3.14) also identified that the design, value and utilisation of WBAs are problematic-themes that relate mostly to an assessment’s validity and purpose. Tooke’s report on medical education in general also expressed a similar stance on WBAs, suggesting there is confusion as to their intended role (2008). In histopathology, this is particularly relevant as pathologists work quite differently to their surgical colleagues and have a unique set of skills. It could be argued that surgical specialties have clearer defined tasks which lend themselves to observation whereas the art of histopathological diagnosis is not as easily observed as it mostly occurs inside one’s head. It is difficult to design a WBA for histopathology when there is no evidence in the literature clearly describing the competencies or steps involved in diagnosis. This might also explain why assessors and trainees had a negative opinion on WBA being able to identify diagnostic competence.

Q2 asked if the number of WBA per stage of training is appropriate with no clear cut results as approximately the same numbers of assessors and trainees were divided equally between the options available (see Figure 3.3 and Table 3.2). It is not possible to draw any clear conclusions from this, but perhaps assessors and trainees feel less polarised in their opinion because the underlying validity of the assessments is the bigger issue. However, research in surgical specialties has already questioned the prescribed number of WBAs for each trainee, suggesting lower quotas may be just as reliable for determining trainee aptitude in a given area (Beard 2008; Marriott et al. 2011). Studies have found that WBAs can become tiresome for trainees (Pereira and Dean 2009) which may result in trainees selecting cases retrospectively when a suitable period of “free time” arises. A traineequestioned the need to *“have to reach an arbitrary minimum number”* of WBAs. Although it could be argued doing ‘more’ may increase the reliability of WBAs this comment suggests the educational value is considered more important by trainees. Minimum numbers are also usually linked to summative assessment, which further illustrates the conflict with others respondents thinking WBAs are formative in nature. Training of assessors can increase reliability and so can utilising a range of assessors. To increase the reliability further, WBAs should be viewed with other items of evidence at the ARCP, so that the trainee’s performance can be triangulated. This is also recommended in the histopathology curriculum, but it may be argued that triangulation by itself does not increase reliability, if the pieces of evidence collated (such as WBAs) are not valid in the first place.

The results from Q3 suggest the time needed to complete WBAs is not a major area of concern in histopathology as it has been suggested for other disciplines in the healthcare literature (Bindal, Wall and Goodyear 2011; Massie and Ali 2015). This might be because histopathology cases are provisionally reported by trainees and then the overseeing consultant checks the case, often with the trainee in attendance. These one-on-one sessions are ideal scenarios for assessors to give feedback to their trainees and unlike a lot of other medical specialties, the discussion can be unrestricted as patients are not present. Nevertheless, even if there is less time pressure in histopathology, the actual “cost” in time will be far greater if the underlying validity and educational impact is not being achieved.

Most assessors also felt their trainees chose appropriate cases (Q10). Linked to this, most trainees also felt their assessors were accessible (Q9), engaged and helpful (Q10) when completing WBAs which is encouraging given that many assessors have significant clinical and/or academic commitments. It therefore seems that there is a disconnect between having what appears to be regular and useful training episodes between trainee and assessor and then translating that on to the WBA. Perhaps assessors already know if a trainee is competent or making sufficient progress but the WBA merely facilitates somewhere to document it, rather than the WBA aiding this judgement process by having suitable performance anchor statements. However, our survey suggests there may be issues with how WBAs are administered as well. For instance, an assessor may know from interacting with the trainee on a day to day basis whether they are making sufficient progress and performing well. Some respondents may argue that the WBA merely asks the assessor to form a judgement of performance on cases that are not representative of true performance. One assessor commented that WBAs are *“specially selected which prove absolutely nothing”* and another commented that “*I often observe of trainees completing the majority during the last month prior to submission of ARCP paperwork”*. In line with the literature, there is therefore a need to monitor WBAs and to ensure trainees do not manipulate who and on what they are assessed (Murphy et al. 2009). For instance, a trainee who never picks challenging or unfamiliar cases will always be well-prepared and not have their true learning needs addressed. It has been shown in other studies that trainees sometimes “select” cases for discussion and avoid difficult cases (Mitchell et al. 2011). Not only will this affect the educational impact of the encounter, but it will be a waste of time and resources (e.g. cost) for the assessor and trainee. Here, the acceptability of the assessment tool to the trainee has had a direct effect on its validity.

Given that only half of assessors believed they have had sufficient guidance to complete WBAs (Q9), this could impact on how assessors are engaging with WBAs and their trainees. In particular, the assessment’s validity will be reduced if assessors are not clear on what they are meant to be assessing. Studies have shown that appropriate guidance for assessors was essential for successful WBA implementation and comprehensive training is required for both trainees and trainers (Ryland et al. 2006; Kirton et al. 2013). This may be particularly important when assessors have had limited training on completing WBAs for trainees and are unsure about how to provide useful feedback. Poor quality feedback would therefore reduce the educational impact of the encounter. For instance, WBAs in histopathology use phrases such as “meets expectations” when rating performance. Clearly, without proper training, the word “expectations” has many connotations and could be interpreted a number of ways. For example, is the expectation related to the year of training, how the individual compares to his peers or does it stem from one’s own personal expectations?

Thus the data from this survey is commensurate with the findings from other specialities, and suggests that current WBA’s used in histopathology do not always meet the criteria of a useful test. They are perceived as being applied variably, produce results with questionable validity and have little apparent educational value.

There appears to be a negative perception of WBAs especially with regard to their capability in identifying diagnostic competence, struggling trainees or trainee progress. Our data shows that WBAs are viewed as being regulatory with their purpose and validity questioned. These opinions are shared by both assessors and trainees. There are conflicting opinions from this study which illustrate that the perception of WBAs is a complex issue which needs further exploration. In concordance with the literature our data shows that designing WBAs which are robust, feasible and challenging is difficult as they must satisfy a range of parameters and users (Prescott et al. 2002). Engaging with assessors and trainees when designing and implementing future WBAs therefore seems a pragmatic approach to try and deal with these issues.

Given the current emphasis placed on using WBAs during training, these results suggest that further work on this topic is required to unravel the issues raised. In particular, the validity of the current WBAs needs to be researched so assessors and trainees know what the assessment is and is not testing. To this end, most assessors and trainees believe WBAs do not identify diagnostic competence, which in turn, explains to some extent why they also thought they could not identify struggling trainees or identify trainee progress. Given that the literature review in chapter 1 was unable to define diagnostic competence or how consultants judge it in their trainees, perhaps it is not surprising that the questionnaire also reflects that WBAs are not perceived to be able to identify it.

## 3.7 Limitations of study

Only a limited range of questions were included in the questionnaire. Further questions may have been able to offer more insight but the research team were aware that more questions could have deterred responses, and given the fairly low response rates, this may have been justified.

The conclusions drawn from the free text comments and questionnaire cannot be generalised to any great extent. Although many issues were raised, the low response rate may have introduced responder bias by selecting a group of respondents that had a greater interest in WBAs (either a positive or negative opinion). Those individuals who were apathetic to WBAs may not have responded. However, the questionnaire does offer insight into the issues of using WBAs and begins to form a context in which competency is judged in histopathology.

Finally, the free text comments are not ideal for qualitative analysis where more thorough lines of enquiry and interpretation are required to explore key issues. For instance, focus groups and/or interviews with trainees and consultants would generate much richer data to achieve this-something that will be used in the next chapter (Chapter 4).

## 3.8 Conclusion

The GMC has stated that WBAs are a summative tool used to assess “competency” and they are used, in part, for the same purpose in histopathology training (General Medical Council 2011; The Royal College of Pathologists 2015). However, the wider literature (Bindal, Wall and Goodyear 2011) and the results from this questionnaire suggest that their effectiveness, educational value and the numbers required to demonstrate competency, requires further evaluation.

It seems reasonable to assume that a good starting point for improving the perception of the current WBAs is to understand the underlying qualities (knowledge, skills behaviours) which are relevant to diagnostic competence and the context in which they are judged. Qualitative research techniques are being used increasingly in healthcare to elicit people’s understanding and the unique meaning they attach to their experiences and actions (Greenhalgh and Taylor 1997). Qualitative research is exploratory in nature and can be used to elicit what qualities are relevant to diagnostic competence and the nature of the judgement ecology which underpins how consultants determine competency in trainees. To this end, although the questionnaire has begun to conceptualise the “judgemental problem” and part of the “judgement ecology”, qualitative methods are needed to fully explore diagnostic competence and will be utilised in the next chapter of this study.

# **Chapter 4**

# **A qualitative interview study using Framework Method to explore diagnostic competence in histopathology**

**Chapter 4: A qualitative interview study using Framework Method to explore diagnostic competence in histopathology**

## 4.1 Background

In the previous chapter, the questionnaire results showed that the majority of assessors and trainees felt that the current WBAs do not identify diagnostic competence in trainees. This is potentially problematic as WBAs are taken to the ARCP (alongside other evidence) to determine trainee competency. It is not clear why WBAs do not identify diagnostic competence but a possible explanation is that the forms do not reflect the knowledge, skills and attitudes that are most relevant to competency in histopathology.

Understanding what diagnostic competence encompasses and the judgement ecology which pertains to this might help to create an evidence base to design future assessment tools. The literature review (Chapter 1) showed that there is no definition or clear description of what “diagnostic competence” involves, but the wider literature suggests medical competence in general can have many meanings, perspectives, interpretations and implications (Ilott and Murphy 1997). Qualitative research methods are for exploring these complexities as they provide descriptive information which can be interpreted to aid conceptualisation and understanding of phenomena (Murphy et al. 1998a). Unlike quantitative research, categories and theory emerge from the data rather than numbers and averages (Bryman 2008). This type of interpretation helps to make the world more visible (Denzin and Lincoln 2000).

In chapter 2, the theoretical background for this study was discussed, including the need to fully understand the judgement ecology and cues that underpin diagnostic competence. As such, to understand “diagnostic competence” the underlying concepts should be explored with a suitable method. To this end, semi-structured qualitative interviews have been chosen as their explorative nature should help to unearth these concepts. They will also, alongside the questionnaire, help to achieve the first three objectives in Cooksey’s framework which steers this study:

* Conceptualise the judgemental problem
* Understand the judgement ecology
* Identify relevant cues
* Sample cue profiles
* Sample participating judges
* Obtain judgements
* Capture individual judgement policy

## 4.2 Aim and objectives

**Aim**

To explore diagnostic competence in histopathology by means of a qualitative interview study.

**Objectives**

To use a Framework Method to explore and understand:

1. What diagnostic competence encompasses in histopathology
2. The cues consultants look for in their trainees when determining diagnostic competence
3. The judgement ecology which pertains to determining diagnostic competency in trainees

## 4.3 Method

The Framework Method by Ritchie and Lewis (2003) broadly fits into the technique of thematic analysis where codes are identified within interview transcripts and organised into themes or categories (Ritchie 1994). It was developed at the UK National Centre for Social Research and is particularly utilised in applied social policy research.

Semi-structured, open-ended interviews were chosen to use alongside the framework method as they provide a useful balance in covering “necessary topics” that are relevant to the construct, while giving participants flexibility in their responses to offer their own insight into it. The “necessary topics” included elements from the literature review which were believed to relate to diagnostic competence and could be explored in the interviews. These were:

1) competency is assessed in stages (Chapter 1, section 1.2.9)

2) competency involves being accurate in diagnosis. (Chapter 1, section 1.2.10.2)

3) competency involved a diagnostic process (Chapter 1, section 1.2.10.3)

Framework method involves creating a chart or matrix, with rows (cases) and columns (codes). This allows the researcher to see the individual views across the different themes, but also how each participant contributed to each theme. This allows for comparison and contrasting across cases and within individual cases. It is fairly structured and systematic in the way data is organised which makes the analysis explicit and capable of being assessed by others (Pope, Ziebland and Mays 2000). However, the researcher must also show flexibility and openness with the data, without trying to quantify it, so that the analysis still reflects the original accounts and observations of the people studied (Ritchie 1994).

Thematic analysis is performed through the process of coding in six phases; familiarization with data, generating initial codes, searching for themes among codes, reviewing themes, defining and naming themes and producing a final report.

A theme represents an element of the data which is relevant to the research question, often something that keeps recurring or gives meaning to the data (Braun and Clarke 2006). Ideally, a theme will be present a number of times within the dataset, but more instances does not equate to that theme being more crucial (Braun and Clarke 2006). Linked to this, the “keyness” of themes is also worthy of explanation. Keyness may capture an important element that is clearly relevant to the research question even if it is only mentioned once (Braun and Clarke 2006). Hence, the final number of themes may not necessarily include those that were the most prevalent in the data but could also be those that had significant “keyness” or importance to warrant inclusion.

The patterns in the data can be identified using a deductive or inductive approach. The latter approach does not stem from applying a particular theory (and its assumptions) but conclusions are derived from the data itself (Patton 1990). This may produce a richer description of the data as it is less open to personal biases or pre-conceived ideas. A quasi-deductive approach was used in this study as the literature review suggested that diagnostic competency was, for example, likely to be assessed in stages according to RCPath guidelines and there were clear research questions to answer. However, an inductive approach was also utilised to explore the concept of diagnostic competence in order to permit unexpected themes to emerge from the data.

### 4.3.1 Participants

The target population were consultant histopathologists who had experience using WBAs and training trainees. These individuals were chosen as they complete the majority of WBAs (where aspects of competency are judged), and may also be involved in appraising trainees and determining ARCP outcomes. The participants, therefore, were initially purposively sampled from an opportunistic sample of consultants who worked within Sheffield Teaching Hospitals NHS Trust. Several consultants from outside the region and two trainees were also interviewed towards the latter end of the process to ensure saturation and triangulation of emerging concepts.

The sample was chosen to encompass as wide a view as possible in order to fully understand the concepts being explored. Therefore, the sample included:

* newly qualified consultants (<3 years’ experience) who had recent exposure to the training environment
* experienced consultants (>3 years and some with 20+ years’ experience) who trained in a different era and whose perspective might differ
* current or previous training programme directors who had breadth of experience and knowledge of the ARCP process
* current or previous educational supervisors who had experience of assessing trainees and gathering evidence for the ARCP
* NHS and academic consultants to offer insight into clinical and academic qualities that might be part of diagnostic competence
* histopathology trainees (for further insight into the judgement ecology)

This ensured that participants represented a range of demographic factors and the perspectives of all stakeholders (Ritchie and Lewis 2003).

The study had ethical approval from the University of Sheffield Ethics Committee (ref. 008042) Recruitment of participants was via email from the researcher, introducing the project and attachment of a participant information sheet (see Appendices 3 and 4). A cooling off period between receiving information on the study and providing consent was used to ensure voluntary participation. A meeting was held with those participants who were willing to be interviewed, so any questions could be answered. They were reminded that participation was voluntary.

Anyone who declined to be interviewed or did not provide informed consent was excluded from the study.

The study was eventually composed of 13 participants, summarised in Table 4.1 below.

|  |  |  |
| --- | --- | --- |
| Gender | 8 males | 5 females |
| Age range | 30-63 years | |
| Background | 11 consultants | 2 trainees |
| 5 past or current TPDs |  |
| 11 past or current ESs |  |
| Mixture of OMFP and general pathologists | |
| Years as a consultant  Range: 1-29 years  <3 years, n= 3  >3 years, n = 8 | |
| Geographic spread | Participants were from throughout the UK | |

**Table 4.1. Summary of participants who took part in the study**

### 4.3.2 Interview guide

A semi-structured interview guide was used to ensure key areas were covered during the interview whilst ensuring participant’s views were accurately reflected in the data (Ritchie 1994), (Appendix 5). The interview guide was tested in a pilot interview to ensure its efficacy but no modifications were suggested. During the interviews that followed, the guide was amended using constant comparative analysis to explore new leads or themes that had emerged (Glaser and Strauss 1967).

### 4.3.3 Procedure

Interviews were conducted in quiet room chosen by the participant, usually their office. Telephones were temporarily switched off and a “Do not disturb” sign fixed to the door ensured no interruptions occurred. Participants signed a consent form (Appendix 6) confirming they were happy to be part of the research study and to be interviewed. Each recording and transcription was given a reference number, stated at the start of each recording. The key to these reference numbers was stored in a locked drawer to which only the researcher had access. Participants were not identifiable from either the recordings or subsequent transcriptions. Transcriptions were fully anonymised.

Interviews were recorded on a digital Dictaphone and field notes were written on the interview guide as lines of enquiry or interesting points were raised during each interview.

The Dictaphone and field notes were stored in a locked drawer. Digital recordings were stored on a password protected computer and backed up on an encrypted USB stick. All passwords were only known by the researcher.

Participants were reminded that they could withdraw from the study at any point as detailed on the consent form. Participants were also given contact details (including name, email and telephone number) on the information sheets should they wish to complain or seek further information about any part of the study.

Interviews were undertaken until saturation occurred. Put simply, this is the point where the most recent interviews offer no further insight and produce no new data (Bryman 2012).

### 4.3.4 Transcription

Interview recordings and field notes were transcribed verbatim as quickly as possible following each interview by the primary researcher (Daniel Brierley). This meant the data was analysed contemporaneously with emergent themes fresh in the memory.

Transcription was undertaken according to the data governance and ethical protocols described by the University of Sheffield Ethics committee. This ensured probity of the research and the protection of the participants.

### 4.3.5 Analysis

Transcripts were organised into question and answer pairs within a table to aid labelling of information and analysis. A very brief framework was created from knowledge gained from the literature review and RCPath guidelines on training. The initial framework had three major themes: “Stage of training” and “Diagnostic accuracy” and “Process”.

Thematic Framework Analysis was then conducted in five stages as per the descriptions by Ritchie and Lewis: familiarization, identifying a framework, indexing, charting, and mapping and interpretation (2003).

Stage 1: Familiarisation: the transcripts and field notes were read multiple times to identify key concepts and recurrent themes.

Stage 2: Identifying a framework involved organising the data in a meaningful and manageable way so that the later stages of analysis were possible. The framework evolved as themes emerged from the data, partly through the previous familiarisation stage and with a degree of trial and error.

Stage 3: Indexing involved systematically applying the framework to each transcript

Stage 4: Charting involved summarising the indexed data for each category followed by organising these summaries into chart form. This format facilitated the data analysis that followed in stage five.

Stage 5: Mapping and interpretation: patterns were identified and sense made of the data from the perspectives of the research questions. This concerned exploring the relationships between the themes to build a framework which explained the data.

The analysis was a dynamic and continuous process which took data from the most recent interview transcripts to help explore any new emergent themes in subsequent interviews.

### 4.3.6 Reflexivity

The researcher had been a specialist trainee in oral and maxillofacial pathology so had considerable experience of the training process. The researcher attended a one day qualitative interview training course before data collection to improve interview technique. The course was run by experts in qualitative research and organized by the Social Research Association. All transcripts were also read and analysed by the whole research team for further input and amendment where appropriate.

The interview guides and subsequent alterations were checked by a more experienced member of the research team (SZS). To ensure breadth and depth of perspectives were reflected in the data, the participants were purposively sampled and reviewed after every third interview.

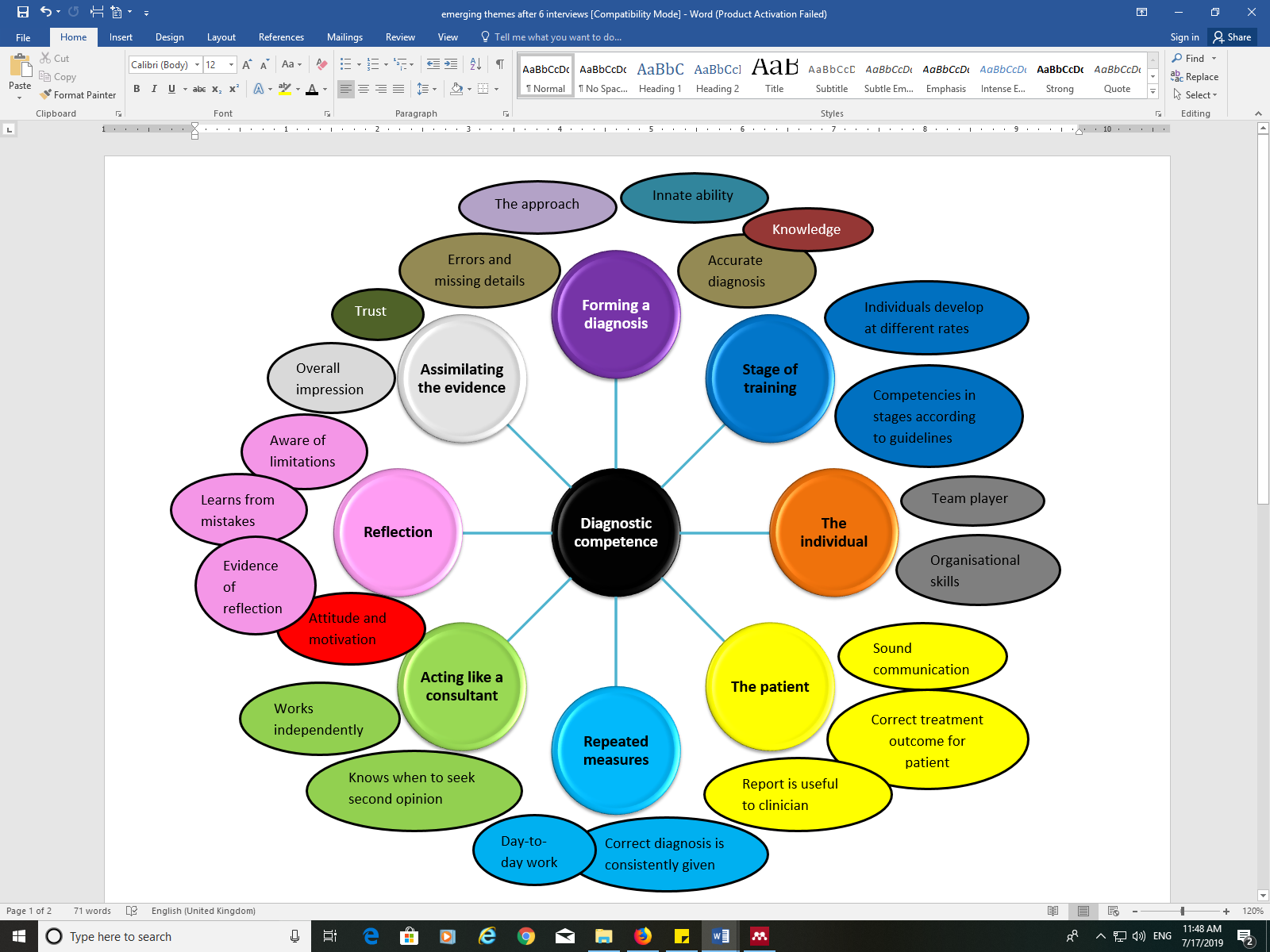
The research team met after every 2-3 interviews to check the transcripts and interpretation of the emergent themes and sub-themes, as this exercise helped to validate the data analysis. For example, as different individuals may see different things in the data or change their perception of the data as themes emerge, these discussion can encourage thoroughness and interrogation of the data (Barbour 2001).

## 4.4 Results

Thirteen interviews were conducted. No new themes emerged after interview ten. The initial framework was changed several times as data were analysed and through an iterative process, all versions were recorded. To ensure consistency, transcripts were re-read and cross-checked each time a modification was made. The transcripts were shared and discussed at each stage of the analysis with the research team to ensure validity of the framework. The raw data and framework are provided on disc 1.

The two initial themes identified from the literature review (“Competency is assessed in stages”, “Competency involves being accurate in diagnosis” and “Competency involves a process”) became 22 sub-headings following 6 interviews arranged under eight categories:

* “Repeated measures”,
* “The patient”,
* “The individual”,
* “Stage of training”,
* “Forming a diagnosis”,
* “Assimilating the evidence”,
* “Reflection” and
* “Acting like a consultant” (Figure 4.1).



**Figure 4.1. Categories and sub-headings after six interviews**

Following discussion with the research team, the 8 categories were reorganised to 6 and the 22 sub-headings reduced to 19 (Figure 4.2).

The six categories were:

“Stage of training”

“Forming a diagnosis”,

“Reflection”,

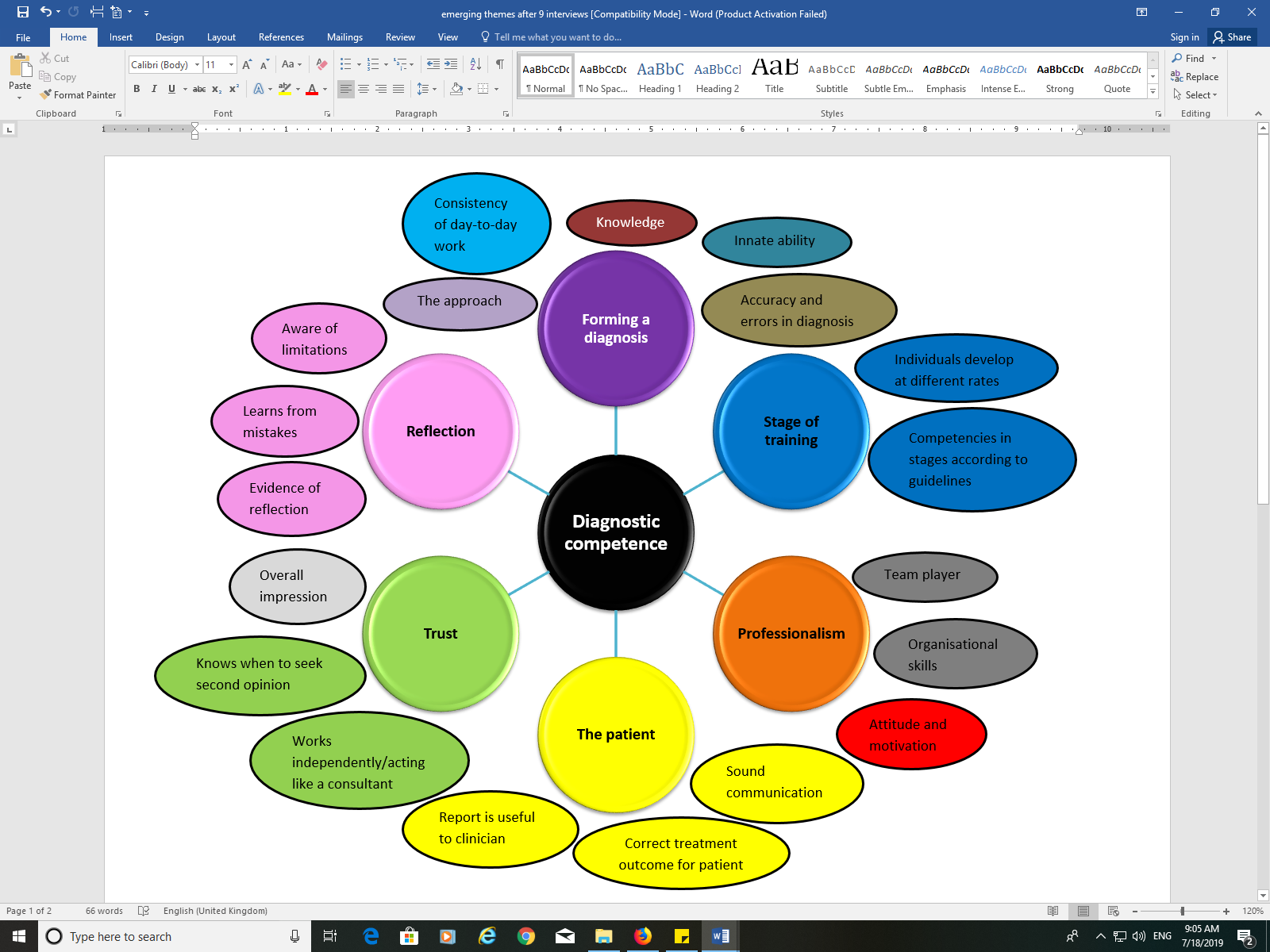
“Trust”,

“The patient” and

“Professionalism”.

The categories were reduced from 8 to 6 because it was felt two categories were conceptually linked and could be merged. To explain:

* The category of “Repeated measures” present in the first iteration included the sub-headings “Day-to-day work” and “Correct diagnosis is consistently given”. Essentially, these related to trainees consistently getting diagnoses correct in their day-to-day work. These two sub-headings were therefore merged to form the sub-heading “Consistency of day-to-day work”. This new sub-heading was moved under the category “Forming a diagnosis”. The category of ‘Repeated measures’ was removed.
* The category of “Assimilating the evidence” included the sub-headings “Trust” and “Overall impression”. On reflection it was decided these headings had similarities with the category “Acting like a consultant” which included sub-headings “Working independently” and “Knows when to seek a second opinion”. To simplify, the categories “Assimilating the evidence” and “Acting like a consultant” were removed and their sub-headings were now placed under the category “Trust” which unified the concepts.
* The category “The individual” was re-conceptualised as “Professionalism” because it was felt that the sub-headings “Teamwork” and “Organisational skills” were aspects of professionalism. In addition the sub-heading “Attitude and motivation” (within the “Reflection” category) was moved to the “Professionalism” category.
* The sub-headings “Correct diagnosis” and “Errors and missing details” were merged to form one sub-heading “Accuracy and errors in diagnosis” as they were felt to represent the same idea, just expressed in a different way. They remained in the category “Forming a diagnosis”.



Finally, after 13 interviews, the data was re-examined in order to determine how the different categories related to one another and if they naturally formed certain groups (Figure 4.3). 

This resulted in 5 major themes as described below.

**The Process**

This incorporated the categories of “Forming a diagnosis” and “The patient”. These two categories were conceptualised under “The Process” as forming a correct diagnosis was related to ensuring patients received the correct treatment. Therefore, ‘The Process’ essentially involved taking the necessary steps to ensure errors were minimised and correct diagnoses were consistently given for the benefit of the patient.

**The Person**

This incorporated the categories of “Reflection” and “Professionalism”. These two categories were conceptualised under “The person” as they both related to personal qualities which, although partly linked to other categories such as “Forming a diagnosis”, they also incorporated more general aspects of professionalism and reflection which would help trainees to improve and learn from their mistakes and act in a professional manner.

**Trust**

Trust was present within the previous coding but also represented a major theme within the data. Trust was built as consultants spent time with their trainees; time needed to observe the “person” and “process”. It was often exemplified by consultants delegating tasks and allowing trainees to independently report. Trust had significant ‘keyness’ to warrant it being a standalone theme rather than being incorporated into professionalism.

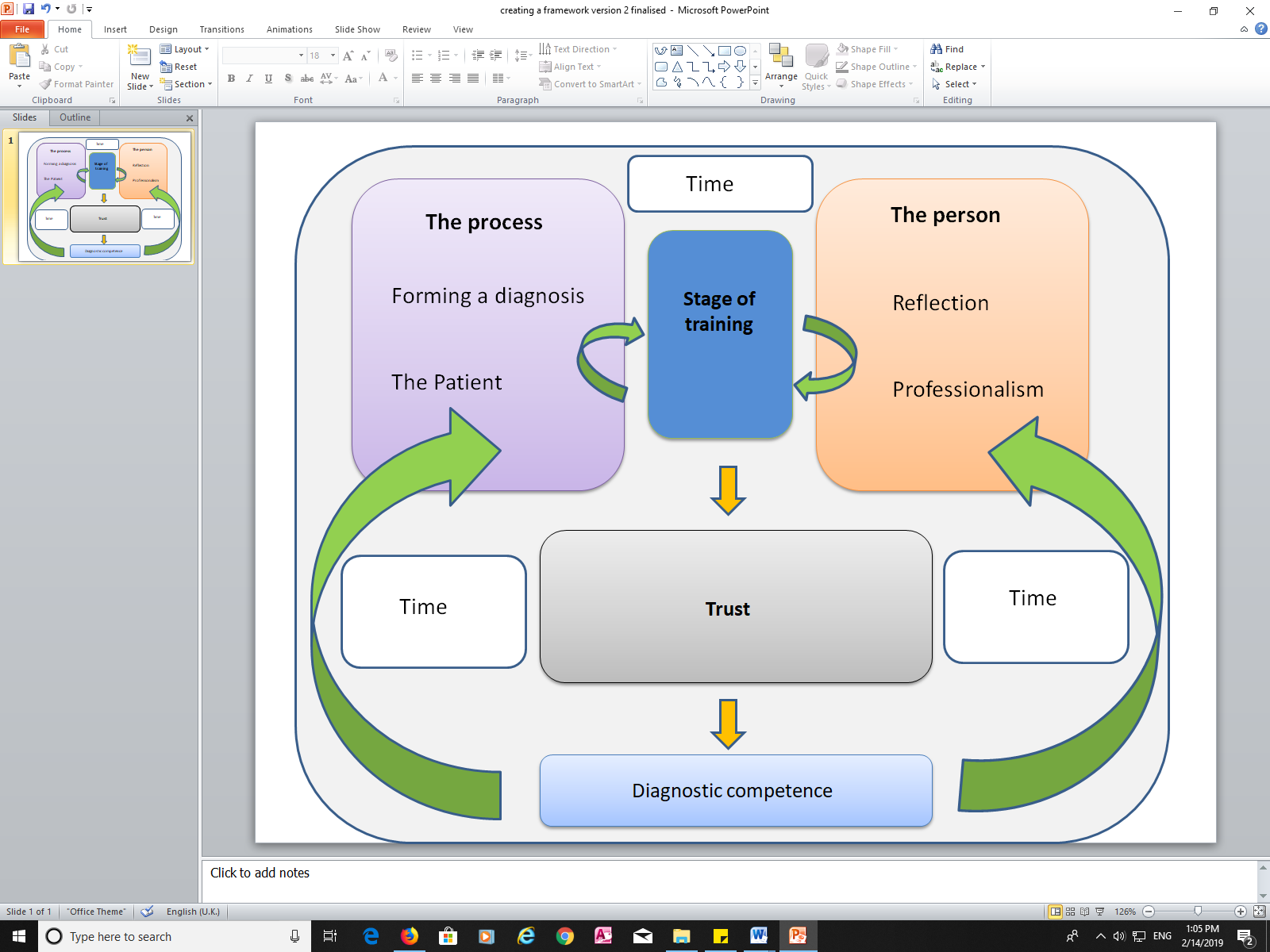
**Stage of training**

“Stage of training” also formed a category within the previous coding but also represented a major theme within the data. Decisions about “process” and “person” were dependent on the stage of training, such that diagnostic competency was stage-dependent

**Time**

“Time” emerged as a major theme only after the four aforementioned themes were examined. For example, interviewees did not state that time was part of diagnostic competency itself, nor was it specifically coded within the data, yet it was actually manifested in all of the previous sub-heading and categories and helped to harmonise the data. Essentially, the judgement of diagnostic competency was longitudinal in nature as time was needed to assess its many elements.

The themes were further conceptualised into a proper framework to explain the judgement ecology (Figure 4.4). The individual codes shown in Figure 4.3 have been removed from Figure 4.4 so the framework is not visually over-complicated. However, they are described further in sections 4.4.1 to 4.4.5.



**Figure 4.4. Conceptual framework derived from analysis of interview data to explain how consultants judge diagnostic competence in trainees.**

To summarise, diagnostic competence emerged from the data as a complex concept, constructed from many different elements which had significant interplay. In order to make a judgement of diagnostic competence, a suitable period of time must be spent with a trainee; along with a period of interaction in which the trainee must demonstrate both processing and personal characteristics which are relevant to the art of histopathology. As trainees progress through training, there is an expectation that their diagnostic competence improves, such that by the end of training, a consultant can trust their trainee to be an independent practitioner.

The longitudinal nature of judging diagnostic competence links to both the personal and processing qualities and how they develop during training. Ultimately, the trainee should reach a level of diagnostic competence in which they can be “trusted” to act like a consultant.

The framework and anonymised quotes from the interviews will be used to detail this further. A bracket after each quote is provided to show which interviewee it relates to. To this end, participants are numbered 1-13, with the number after the decimal place indicating the exact position that quote can be found in the original interview transcript.

### 4.4.1 The Person

The analysis of the data showed that there are several personal qualities which are important to diagnostic competence. These can be clustered into two sub-themes:

* Reflection (comprising the codes “aware of limitations”, “learns from mistakes” and “evidence of reflection”).
* Professionalism (comprising the codes “organisational skills”, “team player” and “attitude and motivation”).

Reflection

The reflective process began with consultants looking to see that their trainee was “aware of their limitations’:

*“…and they will also hopefully become aware of what it was that actually led them coming to an erroneous conclusion” (7.21)*

*“Having an element of insight, a degree of insight, is a very important quality in a pathologist”. (9.15)*

*“It's having the courage to not know something. So when I want to, so when I think a trainee is competent is when I know that they've got the courage or confidence to admit when they don't know.” (8.21)*

This meant trainees would know when to refer or report safely, which links reflection to theme of “acting like a consultant” and “professionalism”:

*“You might not be able to reach a diagnosis. So I suppose competence would also be erm sort of having an awareness that you're not going to be able to reach the diagnosis yourself and handing it over someone else in a different specialty or something like that?” (11.5)*

*“Knowing again, it's coming down to knowing your limitations, knowing that they pick up a case and knowing that they find that sort of case difficult and appreciating that they don’t have to report it there”. (13.22)*

The ability to reflect on your performance was important so trainees “learnt from their mistakes”, which was also part of being “professional”:

*“You do something once is acceptable, do something twice is careless, doing something three times is negligent.” (7.20)*

The link between “reflection” and” professionalism” was also highlighted elsewhere:

*“I look to their reaction to the criticism to whether they are listening to it and taking it seriously…” (3.39)*

This is further supported by comments which suggest there is an expectation that trainees take ownership of their learning:

*“I'm a believer of self-directed study.  That is how it is designed, that is what I encourage…we are not in primary school anymore.” (4.27)*

*“They will go away and they will be self-directed” (2.26).*

*“The most productive times of my life are times when I have had time to think on my own.” (3.32)*

When trainees showed “evidence of reflection” it helped to build “trust” between the consultant and trainee:

*“So you know that so when you get to that point you can say "right you can go and report on your own now". (8.87)*

“Evidence of reflection” was part of what constituted safe practice:

*“I think it’s part of developing as a safe practitioner” (7.43)*

*“I remember somebody saying to me that actually you don't need to be particularly good unless you know when to refer” (5.31)*

*“I think generally a lack of self-awareness then leads into a lack of safety in your reporting”. (13.18)*

Reflection then helped to influence diagnosis-making and how it was expressed:

*“But also having the insight to be able to recognize when I can't make a diagnosis or when there's a grey area and I have to be able to express that ambiguity”. (13.4)*

*“It's also important for the pathologist themselves to feel like they're working within their own kind of realms of what they're, what they're comfortable reporting and what they're comfortable giving an opinion on”. (13.15)*

*“I think it's a spectrum of… of how… how confident you are. So, so part of that competency comes down to confidence and having the confidence to know that you are right or that you're probably right. You need some immunohistochemistry to help or that you might be like to show it to a colleague or that you actually don't know what this is at all. So I think there's this spectrum of confidence and I think that's… that's what kind of ties into this. This idea of ambiguity and then how… how confident you are around those areas”. (13.11)*

This cycle of reflection and improvement was considered highly valuable:

*“The challenges and the mistakes are the things that we can actually gain the most from”. (7.58)*

*“It's not something to penalise them for to be honest…you know, I've made mistakes and from that I've learned a lot.  You learn a lot from your mistakes”. (6.46).*

*“So I've definitely learned more from cases that I remember because I got it very wrong or I completely missed the point of what was going on” (13.12).*

And the fact that following qualification, reflection ensured that you remained a safe practitioner:

*“It has got a role in maintaining that competence because medicine changes all the time.  The types of biopsies that we get in histopathology are constantly evolving” (3.33).*

*“It’s not that you know everything or you can make every diagnosis. But also it’s knowing what you don’t know and also being able to refer err difficult cases…” (1.5)*

*“I think that is a safe part of anybody's practice, you know, an understanding of your limitations and where you reach the end of your diagnostic competence (7.43)*

*“I think without that* (reflection) *there's the risk that you reach a level of competence. Probably defined by having your part 2 and getting your CCST…and then potentially just staying at that level of competence”.(13.21)*

Personality characteristics linked to professionalism

Analysis of the data revealed several personality traits which clustered around professionalism. Sometimes the word “professionalism” was directly used:

*“To sign off a trainee, not only do they have to pass the exam, they have to pass their competences and show their professionalism”. (4.40)*

In other instances, professionalism related to certain behaviours that were described, such as “organisational skills”:

*“I look at the approach to cut up, in how they have prepared, how they have dressed, how they have protected themselves, how they've read the data, prepared themselves before they have looked at any tissue”. (3.34)*

*“Well then there is the organisation of the surgical work and the slides - getting them into a good order”. (3.37)*

*“But also working with a clean workstation where everything has been double-checked so when you look at a microscope slide you know that slide correlates to that piece of paper which correlates to that patient, so everything joins up. (8.103)*

“Organisational skills” involved being efficient and prioritising workload:

*“So I am looking for people who get the complex cases and prioritise them towards the front”. (3.35)*

*“And how proactive they are”. (10.21)*

*“You know, have you thought about the fact that perhaps you did have that sat on your desk for a couple of days longer and did not finish it off”. (7.57)*

*“Well, no, it's not just that they can diagnose it but have they done that in a suitable time frame as well”. (12.35)*

*“So it's… it's not just about being able to get the work done to the right level of correctness. I think it's about being able to get enough of that work done in a given time period as well. (13.13)*

Being a “team-player” was also described:

*“There are people who are diagnostically competent who move job many, many times due to falling out with their colleagues.” (3.15)*

*“Because they will fall out with all their consultant colleagues because they have no mechanism for resolving disagreements or conflict. (3.16)*

*“You know, being able to work within a team.” (9.36)*

*“So I think teamwork and ability to work with people is pretty important. If you can't work in a team and you can't trust and respect your colleagues it means that you're probably going to hide errors. (8.127)*

Having the right “attitude and motivation” were important personality characteristics that consultants looked for:

*“I’m probably looking more for enthusiasm and drive and interest, and engagement”. (10.35)*

*“I think it's things like someone’s passion, their drive, their enthusiasm about a particular speciality. How much they're willing to apply themselves to a certain speciality.” (9.12)*

*“It's obviously the dedication, the aptitude and how much they are going to put into their day to day work”. (6.3).*

Conversely negative “attitudes and motivation” which related to professionalism were also given, such as avoidance behaviours:

*“The traits that you tend to see in them is a bit of avoidance”. (7.22)*

*“They will not necessarily be happy just to sit round a microscope and actually have their direct thoughts on a particular diagnosis that they have made challenged”. (7.22)*

Or, individuals who responded poorly to feedback:

*“The problem is not with actual diagnostic ability; it is to do with taking criticism and being able to modulate their own behaviour”. (5.29)*

*“And they won't listen to others, their opinion or the way they are thinking.  I think that's a difficult trait to deal with and we have had people who have done that and they won't change from what they think is right”. (6.58)*

*“Where you explain that what they've done is inaccurate or in cut up, if the blocks taken are not high enough quality and they don't seem to understand, that and don't want to understand how to improve then that would be a red flag in terms of competence”. (13.52)*

This sometimes resulted from a lack of reflection, affecting both professionalism and diagnostic accuracy:

*“And I think that is one of the most important areas of competence, the, the unsafe trainee is one who confidently makes a diagnosis on virtually every case”. (8.22)*

*“I think being able to take feedback on board and reflect on it and to act on that feedback is a good important part of professionalism because again it goes back to what we were talking about, insight…then there's a lack of insight in there which affects your professionalism which could become patient safety issue”. (9.22)*

*“If you see a trainee is disorganised and got it all over the place that affects their diagnostic ability because they couldn't put things in order” (4.6).*

*“I suppose that's because you're then starting to understand that person doesn't necessarily have a kind of learning mentality and how are you supposed to improve if you're not willing to learn”. (13.52)*

This was particularly evident in comments which described how personality characteristics affected the ability of the trainee to follow the correct diagnostic approach:

*“I think you know when you…get them coming together in terms of the process that you see somebody developing through years as a diagnostician, it is these particular personality traits that I think are brought together that would make that process as efficient as possible.” (7.90)*

And how this affected trust between consultants and trainees:

*“If somebody won't take advice or criticism then they won't, they'll carry on making the same mistakes…you'll always worry that the higher up they go that they will be independently reporting things which they don't have the competence to independently report”. (5.30)*

### 4.4.2 The process

Analysis of the data showed that diagnostic competence revolved around a process in which trainees demonstrated:

* The ability to form a diagnosis (comprising the codes “knowledge”, “the approach”, “innate ability”, “accuracy and errors in diagnosis” and “consistency of day-to-day work”)
* An awareness of how their diagnosis and report can affect patient care (comprising the codes “sound communication”, “correct treatment outcome for patient” and “report is useful to clinician”.

Forming a diagnosis

An important part of forming a diagnosis was “accuracy and errors in diagnosis”. Sometimes the “accuracy” was stressed:

*“I guess if I had to define it, one, that they get the diagnosis right”. (10.6)*

*“The ability to derive a correct diagnosis…and accurate manner”. (2.3)*

*“The short answer is getting it right and actually being able to define what the entity that you're looking at is.” (7.5)*

*“Well I suppose pathology is about diagnosis, isn't it? So the aim would be to get the diagnosis right not wrong” (11.5)*

But sometimes the need to minimise “errors in diagnosis” was described, although the net effect was similar:

*“Erm, also the amount of the changes you're making to the report” (10.24)*

*“So the the (syc) amount of error we're allowed in our job as histopathologists is actually almost none.” (1.7)*

Part of “accuracy” also related to being in line with your peers’ diagnoses:

*“I guess it's important because that's how we are assessed as consultant as well, so when you become a consultant and you're doing your EQA, that's how your competence is assessed.” (10.14)*

*“Whether you are in consensus with a reasonable body of, erm, experts or peers in that particular area where you practice.” (9.5)*

Accuracy didn’t always mean getting an exact diagnosis:

*“Having the sort of the skills to accurately reach either a definitive diagnosis or erm a good working diagnosis of a disease category or a condition”. (12.4)*

*“So I suppose my short answer is I don't think diagnostic competence is about getting it right all the time.” (13.4)*

The “approach” was sometimes considered more important than getting the diagnosis correct:

*“At least they've got the right approach, erm, what sort of things they are going to do or investigations they might actually use to get to the right answer.” (10.6)*

*“So I want to understand how they got there. I don't mind whether they do it heuristically or whether they do it by gestalt reasoning so they can do it by wallpaper matching…but they have got to enunciate it to me”. (3.7).*

The “approach” sometimes was described using explicit criteria, such as the right ordering tests or information garnered from the macroscopic dissection of the case:

*“The ability to order appropriate special tests.” (8.69)*

*“Yes, yes it is because you can't make a diagnosis without dissecting something properly and so how have they taken on board the data, have they done the dissection competently”. (3.35)*

*“You might get to the right answer you probably would in the end but you'd want to sort of narrow it down and erm sort of take a more logical approach to sort of choosing what tests extra tests - you're doing things like that”. (11.9)*

*“I'm thinking more trimming and sort of physical cut up stuff managing my macro, knowing block selection what's important, how that's going to relate to your reports and things.(11.22)*

Sometimes the “approach” involved clinico-pathological correlation:

*“Importance of clinical findings in the diagnostic process”. (8.24)*

*“If it is a breast report, have they looked at the mammogram, are they looking for calcification or not, have they read the request form, have they understood the data?” (3.34)*

*“I think red flags for me would be for example missing clinically relevant findings”. (1.20)*

*“You look for the ability to integrate clinical data, the macroscopic interpretation of the information available” (2.5)*

This clinical information often helped inform the range of possible diagnoses:

*“They may be more limited in what they're considering a diagnosis could be so I think they perhaps… the breadth of differential diagnosis will be much less and there is I guess the potential that you could miss a diagnosis if you haven't thought about it in your differential then you're not very likely to diagnose it”. (12.1)*

Yet sometimes it was difficult to describe certain elements of the “approach”

*“But I find it very difficult to quantify what that actually is (7.73)*

*“There are certain things that I would just pick up on a slide that almost never reach cortex, and that's difficult from a training point of view because you know I've got to try and vocalise stuff that just tends to go on in my head underneath the surface”. (7.65)*

These intangible elements were sometimes correlated to “innate ability”:

*“Sometimes those at the top end aren't actually very knowledgeable, don't work very hard but they are just capable of doing it, so they have got a talent for it”. (3.18)*

*“When someone's got the eye for it or they've got the right approach it's very difficult to define or record but you know it when you see it”. (8.97)*

*“Because they couldn't interpret… for everything there is also a dimensional issue where understanding how something that's three-dimensional ends up as a two-dimensional image in different colours on a slide. They just couldn't actually erm sort of acquire the basic level skills despite additional help and remedies and they agreed that they, they just... it wasn't for them and they have left...so I think not everyone can do it”. (12.15)*

Linked to this, the way a trainee or consultant examines a slide was part of their approach:

*“How you approach looking at slide, I think so, I don't think it's necessarily done consciously by most people but if I look at most consultants and most experienced consultants, they do a lot of their worker very low power”. (13.13)*

*“But I think looking at a first-year trainee versus a consultant in their approach to how they physically interact with a slide is very, very different and maybe you could do that sequentially over someone's training and if by ST3 someone still wasn't developing that ability to zoom out and think then that allows a targeted training opportunity to make them consciously change a little bit”. (13.14)*

And the ability to identify patterns on slides as a means to refine diagnoses:

*“But I think very much pattern is quite important and I think it's recognizing that pattern as well as individual cell features, there are architectural features. There's a pattern”. (12.13)*

*“You spend a lot of time looking at the details, maybe not seeing the wood for the trees and then over time that gradually zoom out if you like and learn to do more of your work at lower power. One thing that I consciously did that another trainee taught me was if you identified a feature at a high magnification like x40, then go back to x 2 and see if you can still spot it. So I consciously trained myself to try and look for details”. (13.13)*

*“I think some people will never get it because it's a visual special pattern recognition thing”. (3.18)*

*“And my ability at a low power magnification to say “I think it fits into this basket”. So I was almost using like, trying to consciously develop this heuristic approach at low power. “What are the 10 things I need to be thinking about” rather than there are two hundred things”. (13.11)*

The “approach” appeared to facilitate a diagnosis:

*“To derive an answer and then putting it together”. (2.6)*

*“Well to reach a diagnosis you have to develop obviously a way and the mechanism to reach it”. (4.7)*

*“I think, there are, there is a process behind that…” (7.7)*

Consultants examined this “approach” in trainees to gauge their understanding of cases:

*“I will then begin to test them in terms of the basics to see whether they have issues in terms of reasoning or interpretation or even just recognising pathological items”. (2.24)*

*“Whether their approach to something is systematic enough so that they could actually stand up and say you know, when I report a neck dissection, I do this so I know I have confidence”. (7.59)*

It appeared to help consultants understand how “errors” might have occurred and how they may be addressed:

*“Now, if they get the diagnosis wrong, which may happen, but they base their diagnosis on the evidence seen, then we can correct it”. (4.13)*

*“Even if they say oh I thought this looked malignant and it's not, then you say well why did you think it looked malignant, well it's got mitotic figures.  Well can you get mitotic figures in inflammation or other features of inflammation”. (5.26)*

*“Now, if they get the diagnosis wrong, which may happen, but they base their diagnosis on the evidence seen, then we can correct it.  But if they get the diagnosis wrong because they didn't take into account anything, then it is very obvious that they haven't studied correctly that stage…in which case they are incompetent”. (4.13)*

Enunciating the “approach” ensured correct diagnoses did not stem from guessing:

*“Some people can arrive at the correct diagnosis and can't explain why and have arrived there by guessing.” (3.6)*

*“It depends on whether they got it right just by fluke or whether they were actually thinking along the right lines, erm, what's the sort of approach and how they got to the answer”. (10.65)*

“Knowledge” was considered important and was needed to form a diagnosis:

*“But knowledge too I think is important because I think, particularly in terms of disease process, you can't gain everything from just looking at the slides.” (7.75)*

This was sometimes described as basic or core knowledge:

*“You can clearly see the trainee who does not have the biomedical knowledge of course”. (4.34)*

*“I think to start off with you should have a very good basic understanding of general…if you know the normal you can do the abnormal, you know that's the basics of what you need to understand”. (6.1)*

Or, new developments and guidelines that should be synthesised into histology reports:

*“So really you have to be pretty well up on the literature in terms of classification and current techniques for reaching a diagnosis”. (8.24)*

Aspects of professionalism, such as a lack of care, sometimes affected the ability to form a diagnosis:

*“The mindset where errors are virtually eliminated through care and attention.” (10.3)*

*“Perhaps there is no one right way to look at a slide but at least you can see whether the coverage is meticulous”. (3.35)*

Thus, forming a diagnosis was hindered if trainees lacked the right approach and reflective qualities to improve their approach:

*“First of all they don't appreciate the importance of clinical findings and even after you've told them, repeatedly, they still don't do it”. (8.109)*

Forming a diagnosis also meant being “consistent in day-to-day work” as accurate diagnoses had to be given consistency to be competent:

*“When I was doing them as a trainee I just felt like I just got lucky with a few because I only knew so much, but that doesn't make me competent”. (10.12)*

*“But experience, are…are looking and seeing the correct features for yourself and seeing it repeatedly over time and in slightly different cases”. (12.12)*

*“I think there's that the consistency of their competency, it's not a one-off competency. They need to be consistently competent”. (12.22)*

To this end, comments related to WBAs and exams only examining distinct points in time and to assess competency one needed a more longitudinal, representative approach to monitor for “consistency”:

*“So I guess more like a long-term thing and these assessments are just like a snap-shot in time so I'm not really sure how representative they are of competence at that time, exact sort of instant. (10.11)*

*“It is an artificial environment you are working under pressure; you're not necessarily practicing what would you would do normally.” (9.31)*

*“It's the day to day work that you do and that's reflected in your performance… not from any pieces of paper stuck in your file to say, “Oh yes, you've done that” so you've been ticked off for that competence”. (6.55)*

*“Exams that are external validations provide the snap shot but then on an exam environment, you're adding on the performance of the day which is not necessarily reflecting the trainee's abilities and competences”. (4.23)*

*“The chances of coming across the same case, having that same assessment and consultant… so much confounding that I don't think you really get that progression documented in them”. (13.32)*

This idea of consistency of competence was also reflected in comments that criticised trainees for selecting cases for WBAs which might not represent their true competence:

*“The weaknesses of the system (WBAs) are the trainee often looks at things that are sort of guaranteed to give a positive response”. (2.9)*

*“I think the main changes (to WBAs) would be, I guess there is no hindrance to us implementing it in this way it is actually trainer led rather than trainee led and that then would allow us just to avoid that level of pre-selection the trainees have…” (7.39)*

*“They would very rarely choose a case which may result in a negative outcome. Would they?” (8.48)*

An awareness of how the diagnosis and report affect patient care

There was an acknowledgement regarding the pivotal role pathologists have in patient care and hence it is important that the patient receives the “correct treatment and outcome”:

*“Because the buck stops with you then and you've got to be sure that it's benign, not 95%, not 98%, you've got to be 100% sure that it's benign”. (5.24)*

*“But also there's a patient, patient safety overarches everything anyway”. (9.18)*

*“Pressing the button, well you could send something out which is wrong (with errors in) and that could have an impact on the patient. So it's important.” (11.69)*

The diagnosis clearly had an effect on the “treatment and patient outcome”, whether explicitly said:

*“Allow or facilitate, err, the correct or optimal treatment for the patient. (1.5)*

*“Absolutely, what the impact of their reports and what they have done is on the overall management of the patient”. (6.21)*

*“I mean the key factor is if it changes the treatment plan for the patient”. (10.7)*

Or, even if the patient wasn’t mentioned, a risk to the patient was implied:

*“Incompletely excised BCCs, missing positive margins, grossly over calling or under calling melanocytic lesions I would consider serious concerns.” (1.20)*

*“The ability to derive a correct diagnosis in a timely fashion.” (2.3)*

However, the diagnosis at the end of the report was not the only area of importance. The message conveyed in the report had to be “useful to the clinician”

*“They need to convey the information in a useful way to the clinician”. (1.15)*

*“Some trainees have the propensity to produce overly long reports with too much complicated terminology which is of no clinical relevance”. (1.15)*

*“It's getting the message across as well so whether they can diagnose it is one thing but can they write it in a way as well so the users actually realise what they're trying to tell”. (10.66)*

“Sound communication” was recognised as a facilitator to good patient care:

*“The point of the report is to communicate”. (2.21)*

*“But also how you communicate that diagnosis to your colleagues and often times it’s not a black and white diagnosis sometimes there are degrees and layers of what we do and how can you convey that message.” (9.27)*

*“Communication is... it's a very important part because essentially we're communicating our diagnosis via a report. It's not a histopathology report going out it's a diagnosis”. (12.41)*

*“I think communication is probably the key attribute in terms of practicing safely”. (13.21)*

In particular, language within “sound communication” was deemed important when conveying diagnoses:

*“You know, a lot of people I've known have failed it and then passed. It's been a language problem.” (5.66)*

*“I am looking for quality of presentation of the report and language”. (3.41)*

It was acknowledged that the surgeon must “find the pathology report useful” as it is their interpretation that impacts on patient care:

*“I like practical situations like taking a senior trainee to an MDT meeting and seeing how they present the cases, to see whether they have got an empathy with the clinicians”. (3.20)*

*“I want a hierarchy of information so that if there is a piece of major information that has to be conveyed, it's not buried in an essay… that's a failure of empathy with the surgeon”. (3.24)*

*“What matters is what you communicate to the surgeon and you need to know what that means that they are going to do”. (5.40)*

There was a strong link between the sub-themes “forming a diagnosis” and “the patient” as illustrated by comments such as:

*“Does the approach show that what they are doing is safe.” (8.77)*

*“I look at how they've set out the report. I look at how they explain it and how they explain how they prioritise what they have put in”. (3.34)*

*“Because ultimately there's a patient at the end of what we're doing. So I think I think safety and probably accuracy goes hand-in-hand with safety I think you're getting a diagnosis right or giving a report that conveys the right level of ambiguity are both safe approaches to reporting a slide or reporting a case”. (13.15)*

Hence, the need for “accuracy” particularly related to whether the “error” would affect “patient treatment and outcome”

*“For example, pleomorphic adenoma and myoepithelioma maybe doesn't make massive difference, but whether it's a pleomorphic adenoma or a carcinoma in pleomorphic adenoma does make a big difference”. (10.10)*

Sometimes “patient treatment and outcome” was correlated to “professionalism”, as trainees acknowledged their role in patient care:

*“So if you are aware of that early on and you don't become kind of working in this silo where you feel that you're constantly protected by your seniors…then it drums home very early on that there's a patient at the end of what you're doing…you're responsible for your pathology report”. (9.35)*

### 4.4.3 Stage of training

The analysis of the data showed that expectations of diagnostic competence vary depending on the stage of training. Stage comprised the codes “individuals develop at different rates” and “competency in stages according to guidelines/standards”.

There was an appreciation that “individuals develop at different rates”:

*“People develop at different stages”. (1.26)*

*“I think trainees move at different paces, that they move at different paces in different arenas”. (2.13)*

*“So someone progressing slowly, we have had people who progress very, very slowly and in the end they become very useful consultants”. (3.18)*

However, there was a general expectation that trainees should improve with time:

*“No, I think I'd look for more as they go through. So you'd want to see it evolve in certain ways because obviously in the beginning somebody's not going to know what they're doing at all really”. (11.31)*

*“I think again depending on the sort of the year of traineeship…I think… I think the expectations will increase erm as you work through training and become a more senior trainee”. (12.6)*

This expectation was influenced by “guidelines/standards” within the RCPath curricula:

*“Well it's a big difficult to define, but to be able to err be competent in the job at a particular level or standard that has been set by the training bodies involved and the regulators.” (9.4)*

*“So I suppose to measure competence, one way of measuring competence is to say you have a gold standard so it's their ability to reach a definitive diagnosis against the gold standard”. (8.11)*

*“And if you're ST1 and you miss something complicated then they're not going to dress you down for that. They're going to talk to you about it. Whereas if you're a senior trainee and you miss something that I really think you should have gotten they're going to explain to you that, you know, that's not really acceptable”. (13.7)*

*“My own opinion with EQA schemes is that all trainees should do them. Maybe not right from the start, maybe from Stage D”. (13.32)*

And also objective evidence such as exam results and WBA outcomes:

*“There are the exams that keep the goal posts throughout the training and there are also the WBAs which may take the form of a summative assessment but they may be descriptive as well. We use the WBAs to see where about the trainee is in the training and we ask them to produce six of each so we can have eighteen per stage that they can give us that information. ”. (4.5)*

*“The second tool we use is the end of placement assessments and they give a lot of information because they are more descriptive”. (4.5)*

But consultants agreed the need for flexibility when judging diagnostic competency:

*“So you have to look at each trainee as an individual and you have to adapt to the trainee rather than assume that everyone is the same”. (2.7).*

*“You may have a first year trainee actually perform better than a second year trainee.  In which case I admit they are more advanced and as the training comes is individualised so the trainee may sit an exam earlier than another one depending on where they stand and they may start independent reporting earlier”. (4.4)*

*“I don't think you can pin a definitive timeframe on it and it depends on the individual as well. I think some people perhaps pick things up more readily and become more confident more quickly than others. Someone may need to see 10 fibro-epithelial polyps and yeah, I know what that is, I can recognize that or whereas someone may still need to see 20 before they reach the same state so I don't think it's a one fit all type. (12.19)*

Markers of reaching “standards”, such as passing exams, appeared to sometimes influence when responsibilities were given to trainees:

*“I think the safest time to independently report is after FRCPath so at least you’ve got some documentary evidence. (1.22)*

*“But I think post FRCPath in order to err well we have a list of cases pre-CCST types cases that they can report according to Royal College guidelines” (1.6)*

*“They would prefer trainees not to be writing up reports until they’ve reached a level of err competence which they define by exam”. (2.29)*

Having stages in training with prescribed “guidelines and standards” correlated with protecting the patient:

*“For patients' safety you have to have some sort of process behind it whereby if an issue arises you are able to say well we did this and this, so there is some process rather than such and such a person had a feeling that this would be alright”. (7.52)*

*“So, medico legally it is quite useful to have Part 2 as the, as the sort of barrier”. (8.18)*

This was also reflected in comments suggesting a need to have objective criteria when making judgements of competency:

*“But I think that it helps to have a form where you have objective criteria to look at”. (9.21)*

*“I don't believe in gut feelings.  I think they are the products of the evidence” (4.45).*

*“You know when you are signing off that someone is capable against the standards set in that curriculum, you actually do need to have some way of actually objectively showing that”. (7.49)*

### 4.4.4 Trust

Analysis of the data showed that the judgement of diagnostic competence centred around consultants trusting their trainee and comprised the codes “overall impression”, “works independently/acting like a consultant”, “knows when to seek second opinion” and “trust”.

The theme of” trust” tended to be described as a gradual process which developed through training:

*So it’s a case of slowly letting go rather like teaching a child to ride a bike. You run with them alongside, you've taught them how to push the pedals, you've taught them how to steer, you run with them alongside, eventually you let go of the handle bars and you either fall over or they cycle off into the sunset. (2.30)*

“Trust” appeared central to training:

*“Because if you did not trust them you would not be able to give them work”. (6.18)*

*“You've got to trust them”. (3.51)*

*“What I have seen around histopathology and around medicine in general is that people who can't be trusted either in the workplace or in life in general, there is a generalisation of failure of trust”. (3.50)*

However, “trust” tended to be fully realised towards the end of training when an “overall impression” had been built up:

*“You're looking for a safe pair of hands, then you know if they have been through Part 2, then you know they've been over rare cases, they've read the books properly, they've looked at all the black boxes.”(5.22)*

*“What do we assess competence by: they've done the job.”(9.17)*

*“It’s the overall impression, I don’t think you can use only one thing”. (2.27)*

*“So it's not just the diagnosis it's the whole erm, right from the specimen arriving to the report going out. So it's the whole thing. It's a different aspect it's not the same as diagnosis but it is... part of their all-round competence I'd say. (12.39)*

*“I think when you've been in the department for a while consultants get to know you, they know what you're good at. They know what areas you like and what areas you don't and they can take that in context”. (13.8)*

And when trainees began “acting like consultants”:

*“If they can identify hidden traps in terms of making a diagnosis and other information that may be relevant which is not automatically being sought”. (2.6)*

*“Reproducibility comes as they approach consultant level”. (4.8)*

*“To see them be able to handle a case from start to end, sort of independently. Erm be able to make all those decisions with; do you need extra levels, deepers, full face, which test you need and why and then come to their own conclusion and write a full report as a consultant would”. (11.32)*

Or, by consultants feeling that no harm would come to patients:

*“I think trust in terms of competence is actually you relying or thinking that the trainee is again, safe.” (10.53)*

But” trust” also related to feelings consultant’s had about trainees:

*“Quite often people do have a feeling but they cannot pin it down to something”. (4.44)*

Hence, trust sometimes involved combining objective evidence with feelings to come to a judgement on a trainee’s competency:

*“You need to take everything into consideration. Yes. Everything”. (4.43)*

“Trust” was embodied in certain actions such as allowing trainees to “independently report” or requiring less supervision (“acting like a consultant”):

*“In order to trust someone to report independently you have to have faith”. (1.7)*

*“It's the fact that they're able to handle more difficult cases with less and less supervision. Less input from me. I'm there to actually just guide but actually hold their hand and take them to the answer”. (10.44)*

*“Yeah, to independently report…does that have a role, trust. Absolutely. I think if you know a consultant trusts you to do something then you feel that they feel that you are competent if you see what I mean”. (13.54)*

“Acting like a consultant” also related to trainees mirroring consultant behaviours such as having confidence in their diagnoses:

*“Whether they can actually stick with their clear reasoning to a diagnosis”. (2.18)*

*“I make people completely commit to what they would have sent out”. (3.45)*

*“Someone who is offered as a consultant to trust, if people don't believe in them, they are probably not going to succeed”. (3.20)*

*“But I think I would expect somebody who is competent to have confidence with the majority of the stuff they're doing or a lot of the stuff”. (11.47)*

*“There's no point in being right if you can’t explain why and you can’t kind of influence others decisions if they’re not agreeing with you”. (13.38)*

And being efficient with clinical resources and time:

*“There's a cost element that is, you know, if you're going to order more tests, you've got to sit down and analyze the results you burn. Somebody's got pay for them all”. (11.17)*

*“Because the reality is there is an amount of work you have to turn around if you have to spend 20 minutes looking at every slide at high power and crawling over every cell then you're not going to be diagnostically competent because I would argue if you can only report one case in a day then you're no you're not diagnostically competent. Because you're not able to function how a consultant needs to function in the unit”. (13.13)*

*“If you're reporting, say, a normal duodenal biopsy, if you've been spending five times as long to come to the same relatively simple conclusion then I would say that your diagnostic competence isn't as developed as the person who’s found that balance between quality, safety, and speed”. (13.25)*

“Trust” was also manifested in the way some consultants interacted with their trainees:

*“I think you would see that the manner of communication would probably be quite different…in many ways these interactions would be far more trainee led I think”. (7.80).*

*“And we sit round a microscope and I'm almost incidental to it”. (7.80)*

*“I mean part of the feedback is you agreeing the report and signing it out is the actual confidence that you're giving the trainee that you're happy with what they've done and that it's good enough for your name to go on it” 10.24)*

“Trust” involved consultants seeing that their trainee knew “when to seek a second opinion” which is also related to the sub-theme of “reflection”:

*“Then your job is can I report this independently or not and you want to know that they are not over confident in that way”. (5.22)*

*“If they don't tell you when they've had uncertainties or diagnostic quandaries about certain aspects of the slide, then how can you trust them to independently report something if you can't see inside their thought processes”. (5.57)*

*“They have to be able to ask.  Even consultants can get to that point”. (4.37)*

Without seeing this, consultants struggled to “trust” trainees to “independently report”:

*“So if you've got a nagging, you've always got a sort of nagging uncertainty that they would ask you for help when they are unsure, then it's difficult to let go of the reins”. (5.58)*

Consultants appeared to want to hand responsibilities to senior trainees (such as “independent reporting”), and this acted as a surrogate marker for “trust”:

*“Even if it's independent reporting that is checked by a consultant, like Nadine now, she is bringing me stuff that she says, you know, she's done a report or even if you're doing that, that's a higher level of stress and then you need, you need that”. (5.24)*

*“You know, if you throw somebody in at the deep end, you learn to swim, you learn to do your own thing, you learn to take responsibility”. (6.63)*

*“You will be actually signing out reports with your own name on the bottom of it, that process you know, you have to get used to that sort of pressure”. (7.32)*

*“And then there's slow ramping down of the amount of input that consultants would have to…to my reports. So that's a very graded approach to signing things out independently, which I suppose is an assessment of my competence”. (13.47)*

This was deemed important as the training environment was sometimes unauthentic:

*“In other words, training doesn't replicate working as a consultant because it is a highly supervised environment”. (3.12)*

And upon qualification, trainees needed to be used to “act like a consultant”

*“They need you to be a consultant up and running”. (6.63)*

*“It's putting all of the responsibility on you and it's sort of making you… making you step up”. (13.49)*

### 4.4.5 Time

Although time was not separately coded within the transcripts, through designing a conceptual model of diagnostic competency, it became apparent that it tied many of the elements of competency together. As such, it was manifested in the themes and sub-themes previously described, such as the person, process, stage of training and trust.

For instance, reflection was strongly linked to the theme of time as “time” is needed for reflection to occur:

*“It is then down to the trainee to sit down and read.  It is quite obvious if they haven't read it the next time you see them (4.27)*

Reflection was also used by consultant to check if errors were minimised in future interactions with the trainee. This demonstrates the interplay between “reflection”, “time” and “errors”:

*“… and the next time that situation arises that if they have been reflective, if they've gone away and done the necessary building blocks of knowledge that that, the particular issue which has arisen will not arise again” (7.87)*

“Time” was also required by consultants to properly judge whether trainees followed the correct diagnostic approach:

*“I think that is a process that you actually only really get an understanding of it get an insight into it, if you actually spend time with a junior pathologist round a microscope”. (7.9)*

*“But if someone reaches the final diagnosis without synthesising what they have seen in different stages then that worries me”. (4.32)*

*“I think it's quite useful to have a similar dialogue and not just sort of a one sided conversation- a dynamic sort of relationship when working on a case.” (10.18)*

*“You have to spend time with your trainees... it is down to the discussion, it's being able to actually discuss cases round a microscope” (7.69).*

The “time” needed to review the diagnostic “approach” was closely linked to ensuring the trainee didn’t put patients at risk:

*“In the few where he doesn't know the answer, he's referring appropriately i.e. either asking the trainer or showing it round the department, so I'm pretty confident that 100% of the time, I'm confident that that person is safe”*

Sometimes “time” was needed as a means for the trainee to generate experience. Experience, in turn, fed into reflective practice and having the correct “approach”:

*“But gradually with experience you can exclude the ones that it is not likely to be.”(8.9)*

*“Bringing experience to bear on it”. (7.71)*

*“For example nevoid melanomas is well known as a catch and it’s often difficult to spot so unless you’ve got experience you may easily miss something like”. (1.22).*

*“There is experience to go into that” (7.7)*

*“I definitely felt myself getting more competent and that was looking at slides, that was just experience and it was looking at slides getting almost instant feedback from teaching sets where I would have the answer or from an online resource where there might be a differential explanation, but it was that it was that cycle of looking at cases getting something wrong or right reading about that, going back doing it again. So it was that repetitive cycle of to use a pathology phrase “shifting glass””. (13.11)*

By spending “time” with trainees, “trust” was allowed to develop:

*“I think the reason it (independent reporting) doesn't happen in every case before the actual consultant appointment especially in a district general hospital is for that exact reason that the consultants don't know the trainees well enough”. (8.88)*

*“But different people are trained differently, different people erm, approach things differently, so I would like that, sort of, peace of mind thing, spending some time with them before I could actually, I would be comfortable with them signing my reports out.” (10.60)*

*“I think I would have to have sufficient time for, and this I guess comes back into more feeling than paperwork, that you have spent sufficient time to actually understand the acumen, you know the range of mistakes, you know where their weaknesses are”. (7.52)*

“Time” was also needed to assess trainee “professionalism” as time was needed to build a picture of someone’s behaviour:

*“It's obviously the dedication, the aptitude and how much they are going to put into their day to day work.  You can actually see every day who are hard-working and who are not.  Erm, who would cut corners and you know try and get answers from others”. (6.3)*

“Time” was needed to triangulate feelings and evidence in order to determine if trainees were progressing through the “stages of training” appropriately:

*“It's just asking the right questions really and I think if they're not showing any progression with time that tells me that they're probably not doing very well” (10.44).*

*“They failed to make any progress really; their feedback on every placement was sub-standard repeatedly”. (1.27)*

*“Sometimes it is a lack of confidence and you probe and you push a little bit you get there and that's good and other times you just very quickly come to the idea "you really don't know what you're looking at do you?"”(7.87)*

Time was needed to assess the “consistency” of “accurate diagnoses”:

*“It's the ability to arrive at the correct diagnosis in a reasoned and reproducible way”. (3.5)*

*“The ability to derive a correct diagnosis…in a consistent fashion. (2.3)*

*“Report those accurately 100% of the time”. (8.14)*

*“Yes, you want to be sure that you know they will look at it and get it right and get it right and get it right again”. (7.60)*

*“I want to see them sort of repeatedly getting to the point where you can, they can have the slide and come and say “yes, this is what I think this is” and have their diagnosis and do that you know, repeatedly over a period of time where they... even if they haven't seen something for several weeks and next time we do get the same lesion they can still recognize that diagnosis”. (12.21)*

“Consistency”, in turn, related to the fact that a pathologist can have little “error”:

*“It's not the sort of job where you say “Well erm, yeah, I’m not having a very good day today so I'm just going to blitz through these cases”, it's, you just can't do that, so you have to maintain a certain standard, absolutely. (9.30)*

The analysis of the transcripts has shown that diagnostic competence is a longitudinal judgement, encompassing both “process” and “person” (Figure 4). Expectations of competence are, to some extent, stage-dependent and guided by RCPath guidelines. However, the judgement is complex, incorporating both evidence and feelings about trainees. Diagnostic competence appears to manifest in consultants “trusting” their trainees to be independent practitioners.

## 4.5 Discussion

Although competence in medicine has been difficult to define, it is important to explore and understand as it is of value to trainees, educators, regulatory bodies and the public (Burg, Lloyd and Templeton 1982). This ensures competency is transparent to all parties involved and should be made explicit in each medical or dental speciality where differences in competency may exist. Currently, the literature does not make it clear what diagnostic competence incorporates in histopathology or how consultants determine diagnostic competence in trainees. Despite this, WBAs are frequently used to assess competency in histopathology. Consultants must therefore spend a lot of time with their trainees, especially as up to 18 WBAs are required per year. Consultants also oversee all trainee work regardless of assessment targets. The interaction between a consultant and trainee can hence potentially offer significant insight into the judgement of diagnostic competence. This was particularly important given that the results from the questionnaire (see Chapter 3, section 3.4) suggest the current WBAs are not able to identify it. The purpose of this study, therefore, was to understand and explore the cues which underpin how consultant’s judge diagnostic competency in trainees.

Our data suggest there appear to be many “cues” which underpin how consultants judge diagnostic competency in their trainees. The judgement ecology indicates these cues are inter-related with each one having their part to play. This is mirrored in how interviewees think diagnostic competence should be assessed *“the worst thing for diagnostic competence is having one piece of paper or one assessment that is your proof of diagnostic competence. I think it has to be approached from a number of angles”.*

This indicates that exam results alone are not sufficient to assess diagnostic competence and this is in line with the curriculum that also lists other Stage D competencies which are expected before a trainee can be signed out of training (see Chapter 1, section 1.2.9). Although participants were not definitive as to which “cues” or qualities were most important to diagnostic competence, when asked to define diagnostic competence, it appeared that the ability to form a diagnosis was one of the most important factors. This theme had the richest data with many related aspects described by participants including the pathological approach, accuracy, error, consistency and knowledge. The degree of accuracy needed in the diagnosis was moderated by the stage of training, such that expectations of accuracy and consistency of accuracy grew with training experience. This mirrors the competency-based framework of the pathology curriculum. However, throughout training, consultants wanted to see trainees having the right “approach” to diagnosis, as the approach supported sound diagnoses and could be developed and augmented with knowledge, skills and behaviours as trainees progressed. This “approach” was not only of academic interest to consultants, but it ensured patients were protected by trainees demonstrating they had interpreted relevant clinical information, blended it with their histological findings and then conveyed the diagnosis to the surgeon in an appropriately worded report. Over time, consultants entrusted tasks to trainees and decreased the amount of supervision, such that trainees were able to act more independently as they neared consultant status. Diagnostic competence was not conceptualised as a static achievement but something that had to be maintained throughout one’s career.

It is important to note at this point, that within the theoretical framework, different consultants did appear to favour certain “cues” when forming their own personal judgement of diagnostic competency, even if they didn’t always explicitly state this. For instance, some consultants were less inclined to follow curricula and based their judgements on personal experience whereas other favoured evidence when making competency decisions. To explain, some consultants based their decisions based on exam results, WBA outcomes and similar objective measures. This illustrates that although this framework captures how consultants determine diagnostic competency in their trainees, in line with the literature review, judgements are often subjective and biases can exist (Govaerts, van der Vleuten and Schuwirth 2002). That is why a range of assessors should be utilised when determining competency and clearer guidance on what aspects are most appropriate to use is needed.

To summarise, the analysis revealed that diagnostic competence can be conceptualised within five main themes; 1) the person, 2) the process, 3) stage of training, 4) trust and 5) time. From its empirical base, this study forms a conceptual model (Figure 4.4) to describe how consultants determine diagnostic competence in their trainees and the cues which underpin their judgements. The discussion which follows examines the judgement ecology, discusses the themes identified in the data and compares our findings with the literature and questionnaire where appropriate.

### 4.5.1 The person

**Professionalism**

There has been an increasing emphasis placed on professionalism in medical and dental training, which has probably stemmed from high-profile incidences of mal-practice presented in the media, changes in patient care philosophy, the availability of online medical advice and changes to regulatory bodies due to pressure from government policy (Irvine 1997; The Royal College of Physicians 2005; van Mook et al. 2009). This has filtered down into medical and dental training programmes, such that professionalism is now considered an integral part of training curricula (Field, Cowpe and Walmsley 2017). It is therefore not surprising that participants commented on professionalism either directly, or by stating behaviours relating to professionalism, when describing diagnostic competence. These different stances reflect the fact that professionalism can be explicitly or implicitly described in curricula as shown in the following discussion.

Medical professionalism has been defined as a set of values, behaviours and relationships (The Royal College of Physicians 2005). In particular this includes integrity, compassion, altruism, continuous improvement, excellence and working partnership with members of the wider healthcare team (The Royal College of Physicians 2005). Here, the definition does not just describe behaviours such as “compassion” and “altruism” but also includes aspects of reflection such as “continuous improvement” and vaguer terms such as “excellence” which presumably relates to aspects of knowledge, skills and behaviour which deliver excellent patient care. In contrast, the Accreditation Council for Graduate Medical Education in America (ACGMEA) stated six core competencies for resident education (medical knowledge, patient care, professionalism, interpersonal and communication skills, practice-based learning, and systems-based practice) (Merlin et al. 2014). Here, professionalism is considered a separate competency, distinct from the other competencies such as “communication” and “medical knowledge”. Confusion is compounded when the term “professional competence” is used, as this phrasing tends to incorporate aspect of competency and professionalism, such as:

Professional competence is the habitual and judicious use of communication, knowledge, technical skills, clinical reasoning, emotions, values, and reflection in daily practice for the benefit of the individual and community being served (Epstein and Hundert 2002, p. 226).

Given the complex nature of professionalism it is not surprising that it was sometimes difficult to understand when participant’s comments related to professionalism or another theme such as reflection. Indeed, professionalism has been described as a second-order competence and is manifested in all that we do and say (Zijlstra-Shaw, Roberts and Robinson 2013). Nonetheless, aspects of professionalism were certainly part of diagnostic competence, similar to other studies that have examined competence in medical education (Epstein and Hundert 2002; Frank *et al.* 2010b; Matveevskii, Moore and Samuels 2012). From an interpretative point of view, analysis of the data showed that participants appeared to segregate professionalism from other aspects of competency. For instance, the theme such as ‘reflection’ embodied enough complex ideas and richness to stand-alone as a separate theme. This is similarly true for the themes of ‘the process’, ‘the patient’ and ‘trust’, which contained links to professionalism but had their own ‘keyness’ to warrant a framework displaying these as separate themes.

Those cues that did cluster under professionalism included timeliness and organisation that may be symptoms of other professional attributes described in the literature such as accountability, working with others and excellence (Cruess and Cruess 1997; Arnold 2002; Brown, Manogue and Rohlin 2002; The Royal College of Physicians 2005; van Mook et al. 2009; Cowpe et al. 2010; General Dental Council 2013; Ferguson 2014). For instance, it was considered important to work to deadlines as the pathologist is accountable to the surgeon, patient and hospital trust to deliver accurate and timely histology reports. Disorganisation was linked to reports being delayed, which could affect professional relationships (teamwork) and directly affect patient care (excellence). Teamwork has often been described as being part of professionalism and this was demonstrated in our data (Cowpe et al. 2010; General Dental Council 2013; Ferguson 2014). For example, certain personality types were linked to not being able to work part of a team such as *“Because they will fall out with all their consultant colleagues because they have no mechanism for resolving disagreements or conflict”.* Responding appropriately to feedback was also part of being professional and mostly related to taking the right course of action (reflection) and showing respect to those giving feedback, both of which appear in professionalism literature (van de Camp et al. 2004; The Royal College of Physicians 2005; Verkerk, de Bree and Mourits 2007; Trathen and Gallagher 2009; Cowpe et al. 2010; General Dental Council 2013). This is supported in our data with comments such as *“I think being able to take feedback on board and reflect on it and to act on that feedback is a good important part of professionalism”.* Trainees who showed high levels of motivation and a willingness to learn were also valued by participants, and are positive traits which have been correlated to professionalism (The Royal College of Physicians 2005; Cowpe et al. 2010; General Dental Council 2013; Ferguson 2014).Therefore, the aspects of professionalism that make up the theme of professionalism were not particularly specific or novel to histopathology. However, it shows that professionalism is an integral part of histopathology, mirroring the wider dental and medical literature. Further, this study is the first to demonstrate that professionalism is an integral part of diagnostic competence. This is important to appreciate as one might consider that diagnostic competence relates to the “ability to make a diagnosis” and professionalism is a separate consideration. As described earlier, competency and professionalism are inexorably linked both in the literature and in the data. Assessments which measure diagnostic competence must therefore reflect this stance, incorporating professionalism descriptors. This is highlighted by one participant who explained *“it's not just the diagnosis it's the whole…right from the specimen arriving to the report going out. So it's the whole thing. It's a different aspect, it's not the same as diagnosis but it is... part of their all-round competence I'd say”.* This is replicated in the conceptual framework as it is both the “person” and “process” that are needed to make a diagnosis.

**Reflection**

Reflection can be described as a metacognitive process that creates a greater understanding of both the self and the situation so that future actions can be informed by this understanding (Sandars 2009). It has been extensively linked to being an integral part of professionalism as previously discussed (Schon 1983; Atkins and Murphy 1993; Eraut 1994). Reflection, either as part of professionalism, or as a stand-alone quality, is also an important part of clinical competence (Epstein and Hundert 2002; Sandars 2009; Kaufman and Mann 2010; Grieveson et al. 2011; Mitchell et al. 2011; Chelliah and Arumugam 2012). Reflection in our data was described as serving many purposes but essentially it facilitated learning, improvement and encouraged trainees to adopt a mentality which would ensure they worked within their competency and asked for help when they needed it. This is illustrated in the data by the following comment “*I* *learned more from cases that I remember because I got it very wrong or I completely missed the point of what was going on*”. In particular, the ability to reflect helped trainees to frame diagnoses which were not straight-forward, for example “*having the insight to be able to recognize when I can't make a diagnosis or when there's a grey area and I have to be able to express that ambiguity*”. This was considered important when trainees became consultants as it meant they updated their knowledge, referred appropriately and remained safe practitioners. Reflection often involves three stages: awareness or interpretation of the situation, analysis or selecting options and thirdly a perspective change or moral judgement (Rest 1984; Atkins and Murphy 1993). This pattern of reflection was shown in our data with comments such as “*they will also hopefully become aware of what it was that actually led them coming to an erroneous conclusion*” which is awareness, “*Having an element of insight, a degree of insight, is a very important quality in a pathologist”* which is analysis and then the judgment that is made based on this, *“It's the ability to arrive at the correct diagnosis in a reasoned and reproducible way, consistent with their own limitations”.* Reflection ensured competency developed and was moulded by experience, as competency was not viewed as a static ability, but something one had to maintain throughout their career. Our data is therefore consistent with the literature on reflection in medical education and the ability of a trainee and consultant to reflect on their performance was considered to be a central part of diagnostic competency.

### 4.5.2 The process

**Forming a diagnosis**

Various studies have examined accuracy in histopathology, usually by means of comparing diagnoses to expert diagnoses (Cross, Dennis and Start 2002; Shanteau et al. 2002; Parham et al. 2010). Other studies have sought to understand how pathologists scan a slide to understand the visual-spatial process that leads to a diagnosis (Tiersma et al. 2003; Krupinski et al. 2006; Roa-Peña, Gómez and Romero 2010; Krupinski, Graham and Weinstein 2013; Brunyé et al. 2014). Accuracy has also been tied to diagnostic error and diagnostic errors comprise a substantial and costly fraction of medical errors (Graber, Gordon and Franklin 2002; Studdert et al. 2006).

Given that the primary role of a pathologist is to provide a diagnosis to a surgeon so that the patient receives the appropriate management, it is probably not surprising that forming a diagnosis was considered to be very important in this study. If a trainee is incapable of making accurate diagnoses, then there is a risk to patients (Goldstein 1999). Our data suggests that accuracy sometimes relates to getting a specific diagnosis while at other times, inaccuracy wasn’t as problematic so long as it didn’t affect patient care. This echoes negligence law whereby the “Cause in Fact” states that the plaintiff must show that their injury resulted from that erroneous diagnosis (Owen 2007); therefore, if the erroneous diagnosis was still a benign neoplasm that required the same treatment, there would be no case. The expectation of a consultant increases when trainees progress through training, such that inaccuracies were less acceptable towards the end of training.

Often the diagnosis was considered less important than the approach that had led to that diagnosis. Even in the first year of training, consultants wanted to see that trainees had the right “approach” to diagnosis even if the diagnosis was incorrect. For instance, *“Now, if they get the diagnosis wrong, which may happen, but they base their diagnosis on the evidence seen, then we can correct it.  But if they get the diagnosis wrong because they didn't take into account anything, then it is very obvious that they haven't studied correctly that stage…in which case they are incompetent”.* Consultants sometimes used objectives measures to try and tease out the approach used by the trainee (such as seeing if a trainee had spotted a particular histological feature or interpreted a special stain correctly), yet other times consultants found the approach most difficult to describe, mainly because the art of making a diagnosis occurs inside someone’s head and is not as easily demonstrable as those specialties which are more practical by nature. Sometimes the data correlated these intangible skills as being innate ability-that some individuals are just better at visual-spatial awareness and pattern recognition. This is illustrated by the comment *“When someone's got the eye for it or they've got the right approach it's very difficult to define or record but you know it when you see it”.* The problem with innate ability is that it is not something you can necessarily teach, so although it helps one to become diagnostically competent, it is probably not something you can assess during training. A more pragmatic approach would be to assess basic levels of aptitude early on in training, so that those individuals who lack any ability to form a diagnosis are directed towards other career paths. Currently, this is one of the functions of the OSPE undertaken in the first year of general histopathology as it acts as a gateway to continue with the rest of training.

The approach was also explored by consultant asking trainees to explain why the diagnosis was what it was, why they reached that diagnosis, how other entities had been excluded and if they had correlated clinical findings with their microscopy. However, others aspects related to consultants just “feeling” that their trainee was competent. These feelings may reflect that some aspects of competency are not easily compartmentalised into specific traits or behaviours, but rather a more global approach to competency is favoured, where the overall judgement is greater than the sums of the parts. It was noted that although consultants may identify errors in trainee reports, the nature of how the error occurred is not always made explicit to the trainee: “*they do pick up things that I've missed and they give me pointers on how to better write reports, but there's no explicit feedback…it's more you have made this error*”. This could be happening for a number of reasons, including a lack of time or prerogative on the consultant’s behalf. However, the results from the questionnaire (see Chapter 3, section 3.4) suggested there is enough time to complete WBAs and consultants are often very helpful in providing feedback. This comment might actually reflect that the “error” has stemmed from an incorrect “approach”, and being able to explain this approach to a trainee is difficult. Again, this feeds back to the invisible thought processes which occur when making a diagnosis which one can’t demonstrate through modelling. As illustrated by the following comment, this might explain why trainees often learn by *“shifting enough glass”,* as the very nature of seeing enough cases gives rise to “trial and error”. The approach to diagnosis is then gradually demystified through experience by the trainee where they are able to *say “I think it fits into this basket”* or *“What are the 10 things I need to be thinking about?” rather than there are two hundred things”.* Here, the experience of “shifting glass” exposes the trainee to the range and depth of diagnoses so they can eventually conceptualise entities into categories and begin to form a differential diagnosis.

The judgement ecology also showed that the regular episodes of one-on-one teaching between trainee and consultant are highly beneficial for the trainee to become diagnostically competent as the consultant can help to guide and refine the differential diagnosis. Hence, the current WBAs only detect minute episodes along this continuum, and although they might offer some insight into diagnostic competence, they may not reflect the full picture because competency is longitudinal in nature. As identified in the questionnaire responses to questions 4, 5 and 6 (see Chapter 3, section 3.4), most assessors and trainees felt that the current WBAs did not identify diagnostic competence, struggling trainees or reflect performance and although part of this might be because the forms do not reflect the knowledge, skills and behaviours that are most relevant, it also reflects that the judgement is longitudinal.

Knowing trainees could follow the correct diagnostic approach was also important to ensure guesswork wasn’t involved as some diagnoses are common by their very nature e.g. a lump on the buccal mucosa is nearly always a fibro-epithelial polyp. It was also important because this approach underpinned a safe trainee, so that errors and patient harm would be minimised. One study found that qualified doctors valued knowledge and clinical skills most highly when assessing postgraduate competence (Hojat et al. 2007). This mirrors our findings in that the ability to form sound diagnoses supported by appropriate levels of knowledge is a highly desirable component of competence in histopathology. However, our study is the first to use qualitative means to try and understand the nature of diagnosis and the factors that underpin it. This study also allows one to understand how accuracy of diagnosis sits within the wider framework of diagnostic competence, instead of examining it in isolation.

**The patient**

The need to protect patients is made explicit in histopathology training assessments and curricula, as well as the regulatory bodies which govern the profession and by working within the NHS (Teasdale 2002; General Dental Council 2013; General Medical Council 2015; The Royal College of Pathologists 2015; The Royal College of Pathologists 2018). Although it is mentioned in the histopathology curriculum, generally, our data reflect that protecting patients is considered a fundamental part of diagnostic competence. Consultants were very aware that their role involved ensuring patients received the correct treatment, something they wished their trainees to appreciate as well. This theme related closely to being accurate when making diagnoses, so patients came to no harm. However, it was also considered important to ensure reports conveyed the right message to surgeons, especially when the diagnoses were not clear-cut and needed phrasing in a particular way. Reports which showed the pathologist had excluded other clinical differential diagnoses were also useful and showed empathy with surgeons. Ultimately, the diagnosis, histology report and any form of communication which could affect patient care was considered significant. Indeed, protecting patients was mostly described in relation to providing diagnoses to surgeons, but it is important to mention that it was also manifested in the theme of reflection (learning from mistakes and self-improvement) and professionalism (as timeliness, organisation and proper communication). For example, *“the reality is there is an amount of work you have to turn around if you have to spend 20 minutes looking at every slide at high power and crawling over every cell then you're not going to be diagnostically competent”.* Here, the act of taking too long to examine slides would delay diagnosis and in turn, patient care. Taking too long could stem from lack of confidence and having the right amount of confidence when making diagnosis appeared important to competence. For example, *“there's this spectrum of confidence and I think that's… that's what kind of ties into this* (diagnostic competence). *This idea of ambiguity and then how confident you are around those areas”.* Ultimately, to be diagnostically competent there had to be a *“balance between quality, safety, and speed”*.

The need to protect patients conflicted with data that showed that it was important for trainees take ownership of cases and independently report, for instance *“You know, if you throw somebody in at the deep end, you learn to swim, you learn to do your own thing, you learn to take responsibility”.* This paradox has also been discussed in the literature (Kilminster et al. 2007); the idea that patients deserve the best care, but in order to train clinicians, experience must emerge in trainees who start without experience. These moments of anxiety where trainees solve problems independently can give a sense of autonomy and build confidence and motivation (Dunphy et al. 2010; ten Cate, Kusurkar and Williams 2011), supporting data such as *“In other words, training doesn't replicate working as a consultant because it is a highly supervised environment”.* That is why the theme of “trust” and “consistency of competence” (see sections 4.4.4 and 4.4.5) bridge the gap between being competent as a trainee and being competent to work as a consultant pathologist. This paradox is partly addressed by EPAs beginning to be utilised in postgraduate training, as they ask consultants to entrust tasks to trainees only when they are entirely happy that the “risk” has gone. However, practical tasks are easier to entrust whereas delegating certain “lesions” for a trainee to independently report is more difficult. To explain, if a trainee surgeon is entrusted to place a catheter in a patient, he can collect the necessary equipment and carry out the task provided he has been trained to do so. While that task is clearly defined, a trainee pathologist might be handed a biopsy of a possible benign polyp, when in fact, it turns out to be metastatic breast cancer. To this end, the suspected clinical diagnosis does not always match what is seen histologically. This is a dramatic example, but the nuances and complexities in diagnosis are potentially problematic. This explains why consultants wish to work closely with their trainees, assess all the evidence and only gradually release them when that trust is gained “*So it’s a case of slowly letting go rather like teaching a child to ride a bike. You run with them alongside, you've taught them how to push the pedals, you've taught them how to steer, you run with them alongside, eventually you let go of the handle bars and you either fall over or they cycle off into the sunset”.* This approach reflects the important role that pathologists play in patient care as it is the pathological diagnosis which ultimately determines the type of treatment a patient will receive. So despite the irony that pathologists don’t tend to meet patients, the need to protect them is substantially manifested in how consultants determine diagnostic competence in their trainees.

### 4.5.3 Stage of training

Competency-based education has been described by Frank and colleagues (2010b, p. 636) as:

An approach to preparing physicians for practice that is fundamentally orientated to graduate outcome abilities and organised around competencies derived from an analysis of societal and patient needs. It de-emphasises time-based training and promises greater accountability, flexibility and learner centeredness.

Analysis of the data showed that different levels of competency (incorporating different abilities) are also expected of histopathology trainees depending on what stage they are at in training. The data did not view first year trainees as being incompetent, but rather the expectation of what would constitute competency in the later years of training would increase in line with their extra experience. This step-wise measure of competency is very similar to Miller’s Pyramid (Miller 1990), where a first-year trainee would “know” in theory how to do something but a final year trainee would have the necessary, knowledge, skills and behaviours to actually “do” the task given to them.

Within the UK, the RCPath also divides histopathology training into stages (A-D) with suggestions on competencies required in order to progress to the next stage (The Royal College of Pathologists 2015). A competency-based framework is also utilised to guide trainers and trainees on the types of pathology that can be “signed off” at each level of training and then reported by the trainee independently (The Royal College of Pathologists 2009). For instance, “fibroepithelial polyps” can be signed off at level 1 but it is not until level 4 that a “granular cell tumour” can be signed off. These guidelines were reflected in the data, with comments such as *“To sign off a trainee, not only do they have to pass the exam, they have to pass their competences”* and *“But I think post FRCPath in order to err well we have a list of cases pre-CCST types cases that they can report according to Royal College guidelines”*. Our data therefore parallels the wider healthcare literature which show how competency-based education is influencing how we assess competence in histopathology (Harden and Crosby 1999; Smith and Dollase 1999; Long 2000; McKee 2008). The judgement ecology also demonstrated that trainee diagnostic competence was partly determined by examination results and WBA outcomes. For instance: “*there are the exams that keep the goal posts throughout the training and there are also the WBAs which may take the form of a summative assessment but they may be descriptive as well”.* However, this was often complemented by looking at the day-to-day work of trainees or using *“end of placement assessments”* such that WBAs weren’t particularly useful on their own. Given that the questionnaire found that WBAs don’t appear to identify diagnostic competence (with similar opinions also found in the interviews), and there being only 3 exams during training, it is probably the day-to-day observations which provide the most regular feedback to consultants about how their trainees are progressing. These regular interactions were valued by consultants and trainees as they also demonstrated consistency *“even if they haven't seen something for several weeks and next time we do get the same lesion they can still recognize that diagnosis”.*

The traditional alternative to competency-based education is the “time-spent” model where trainees spend a minimum period of time in certain sub-specialties, working on real clinical cases such that the trainee would become competent through experience (Hodges 2010). Critics of competency based-education would argue that its reductive nature can limit assessments to only those behaviours that can be measured, such as technical skills (Harden and Crosby 1999; Carraccio et al. 2002; Swing 2007; McKee 2008). This concept is partly borne out in our data where there appear to be some intangible competencies which are not directly measured yet are needed to develop trust between trainer and trainee. This is discussed in the next section.

### 4.5.4 Trust

An important aspect of diagnostic competency was trust and being able to trust a trainee was often described as a surrogate marker for competency. For example, consultants delegated tasks to trainees that they trusted and were more willing to let them independently report. However, it appeared that even when trainees passed their exams and were signed off for certain competencies, trust wasn’t automatically exemplified in the trainer-trainee relationship. Sometimes trainers wanted to spend more time with their trainee to make their own judgement of whether they thought they could be trusted to report independently, which suggests that some aspects of trust are intangible and do not relate to evidence alone. This mirrors other studies which have found that certain traits are tacit by nature, such as trustworthiness and self-awareness, requiring time for them to be imparted to another person as they are used without conscious awareness (Zijlstra-Shaw, Roberts and Robinson 2013). This is in contrast to overt personality characteristics, such as responsibility and accountability which are easier to explain in verbal terms and can be observed and interpreted through certain behaviours (Zijlstra-Shaw, Roberts and Robinson 2013).

Instead of looking at what trainees can and cannot do, or how they might behave, the alternative approach is to examine how the trainer interacts with their trainee and what they allow their trainees to do. Again, EPAs reflect this change in how competency is judged as they ask trainers when they are able to trust their trainees to undertake certain tasks (ten Cate 2013). This could be described as the “acid test” as the implications of entrusting tasks to trainees is that you trust that no harm will come to patients by doing so. However, these trust decisions are context- and task-specific (Mayer, Davis and Schoorman 1995) and it is therefore understandable why participants wanted to spend time with trainees; they needed to assess “risk” (Damodaran, Shulruf and Jones 2017). In other words, being signed off for different competencies did not always align to the complex, risky reality of medical practice in which the clinician must make decisions “in the moment” where “trust” is more important than “competency”. This sentiment of pitting trust against competencies is embodied in the literature (ten Cate 2005; Damodaran, Shulruf and Jones 2017) and is echoed in some of data such as *“You need to take everything into consideration. Yes. Everything…it's still difficult because that feeling you said* and *“Quite often people do have a feeling but they cannot pin it down to something”.* This doesn’t necessarily mean the trainee is not diagnostically competent at this point, but shows that when allowing someone to independently report, the consultant has to be 100% happy that there is no risk to patients. And to inform that judgement, consultants wish to spend time with trainees, checking for consistency of diagnosis and consistency of behaviour. To this end, trusting a trainee often related to repeated behaviours that illustrated they had both “process” and “person”.

### 4.5.5 Time

In order to make a judgement of a trainee’s diagnostic competency, analysis of the data showed that a period of time is needed. The data did not indicate a specific number of hours, days or weeks this should be, yet it was clear that competency had to be assessed over a period of time. This is probably partly due to the complex nature of diagnostic competency. For instance, the data revealed there are many cues to competency, so it follows that they cannot all be observed (or are very unlikely to be observed) over just a few cases. Competency was also linked to being consistent, and consistency, by definition, cannot be measured on a single case. The judgement ecology, being longitudinal in nature, was also reflected in comments relating to WBAs. For example, it can be seen from the data that WBAs were mere “snap-shots” of performance or even specially selected by trainees, so therefore were not a true reflection of their day-to-day work. This ties in with comments from the questionnaire (see Chapter 3, section 3.4) which said WBAs were “specially selected” or “may not be representative of true performance” and similar comments from within the data in the current study which related consistency to personal development *“The chances of coming across the same case, having that same assessment and consultant… so much confounding that I don't think you really get that progression documented in them”.*

Consistency is very important to competence as trainees must show that they are not only competent once, but are competent over a range of tasks on different days of the week. That is because performance with one patient (or diagnosis) does not necessarily indicate performance on other cases (Cohen, Halvorson and Gosselink 1994). Consistency needs to be maintained as inconsistency could put a patient at risk from harm. This is of paramount importance in histopathology as diagnoses determine treatment outcomes for patients and carry a significant burden of responsibility.

The importance of consistency of is also inferred in the curriculum as there are required quotas of WBAs for trainees to complete per year. In other words, these minimum numbers are required to increase the reliability of the outcomes. This is also why WBAs should be undertaken with a range of assessors (Sloan et al. 1993).

Given the vast amounts of knowledge, skills and attitudes that a trainee must acquire and develop during training, it is understandable that competence is acquired over time (Cash 1995). The very nature of competency-based education requires time for trainees to progress through the various stages of training, especially as certain tasks or competencies will not present on a daily basis. Time must pass for the trainee to gain experience and be exposed to all that the curriculum requires.

This is equally true when someone is judging whether a trainee is capable of reflection, which is a cyclical process with time at its core. Competency judgements also require time as we have discussed how competency is often judged in stages, requiring trainees to be signed off for the listed competencies for that year of training. Clearly, and in line with how competency develops, time is needed to develop the necessary knowledge, skills and behaviours of the profession (Rasmussen 1983; Dreyfus and Dreyfus 1986; Miller 1990; Rethans et al. 2002). The analysis also shows that time is needed to ensure one maintains competency after they become a consultant as competency evolves. This explains how reflection augments competency over time. Competence is not the final point, as competency must be maintained and the nature of the “real world” outside of training curricula is that it is infinitely complex, such that trainees must continue to learn, reflect and sustain professional growth (Benner, Tanner and Chesla 1996).

## 4.6 Limitations and considerations

It is worth stating that the data relates to UK pathologists and UK training. Diagnostic competence may differ in other countries, where there are differing expectations, contexts and training pathways to consider. However, this study utilised a range of consultants and trainees, which should add to its generalisability.

The literature review showed that little is known about diagnostic competence in histopathology, such that the majority of themes identified stemmed directly from the interview data. To this end, the emerging framework changed several times as interviews continued, indicating a response to the data. The emerging themes were also triangulated with the rest of the research team where further discussion and reflection occurred. Many aspects of diagnostic competence shared similar characteristics to the literature on competency in general, adding further validity to the study.

It is important to reiterate that participants were asked specifically about diagnostic competency. This phrasing was used to avoid discussions about competency in general as the study aimed to examine those competencies which were integral to making diagnoses. There may, therefore, be other competencies, unrelated to diagnosis-making, which should be examined during training and could be explored with further interviews.

## 4.7 Conclusion

This is the first study to explore in depth the behaviours and actions which demonstrate these qualities as well as the relationships and context which underpin them. This data and the resultant conceptual model can begin to form a basis to design an assessment tool which tests diagnostic competence in histopathology trainees. However, as previously discussed, the relative importance of the different cues was not clearly described by participants. Often, the different “cues” supporting the judgement of diagnostic competence appeared equally important. Although this may be the case, the volume and richness of data that stemmed from “forming a diagnosis” suggests this element is either very complex or more important than some of the other “cues”. Accordingly, the last few stages of Cooksey’s framework relate to capturing judgements so one can see individual judgement policy. In other words, what people say is not always the same as what they do. Hence, the next chapter of this thesis is concerned with exploring how consultants view these cues and their relative importance in the judgement process.

# **Chapter 5**

# **A Delphi study to triangulate the findings from the qualitative interviews for assessing diagnostic competence**

**Chapter 5: A Delphi study to triangulate the findings from the qualitative interviews in assessing diagnostic competence**

## 5.1 Background

One of the main objectives of this chapter is to determine the importance of the cues identified in the qualitative interviews, in assessing diagnostic competence. This will then satisfy the last four stages in Cooksey’s theoretical framework (1996b), which are shown below:

* Sample cue profiles
* Sample participating judges
* Obtain judgements
* Capture individual judgement policy

Case vignettes are often utilized to tease out those cues that participants rate most highly when making decisions. These case vignettes, or scenarios, ask participants to make decisions on information provided. The scenarios must be realistic to work best, reflecting the sort of scenarios that would occur in the real world. By altering the scenarios and the information (or cues) given to the participants, regression analysis is used to determine the importance of the different cues. However, our data indicates that the judgement process and ecology relevant to diagnostic competence is extremely complex and multi-faceted. The interviews also revealed that the judgement of diagnostic competence is longitudinal in nature and as such it is best assessed over time. Hence, the classical application of Cooksey’s framework would not be appropriate for at least two reasons. Firstly, creating scenarios and designing cues that reflected the real judgement ecology would over-simplify the judgement task. Secondly, presenting consultants with scenarios to judge would effectively be snapshots in time and not reflect the true judgement ecology.

However, there are several other ways to determine how individuals rate the importance of different cues, which broadly fit into the category of consensus methods. In these a group of experts are utilized to measure and develop consensus on a particular topic or set of items. The three main methods are Delphi, the nominal group technique (NGT) and the RAND corporation/University of California Los Angeles appropriateness method (RAND/UCLA) (Humphrey-Murto et al. 2017).

The Delphi method was originally used in the 1950s by the RAND Air Force Corporation in America (Murphy et al. 1998). Within the field of medical education, it is the most commonly selected consensus group method, accounting for approximately 75% of papers (Humphrey-Murto et al. 2017). A Delphi study involves using a panel of experts in a given field to come to an agreement on a particular topic or set of items. The participants are presented with information on the topic and asked to determine, individually, how they rate each item (using a Likert-type scale or similar scoring method). Participants are also encouraged to provide feedback on items so they can be refined in future rounds of the Delphi if needed. The researcher collates the responses to the items and is able to see which items participants rated highly (or agreed on) and which ones did not reach agreement and need refinement. In subsequent rounds, the participants are provided with data from this previous round where they are able to continue to refine and perhaps reach agreement on the remaining items. Delphi has several integral features: anonymity, iteration, controlled feedback, statistical group response and structured interaction (Jones and Hunter 1995; Murphy et al. 1998b).

NGT is a structured interaction between individuals conducted face to face. Essentially, these individuals have to express their preference (e.g. 1st, 2nd 3rd) on a particular course of action or policy. Often the least popular preferences are removed for the next round where further discussion can take place regarding the relevant merits of the remaining policies (Humphrey-Murto et al. 2017). RAND/UCLA is essentially a hybrid of Delphi and NGT. The first stage is conducted like a Delphi with anonymous scoring of items. The second stage involves participants meeting face-to-face where the collated results are discussed. NGT and RAND/UCLA were not chosen as it was felt that 1) dominant personalities could unduly influence debate if it occurred face-to-face 2) the target participants for this study (see later) were all extremely busy individuals (and geographically dispersed) who would find it easier to complete the Delphi in their own time, rather than on specified dates and locations.

A Delphi was therefore chosen to capture the judgements of participants, thus fulfilling the last four stages of Cooksey’s framework. Delphi can help to triangulate the cues and judgement ecology identified in the qualitative interviews. However, it will also determine the importance of the different cues, similar to the way case vignettes and regression analysis are often used. The difference is that the cues will be apparent to participants in the Delphi study and not manipulated within scenarios. However, the Delphi will still help to explore the judgmental process because participants have anonymity and are encouraged to provide feedback on the cues in each round. For example, if they are asked to rate the importance of “being accurate” when making diagnoses, there may be elements that participants wish to discuss or clarify with the research team or other Delphi participants. These comments therefore offer insight into how these cues might be interpreted just as case vignettes can fulfil this purpose in Cooksey’s framework.

## 5.2 Aim and Objectives

**Aim**

To use a Delphi study to triangulate the findings from the qualitative interviews in assessing diagnostic competence.

**Objectives**

* To determine if the cues identified from the qualitative interviews are valued by experts.
* To determine the relative importance of individual cues
* To identify any additional cues

## 5.3 Method

### 5.3.1 Participants

Participants were purposively selected in order to create an “expert” panel. In line with recommendations regarding how to select “experts” (Humphrey-Murto et al. 2017), it was felt that TPDs were most knowledgeable in this area, representative of the area of inquiry and had considerable practical experience to warrant them “experts”. In particular, they:

* are responsible for the overall training of trainees
* are responsible for interpreting information at the ARCP and ultimately deciding if trainees are progressing appropriately in training or require remedial action
* are usually senior individuals who have previously held educational supervisor roles
* often complete WBAs for trainees

In addition, further stakeholders were chosen who might offer valuable perspectives given their role in training. These were consultants who had roles within the deanery (other than TPD or educational supervisor) and those who worked with the RCPath in developing curricula or training. To protect participant anonymity, further details will not be provided.

There is no stated recommended number of participants for Delphi studies (Humphrey-Murto et al. 2017). Less than 6 is not considered ideal (as there are too few participants for useful discussion) but 12 or more is considered reasonable (Murphy et al. 1998b). It is important to stress than the number of participants is less important than the makeup of the individuals selected; they should offer expertise and a range of viewpoints to be most useful (Bloor et al. 2015). Twelve participants were chosen for this study in line with the aforementioned guidance. It was considered reasonable to have a relatively small panel for two principle reasons. Firstly, diagnostic competence is a niche topic which is really only fully understood by those individuals who judge diagnostic competence in trainees. Therefore, it would not be appropriate to use other individuals who are often recommended participants of Delphi studies in healthcare research such as patients, providers and payers (Humphrey-Murto et al. 2017), as they would not offer much insight into this complex judgement task. Rather, having a select group of individuals who understand the task fully was more appropriate. Secondly, in some Delphi studies, the first step is generating a list of items to rate in the next round. As the items in this study had already been generated through qualitative interviews, less pooling of ideas and therefore fewer participants were needed.

### 5.3.2 Determining items, rating, ranking consensus, stability and number of rounds

#### 5.3.2.1 Items

These items were generated from the qualitative interview data described in the previous chapter. The themes and sub-themes were re-examined to develop descriptions that would form items that could be rated by participants. Through an iterative manner, the exact wording was changed several times before a list of 27 items was created (see Table 5.1). Some items were self-explicit such as the “The diagnosis is correct” whereas others incorporated more complex ideas. For instance, for the item “Trainee shows evidence of clinico-pathological correlation”, specific examples of what this might include were provided to participants to aid understanding. This was important so that the item didn’t lose its sense of meaning and was made clear to participants. As such, creation of the 27 items was a balance between making sure they represented the themes present in the qualitative interview data and ensuring they were concise enough so the number of items was not too long or over-complicated, as it has been shown this can limit responses (de Villiers, de Villiers and Kent 2005). Items were also grouped under headings to place similar items together (see Table 5.1).

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Definitely not important Definitely important** | | | | | | | | | |
| **Stage of training** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
| 1. The stage of training e.g. ST1 versus ST5 |  |  |  |  |  |  |  |  |  |  |
| 2. RCPath curricula and training guidelines |  |  |  |  |  |  |  |  |  |  |
| **Timing** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
| 3. Placements/cases assessed longitudinally over a period of time. |  |  |  |  |  |  |  |  |  |  |
| 4. Individual cases in an examination setting |  |  |  |  |  |  |  |  |  |  |
| **Professionalism** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
| 5. Trainee is organised and timely when conducting themselves in the department e.g. does not lose cases, is aware of turnaround times, triages urgent cases |  |  |  |  |  |  |  |  |  |  |
| 6. Trainee communicates with all staff appropriately e.g. effectively, timely, politely |  |  |  |  |  |  |  |  |  |  |
| 7. Trainee works as part of a team e.g. works effectively with colleagues and does not create conflict |  |  |  |  |  |  |  |  |  |  |
| 8. Trainee is motivated and has a good attitude |  |  |  |  |  |  |  |  |  |  |
| **Forming a diagnosis** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
| 9. The diagnosis is correct |  |  |  |  |  |  |  |  |  |  |
| 10. Trainee consistently produces a correct diagnosis |  |  |  |  |  |  |  |  |  |  |
| 11. The histopathology report is accurate & does not contain factual errors or omit important information |  |  |  |  |  |  |  |  |  |  |
| 12. Trainee shows evidence of clinico-pathological correlation: clinical information/macroscopic work is used to support/refute possible diagnoses/features. Examples might be correlating radiographic with histological findings; referring to previous pathology reports that are relevant to the current biopsy; recognising that clinical information provided may affect what is seen histologically, e.g. previous radiotherapy |  |  |  |  |  |  |  |  |  |  |
| 13. Trainee follows sound pathological principles to reach a diagnosis e.g. the trainee understands how they reached the diagnosis, has a consistent, accurate way of examining slides and knows the steps involved when making a diagnosis. The diagnosis has not stemmed from guessing. |  |  |  |  |  |  |  |  |  |  |
| 14. Trainee has appropriate level of basic knowledge. e.g. recognises normal histology for different organs/tissues, knows basic pathological processes, can compare normal to abnormal |  |  |  |  |  |  |  |  |  |  |
| 15. Trainee has appropriate level of up-to-date knowledge. e.g. latest papers/datasets |  |  |  |  |  |  |  |  |  |  |
|  | **Definitely not important Definitely important** | | | | | | | | | |
| **Reflection** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
| 16. Trainee is aware of their own limitations |  |  |  |  |  |  |  |  |  |  |
| 17. Trainee reflects on their own limitations/performance |  |  |  |  |  |  |  |  |  |  |
| 18. Trainee shows evidence of improvement following reflection |  |  |  |  |  |  |  |  |  |  |
| **Trust** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
| 19. Trainee can be trusted to carry out macroscopic examination and “cut-up” independently |  |  |  |  |  |  |  |  |  |  |
| 20. Trainee can be trusted to report cases (however, the consultant will check the reports and authorise them) |  |  |  |  |  |  |  |  |  |  |
| 21. Trainee can be trusted to report independently (the consultant does not check the report and the trainee authorises it) |  |  |  |  |  |  |  |  |  |  |
| 22. Trainee asks for help/second opinion when needed |  |  |  |  |  |  |  |  |  |  |
| 23. Your opinion of a trainee’s diagnostic competence is related to an overall impression you have of them, rather than items of evidence e.g. gut-feelings/impressions/ your perceptions about trainees |  |  |  |  |  |  |  |  |  |  |
| 24. Your opinion of a trainee’s diagnostic competence is related to bringing all the evidence together and triangulating findings from exams, workplace-based assessments, day-to-day work |  |  |  |  |  |  |  |  |  |  |
| **The patient’s report** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
| 25. Report is useful to the clinician/surgeon e.g. it does not contain unnecessary information or detail; it shows the surgeon’s clinical differential diagnoses have been considered |  |  |  |  |  |  |  |  |  |  |
| 26. Report ensures the clinician/surgeon receives the appropriate message e.g. when diagnoses are not clear-cut and a differential diagnosis or other message needs to be conveyed. This might include a discussion about the limitations of the biopsy or if further tests or information are needed |  |  |  |  |  |  |  |  |  |  |
| 27. Report is organised appropriately and well-written e.g. it is logically laid out, is not confusing and uses appropriate language |  |  |  |  |  |  |  |  |  |  |

**Table 5.1. List of items to be rated in the Delphi**

**5.3.2.2 Rating and ranking**

Participants were asked to rate the items in terms of importance on a 10 point Likert-type item scale where 10 represented “definitely important” and 1 represented “definitely not important” (see Table 5.1). A 10-point scale was chosen to: 1) ensure there was no “middle option”. This is important because participants have to commit to a positive or negative stance (Garland 1991) rather than remaining neutral. 2) To help with the statistics (see section on stability 5.3.2.4).

Participants were also asked to rate both the top 3 and bottom 3 items which they felt were most important and least important when determining diagnostic competence in trainees. The ranking was limited to six items as it was felt that ranking all 27 items would be too onerous, especially as their importance would also be garnered from their consensus and stability scores (see sections 5.3.2.3 and 5.3.2.4 below).

For both rating and ranking judgements, “free-text boxes” were provided to enable participants to justify their scores or offer other items which were not listed in the Delphi. This information was used to help understand the rationale behind the consensus and offered further insight into the judgement ecology.

#### 5.3.2.3 Consensus

Consensus has proven difficult to define in Delphi studies with a range of possibilities to choose from. In many cases, it is considered an arbitrary figure (von der Gracht 2012). It is important to define consensus before the study begins, not to change it during the course of the study and to ensure it is clearly worded so it is easy to determine whether an item has reached ‘in’ or ‘out’ consensus. Determination of consensus should not just take into account those participants who agree an item is important, but also those who disagree the same item is important, as sometimes a lack of consensus on an item can offer just as much insight as those that reach consensus. Indeed, Delphi is not an exact science and recommendations have focused on trying to make the process of Delphi transparent to those who interpret the results (Humphrey-Murto et al. 2017). Consensus in this study was defined as “consensus in” when >70% of participants scored the item as 8-10 (definitely important) and <15% have scored it 1-3 (definitely not important). “Consensus out” was when >70% of participants scored an item 1-3 and <15% participants scored it 8-10. All other combinations were considered to be ‘no consensus’, in accordance with previous research that used similar 9 or 10 point scales (Harman et al. 2015; Tsichlaki et al. 2017). This consensus method ensures that when >15% of participants score an item 1-3, it can prevent an overall consensus even when the remaining participants score it 8-10. This is important so that outlier’s views aren’t disregarded (and consensus is falsely reached) and has the ability to generate further discussion in subsequent rounds. As a further condition for consensus, it was decided the interquartile range (IQR) should be 2 or less to achieve consensus ‘in’ or ‘out’ (Scheibe, Skutch and Schofer 1975; von der Gracht and Darkow 2010). This is a useful parameter to include because the IQR represents the middle 50% of scores or the difference between the 75th and 25th percentiles and is a further way of ensuring significant outliers’ view are not disregarded.

#### 5.3.2.4 Stability

It is recommended that consensus is not the only measure (Holey et al*.* 2007) to determine Delphi outcome. The stability of participant responses is also important because it shows that views are not changing significantly between rounds. For example, an item could reach consensus in round 1 and then be removed from further rounds of the Delphi in the belief that the opinion on the item was stable. However, if the item had remained for round 2, opinion might have fluctuated. It is therefore important to see that opinion remains stable over rounds to prove that further rounds are unlikely to change the outcome for that item. In this study, stability was determined by comparing the group response per item between rounds. Individual participant responses were not used as Delphi is interested in the group response (von der Gracht 2012). Another advantage of looking for stability is that the consensus measurement defined above uses pooled scores (e.g. 8-10 or 1-3) to determine consensus, therefore truncating opinion. However, the stability testing uses raw scores and detects if fluctuations in opinion occur. Thus, the study utilized consensus and stability when assessing the importance of items.

Likert-type data which is used in this study is ordinal data and is suited to nonparametric tests to determine significance. Furthermore, as the data is paired (first round versus second round), a Wilcoxon matched pairs signed rank test was suitable for determining if there was a significant difference in opinion per item per round. Other statistical tests can be used to measure stability, such as intra-class correlation coefficients (von der Gracht 2012), but a Wilcoxon matched pairs signed rank test was chosen as it was familiar to the researcher and secondly, there is no agreed statistical test for measuring stability in Delphi studies (von der Gracht 2012). Using statistics is a much more objective method to determine stability and number of rounds (see section 5.3.2.5).

Of note, when performing a Wilcoxon matched pairs signed rank test, differences between rounds of “zero” (i.e. when the participant does not change his or her score) are discarded. This is of particular concern if the samples are taken from a discrete distribution such as Likert-type data (Pratt 1959) because if a participant doesn’t change their score, lots of “zeros” can be created. A 10-point scale was utilized to help with this as it provides more choice to participants, potentially offering more scope for participants to change their responses between rounds, something which might prove more difficult on a narrower scale. This therefore helps to reduce the number of zero differences. Furthermore, Pratt proposed a modification to the Wilcoxon test to overcome the issue of zero differences (Pratt 1959), which was employed when analysing the results. In short, the modification randomly allocates zero differences to either positive or negative ranks to overcome the problem rather than just removing them from the analysis.

#### 5.3.2.5 Number of rounds

There is little evidence to suggest the minimum number of rounds required in a Delphi study before it can be terminated but the recommendation is between 2 and 3 (Boulkedid et al. 2011). Ensuring stability in opinion is one way of determining the number of rounds and was adopted in this study because it is a more objective measure (Diamond et al. 2014). If opinion is stable after two rounds, the Delphi stops. If opinion is significantly different (e.g. the results are unstable, see previous section 5.3.2.4), a third round is utilized. No more than three rounds are used as further rounds can lead to attrition (Tammela 2013; Diamond et al. 2014).In addition, there is an argument that conducting too many rounds can suggest to the Delphi panel that they “need to reach consensus” even if there isn’t a naturally occurring one. Indeed, those items which do not reach consensus can offer just as much insight than those that do, highlighting difference in perspective on the topic (Jünger et al. 2017).

### 5.3.3 Procedure

The Delphi was first piloted with an experienced TPD who was not part of the main study. Slight amendments were suggested to the wording of some items to improve how they would be applied to the scale.

The RCPath database of all registered general pathology TPDs was made available and an invitation email (approved by RCPath: Appendix 7) was sent together with a participant information sheet detailing the nature of the study (Appendix 8). A reminder email was also sent. Three TPDs responded to the emails expressing an interest to take part and the RCPath also suggested a senior member of the College who had voiced interest in the project.

Four TPDs in oral and maxillofacial pathology were also sent the email and information sheet and agreed to take part (as they were not present on the email sent by RCPath). A previous general pathology TPD (and therefore no longer on the current TPD mailing list) was also emailed based on their significant experience in training. Three further participants were purposively emailed as they had deanery roles or significant experience in training.

The final Delphi panel therefore consisted of 12 participants whose details are shown in Table 5.2 below.

|  |  |
| --- | --- |
| **Delphi panel background** | |
| Gender | 6 males; 6 females |
| Geography | Panelists were selected from all over the UK. |
| **Roles (past or current)** | |
| TPD | 8 |
| ES | 11 |
| Additional college or deanery role (other than TPD/ES) | 6 |

**Table 5.2. Summary of Delphi panel participants**

All communication between the researcher and Delphi participants occurred over private, individual emails so participants remained anonymous to one another. For round 1, participants were given instructions (Appendix 9) about how to complete the Delphi, a consent form (Appendix 10) and the items (Table 5. 1) to rate and rank (note the word “cue” was changed to “item” on the forms as it was felt the former might be confusing to participants). Participants were requested to return their completed Delphi form within 3 weeks. Reminder emails were sent to non-responders.

Round 1 responses were collated and the following calculated:

* Response rate
* List of items which had reached consensus/non-consensus
* Mean, standard deviation (SD), median, range and IQR for each item
* Top 3 ranking items and bottom 3 ranking items

In addition, free text comments were analyzed to see if items needed amending or additional items required adding to Delphi.

Round 2 of the Delphi was then sent to the participants. They were asked to rate and rank the items again, which included those items that reached consensus in the first round. This was necessary so that statistical stability could be measured between rounds. In addition, participants were given information from the previous round to include:

* Participant’s own scores
* Mean panel score
* Range of scores for that item
* List of items which had reached consensus/non-consensus
* Any amendments to original items from round 1
* Any additional items from round 1

Once all round 2 responses were returned, the following were calculated:

* Response rate
* List of items which had reached consensus/non-consensus
* Mean, SD, median, range and IQR for each item
* Top 3 ranking items and bottom 3 ranking items
* Wilcoxon matched pairs signed rank test to see if stability had been reached

### 5.3.4 Analysis

Data was inputted into an Excel™ spreadsheet. Data governance and ethical protocols were adhered to as per the University of Sheffield Ethics Committee (ref. 008042). This ensured both probity of the research and protection of the participants. Basic descriptive statistics (mean, median, SD, range, IQR,) and inferential statistics (Wilcoxon matched pairs signed rank test) were calculated, the latter using GraphPad Prism version 7.00 for Windows, GraphPad Software, La Jolla California USA, [www.graphpad.com](http://www.graphpad.com). Free text comments were analysed thematically.

Whilst it is acknowledged that Likert data is ordinal in nature and therefore, median and IQR are most appropriate, the mean and SD were also utilized in this study as they offer more insight into the degree of consensus between rounds. For instance, the mean and SD, when combined with the range/IQR and medians, can be used to show whether convergence has occurred, by a movement towards central tendency. The amount of convergence and therefore the strength of agreement is indicated by a comparison of SD (strength of aggregate judgement) while the range tells the researcher if outliers exist. This would not be possible using mean scores alone. This is in agreement with the findings of Greatorex and Dexter (2000) . For example, they have previously recommended using a range of measures to help interpret the data, as this can offer more insight. When this information is used alongside free text comments, a more complete picture is built about what the Delphi panel think of the different items presented to them.

## 5.4 Results

### 5.4.1 Round 1 rating of items

The response rate for round 1 of the Delphi was 100%. The results are summarised in Table 5.3.

A total of 17 out of the 27 items reached “consensus in” after round 1. Ten items reached “no consensus”. No items reached “consensus out”. These items are summarised in Table 5.4.

**Table 5.3. Summary of results from round 1 of Delphi.** Scores rounded to one decimal place. Under consensus column, green indicates “consensus in” and red “no consensus”. Consensus in this study was defined as “consensus in” when >70% of participants scored the item as 8-10 (definitely important) and <15% have scored it 1-3 (definitely not important). “Consensus out” was when >70% of participants scored an item 1-3 and <15% participants scored it 8-10. All other combinations were considered to be ‘no consensus’. In addition, interquartile range (IQR) had to be 2 or less to achieve consensus ‘in’ or ‘out’. Range= difference between highest and lowest score given to that item. Items ranked from 1-27 according to mean score. (1=highest rank, 27 = lowest rank).

|  |
| --- |
| **Stage of training** |
| 1. The Stage of training e.g. ST1 versus ST2 |
| 2. RCPath curricula and training guidelines |
| **Timing** |
| 3. Placements/cases assessed longitudinally over a period of time. |
| 4. Individual cases in an examination setting |
| **Professionalism** |
| 5. Trainee is organised and timely when conducting themselves in the department e.g. does not lose cases, is aware of turnaround times, triages urgent cases |
| 6. Trainee communicates with all staff appropriately e.g. effectively, timely, politely |
| 7. Trainee works as part of a team e.g. works effectively with colleagues and does not create conflict |
| 8. Trainee is motivated and has a good attitude |
| **Forming a Diagnosis** |
| 9. The diagnosis is correct |
| 10. Trainee consistently produces a correct diagnosis |
| 11. The histopathology report is accurate & does not contain factual errors or omit important information |
| 12. Trainee shows evidence of clinico-pathological correlation |
| 13. Trainee follows sound pathological principles to reach a diagnosis |
| 14. Trainee has appropriate level of basic knowledge. |
| 15. Trainee has appropriate level of up-to-date knowledge e.g. latest papers/datasets |
| **Reflection** |
| 16. Trainee is aware of their own limitations |
| 17. Trainee reflects on their own limitations/performance |
| 18. Trainee shows evidence of improvement following reflection |
| **Trust** |
| 19. Trainee can be trusted to carry out macroscopic examination and “cut-up” independently |
| 20. Trainee can be trusted to report cases (however, the consultant will check the reports and authorise them) |
| 21. Trainee can be trusted to report independently (the consultant does not check the report and the trainee authorises it) |
| 22. Trainee asks for help/second opinion when needed |
| 23. Your opinion of a trainee’s diagnostic competence is related to an overall impression you have of them |
| 24. Your opinion of a trainee’s diagnostic competence is related to bringing all the evidence together and triangulating findings from exams, workplace-based assessments, day-to-day work |
| **The patient’s report** |
| 25. Report is useful to the clinician/surgeon e.g. it does not contain unnecessary information or detail |
| 26. Report ensures the clinician/surgeon receives the appropriate message |
| 27. Report is organised appropriately and well-written |

**Table 5.4. Summary of which items reached “consensus in” (green) or “no consensus” (red) after round 1.** “Consensus in”>70% of participants scored the item as 8-10 (definitely important) and <15% scored it 1-3 (definitely not important). In addition, the IQR was ≤2. Some items have been shortened for reasons of clarity.

Table 5.4 shows that no items under ‘Stage of training’ reached consensus. In contrast, all items under both “reflection” and “the patients report” reached consensus and as did three out of four items under “professionalism”. However only one out of two items reached consensus under “timing”, four out of seven items under “forming a diagnosis” and three out of six under “trust”. The items which did not reach consensus are summarised in Table 5.5.

|  |  |  |
| --- | --- | --- |
| **Item** | **Condition 1 met?** | **Condition 2 met?** |
| **Stage of training** | | |
| 1. The stage of training e.g. ST1 versus ST5 | No | No |
| 2. RCPath curricula and training guidelines | No | No |
| **Timing** | | |
| 4. Individual cases in an examination setting | No | No |
| **Professionalism** | | |
| 5. Trainee is organised and timely when conducting themselves in the department | Yes | No |
| **Forming a diagnosis** | | |
| 9. The diagnosis is correct | No | No |
| 10. Trainee consistently produces a correct diagnosis | No | No |
| 15. Trainee has appropriate level of up-to-date knowledge e.g. latest papers/datasets | No | Yes |
| **Trust** | | |
| 19. Trainee can be trusted to carry out macroscopic examination and “cut-up” independently | No | Yes |
| 21. Trainee can be trusted to report independently (the consultant does not check the report and the trainee authorises it) | Yes | No |
| 23. Your opinion of a trainee’s diagnostic competence is related to an overall impression you have of them | No | No |

**Table 5.5. Summary of items which did not reach consensus after round 1.** Condition 1: “Consensus in” is when >70% of participants scored the item as 8-10 (definitely important) and <15% have scored it 1-3 (definitely not important). “Consensus out” was when >70% of participants scored an item 1-3 and <15% participants scored it 8-10. Condition 2: IQR must be ≤2 to reach consensus in or out. Note, some items have been shortened for reasons of clarity.

Table 5.5 shows that items 1, 2, 4, 9, 10 and 23 failed to meet both conditions 1 and 2. Items 5 and 21 failed to meet condition 1 only. Items 15 and 19 failed to meet condition 2 only.

Five participants provided feedback on the 27 items presented to them in round 1. Feedback included observations such as:

*“They are difficult propositions to rate because they are all very important. They are a bit like answering “When breathing, oxygen is important 1 …… 10”.*

*“In this I have focused on diagnostic competence per se and largely ignored other factors which make one a rounded, safe and balanced senior clinician (of which there are a few within these questions).* *Some of these are complexly interrelated and whilst on their own are not important, are functions of other traits”.*

However, more specific feedback which related to how items might be amended for the next round included:

*“The questions do not differentiate between a year 1 trainee and a year 5 one. Some of my answers might be different if stage of training was included”*

*“Many of your statements need clarification with the year of training – because this will make a huge difference”.*

As a result, the following items (under their respective headings) were amended for round 2 to incorporate the stage of training (bold text indicates amendment):

Forming a diagnosis

Item 9: The diagnosis **is commensurate with stage of training**

Item 10: Trainee consistently produces **a diagnosis commensurate with stage of training**

Trust

Item 19: Trainee can be trusted to carry out macroscopic examination and “cut-up” independently **commensurate with stage of training**

Item 21: **Senior trainees** can be trusted to report independently (the consultant does not check the report and the trainee authorises it).

A further comment from a participant was *“Exhibiting ownership of cases, i.e. acting to take responsibility and not as a consultant’s assistant pathologist, independent initiative to sort cases out”.* Therefore item 20 under ‘Trust’ was amended to reflect this stance:

Item 20: Trainee can be trusted to manage cases, **showing ownership and initiative** (however, the consultant will check the reports and authorise them).

**The patient’s report**

In addition, item 25, under ‘The patients report’ was amended as in retrospect it did not read clearly. In round 1 it read: “Report is useful to the clinician/surgeon e.g. it does not contain unnecessary information or detail; it shows the surgeon’s clinical differential diagnoses have been considered”. In round 2 it was changed to: **The trainee ensures the** report is useful to the clinician/surgeon, **only contains appropriate information and shows the** surgeon’s clinical differential diagnoses have been considered. Of note, however, item 25 reached consensus in round 1.

Delphi panel members did not suggest any additional items to the original 27 items provided in round 1.

### 5.4.2 Round 2 rating of items

The response rate for round 2 of the Delphi was 100%. The results are summarised in Table 5.6.

Twenty two items reached “consensus in”. Five items reached “no consensus”. Once again, no items reached “consensus out”. These outcomes are summarised in Table 5.7.



**Table 5.6. Summary of results from round 2 of Delphi.** Scores rounded to one decimal place. Under consensus column, green indicates “consensus in” and red “no consensus”. Consensus in this study was defined as “consensus in” when >70% of participants scored the item as 8-10 (definitely important) and <15% have scored it 1-3 (definitely not important). “Consensus out” was when >70% of participants scored an item 1-3 and <15% participants scored it 8-10. All other combinations were considered to be ‘no consensus. In addition, interquartile range (IQR) had to be 2 or less to achieve consensus ‘in’ or ‘out’. Range= difference between highest and lowest score given to that item. Items ranked from 1-27 according to mean score (1=highest rank, 27 = lowest rank).

|  |
| --- |
| **Stage of training** |
| 1. The Stage of training e.g. ST1 versus ST2 |
| 2. RCPath curricula and training guidelines |
| **Timing** |
| 3. Placements/cases assessed longitudinally over a period of time. |
| 4. Individual cases in an examination setting |
| **Professionalism** |
| 5. Trainee is organised and timely when conducting themselves in the department e.g. does not lose cases, is aware of turnaround times, triages urgent cases |
| 6. Trainee communicates with all staff appropriately e.g. effectively, timely, politely |
| 7. Trainee works as part of a team e.g. works effectively with colleagues and does not create conflict |
| 8. Trainee is motivated and has a good attitude |
| **Forming a Diagnosis** |
| 9. The diagnosis is correct |
| 10. Trainee consistently produces a correct diagnosis |
| 11. The histopathology report is accurate & does not contain factual errors or omit important information |
| 12. Trainee shows evidence of clinico-pathological correlation |
| 13. Trainee follows sound pathological principles to reach a diagnosis |
| 14. Trainee has appropriate level of basic knowledge. |
| 15. Trainee has appropriate level of up-to-date knowledge e.g. latest papers/datasets |
| **Reflection** |
| 16. Trainee is aware of their own limitations |
| 17. Trainee reflects on their own limitations/performance |
| 18. Trainee shows evidence of improvement following reflection |
| **Trust** |
| 19. Trainee can be trusted to carry out macroscopic examination and “cut-up” independently |
| 20. Trainee can be trusted to report cases (however, the consultant will check the reports and authorise them) |
| 21. Trainee can be trusted to report independently (the consultant does not check the report and the trainee authorises it) |
| 22. Trainee asks for help/second opinion when needed |
| 23. Your opinion of a trainee’s diagnostic competence is related to an overall impression you have of them |
| 24. Your opinion of a trainee’s diagnostic competence is related to bringing all the evidence together and triangulating findings from exams, workplace-based assessments, day-to-day work |
| **The patient’s report** |
| 25. Report is useful to the clinician/surgeon e.g. it does not contain unnecessary information or detail |
| 26. Report ensures the clinician/surgeon receives the appropriate message |
| 27. Report is organised appropriately and well-written |

**Table 5.7. Summary of items which did reach “consensus in” (green) or “no consensus” (red) after round 2.** “Consensus in”>70% of participants scored the item as 8-10 (definitely important) and <15% scored it 1-3 (definitely not important). In addition, the IQR was ≤2. Some items have been shortened for reasons of clarity.

Table 5.7 shows that all the items under “professionalism”, “reflection” and “the patient’s report” reached “consensus in”. In contrast, this only occurred for one of the two items given under the “stage of training” and “timing”. Two items out of seven failed to reach consensus under “forming a diagnosis” and one out seven failed to reach consensus under “trust”.

The five items which did not reach consensus are summarised in Table 5.8 below.

|  |  |  |
| --- | --- | --- |
| **Item** | **Condition 1 met?** | **Condition 2 met?** |
| **Stage of training** | | |
| 2. RCPath curricula and training guidelines | No | Yes |
| **Timing** | | |
| 4. Individual cases in an examination setting | No | Yes |
| **Forming a diagnosis** | | |
| 9. The diagnosis is correct commensurate with the stage of training | No | Yes |
| 10. Trainee consistently produces a correct diagnosis commensurate with the stage of training | No | Yes |
| **Trust** | | |
| 23. Your opinion of a trainee’s diagnostic competence is related to an overall impression you have of them | No | Yes |

**Table 5.8. Summary of items which did not reach consensus after round 2.** Condition 1: “Consensus in” is when >70% of participants scored the item as 8-10 (definitely important) and <15% have scored it 1-3 (definitely not important). “Consensus out” was when >70% of participants scored an item 1-3 and <15% participants scored it 8-10. Condition 2: IQR must be ≤2 to reach consensus in or out. Note, some items have been shortened for reasons of clarity.

Table 5.8 shows that all items which did not reach “consensus in” after round 2 were because condition 1 was not met. Two items did not reach “consensus in” under “forming a diagnosis” while “stage of training”, “timing” and “trust” each had one item fail to reach “consensus in”.

Two participants provided feedback on the rating exercise in round 2:

*“Ability to trim and orientate complex surgical resections according to stage of training. This is often assessed here in WBAs (DOPs). Occasionally trainees have problems with orientating a specimen and take few or too many blocks. Covered in q19 but perhaps more specific requirements needed?”*

*“I think you have covered all the important aspects. Under general professionalism, you could have specifically considered Stress Management, Conflict Management, responding to Own Errors (i.e. reflection and not reflex over-correction like calling then next xxx cases malignant)”.*

As the Delphi was stopped after round 2, these comments could not be used to inform a round 3. No more rounds were conducted following round 2 as stability had been reached. Wilcoxon matched pairs signed rank test showed no significant difference between round 1 and round 2 responses for each of the twenty-seven items, see Table 5.9.

|  |  |  |
| --- | --- | --- |
| **Item** | **p value** | **Result** |
| 1 | 0.999 | Not significant |
| 2 | 0.999 | Not significant |
| 3 | 0.999 | Not significant |
| 4 | 0.765 | Not significant |
| 5 | 0.688 | Not significant |
| 6 | 0.703 | Not significant |
| 7 | 0.812 | Not significant |
| 8 | 0.75 | Not significant |
| 9 | 0.23 | Not significant |
| 10 | 0.62 | Not significant |
| 11 | 0.5 | Not significant |
| 12 | 0.999 | Not significant |
| 13 | 0.999 | Not significant |
| 14 | 0.687 | Not significant |
| 15 | 0.25 | Not significant |
| 16 | 0.5 | Not significant |
| 17 | 0.999 | Not significant |
| 18 | 0.5 | Not significant |
| 19 | 0.187 | Not significant |
| 20 | 0.999 | Not significant |
| 21 | 0.999 | Not significant |
| 22 | 0.5 | Not significant |
| 23 | 0.313 | Not significant |
| 24 | 0.999 | Not significant |
| 25 | 0.609 | Not significant |
| 26 | 0.999 | Not significant |
| 27 | 0.687 | Not significant |

**Table 5.9. Wilcoxon matched pairs signed rank test comparing scores for each item between round 1 and round 2.** 2 tailed, with p<0.05 considered significant

A summary of round 1 and round 2 results is shown in Table 5.10.



**Table 5.10. Comparison of round 1 and round 2 results.** Scores rounded to one decimal place. Under consensus column, green indicates “consensus in” and red “no consensus”. Consensus in this study was defined as “consensus in” when >70% of participants scored the item as 8-10 (definitely important) and <15% have scored it 1-3 (definitely not important). “Consensus out” was when >70% of participants scored an item 1-3 and <15% participants scored it 8-10. All other combinations were considered to be ‘no consensus. In addition, interquartile range (IQR) had to be 2 or less to achieve consensus ‘in’ or ‘out’.

|  |
| --- |
| **Item** |
| **Stage of training**  1. The stage of training e.g. ST1 versus ST5 |
| **Professionalism**  5. Trainee is organised and timely when conducting themselves in the department e.g. does not lose cases, is aware of turnaround times, triages urgent cases |
| **Forming a diagnosis**  15. Trainee has appropriate level of up-to-date knowledge e.g. latest papers/datasets |
| **Trust**  19. Trainee can be trusted to carry out macroscopic examination and “cut-up” independently |
| **Trust**  21. Trainee can be trusted to report independently (the consultant does not check the report and the trainee authorises it) |

Comparing round 1 and round 2 results, 5 more items reached “consensus in” compared to round 1 (Tables 5.10 and 5.11). Five less items reached “no consensus”. All items which had reached “consensus in” in round 1 remained “consensus in” in round 2. In addition, the SD and IQR stayed the same or reduced for 23 of the 27 items showing a move towards central tendency and consensus.

**Table 5.11. List of items which moved from “no consensus” to “consensus in” after round 2**

Table 5.10 indicates that for item 1, the mean score increased slightly (from 8.4 to 8.7) showing it was rated more highly and the SD (1.9 to 1.2) and IQR (3.0 to 1.25) reduced, suggesting increasing agreement between participants. This result is comparable to items 19 and 21 which showed a similar pattern. For item 19 the mean increased from 7.8 to 8.3 and the SD (1.5 to 0.9) and IQR (2 to 1.25) reduced. For item 21, the mean increased from 8.0 to 8.2 and the SD (2.5 to 1.7) and IQR (2.25 to 1.0) reduced. However, for item 5, the mean score was identical (so was rated the same) but the SD (1.6 to 0.9) and IQR (2.25 to 0.25) reduced suggesting increasing agreement. However as the mean stayed the same, some participants must have increased their scores while others reduced theirs. For item 15 the mean score increased from 7.7 to 8.0 showing it was rated more highly but at the same time the SD (0.8 to 0.9) and IQR (1.0 to 1.15) increased suggesting slightly less agreement.

### 5.4.3 Round 1 ranking of items

Table 5.12 shows those items which had the highest mean score and also the items which were ranked as being most important when determining diagnostic competence after round 1. Item 16 ‘Trainee is aware of their limitations’ (under the heading Reflection) appeared on both sides of Table 5.12 (e.g. high mean score and rank), but the other items in the table varied and did not appear on both sides. Two items (16 ‘Trainee is aware of limitations ‘and 18 ‘Trainee shows evidence of improvement after reflection’) both under the heading ‘Reflection’ and one item (26 ‘Report ensures the clinician/surgeon receives the appropriate message’) under the heading ‘The patients report’ received the highest mean scores. For those items ranked highest, one (12 ‘Trainee shows evidence of clinico-pathological correlation) was under the heading ‘forming a diagnosis’, one (16 ‘Trainee is aware of their own limitations’) under ‘Reflection’ and one (8 ‘Trainee is motivated and has a good attitude) under ‘Professionalism’.

|  |  |  |  |
| --- | --- | --- | --- |
| **Highest mean score** | **Item** | **Highest rank (rank 1 = most important)** | **Item** |
| 9.6 | **Reflection**  16. Trainee is aware of limitations | 1 | **Forming a diagnosis**  12. Trainee shows evidence of clinico-pathological correlation |
| 9.6 | **Reflection**  18. Trainee shows evidence of improvement following reflection | 2 | **Reflection**  16. Trainee is aware of their own limitations |
| 9.5 | **The patient’s report**  26. Report ensures the clinician/surgeon receives the appropriate message | 3 | **Professionalism**  8. Trainee is motivated and has a good attitude |

**Table 5.12. Items which had the highest mean score or were ranked as being most important in round 1.** Note, some items have been shortened for reasons of clarity.

Table 5.13 shows those items which had the lowest mean score and also the items which were ranked as being least important when determining diagnostic competence after round 1. Item 23 (Your opinion of a trainee’s diagnostic competence is related to an overall impression) under the heading “Trust” received the lowest mean score and lowest rank. Item 2 (RCPath curricula and training guidelines), under the heading “Stage of training” also appeared on both sides of the table (e.g. low mean score and rank). However, item 9 (The diagnosis is correct) under “Forming a diagnosis” (low mean score) and item 4 (Individual cases in an examination setting) under “Timing” (low rank) did not appear on both sides of the table.

|  |  |  |  |
| --- | --- | --- | --- |
| **Lowest mean score** | **Item** | **Lowest rank (rank 1 = least important)** | **Item** |
| 5.8 | **Trust**  23. Your opinion of a trainee’s diagnostic competence is related to an overall impression | 1 | **Trust**  23. Your opinion of a trainee’s diagnostic competence is related to an overall impression |
| 7.1 | **Stage of Training**  2. RCPath curricula and training guidelines | 2 | **Timing**  4. Individual cases in an examination setting |
| 7.1 | **Forming a diagnosis**  9. The diagnosis is correct | 3 | **Stage of training**  2. RCPath curricula and training guidelines |

**Table 5.13. Items which had the lowest mean score or were ranked as being least important in round 1.** Note, some items have been shortened for reasons of brevity.

### 5.4.4 Round 1 comments relating to ranking items

Where applicable, these comments have been compared to Tables 5.12 (highest mean score or ranked items) and 5.13 (lowest mean score or ranked items) to see their relationship to the results.

*“A good trainee seems no limits on what they might learn or what interests them, they cover the curriculum almost without thinking about it and range well beyond it. Reflective learning – over valued and not applicable to all learning styles (no improvement after reflection is important). Overall impressions are too easily swayed, though as training progresses I hope my overall impression converges with the evidence base”.*

This comment supports item 16. “Trainee is aware of limitations” and item 8. “Trainee is motivated and has a good attitude” which are under the headings ‘Trust’ and ‘Professionalism respectively. These were both rated and/or ranked highly in Table 5.12. However it does not support item 18. “Trainee shows evidence of improvement following reflection” (under the heading ‘Reflection’) which received the joint highest mean score for importance.

*“I put 25 first because a trainee’s report is structured around exam requirements and tends to include superfluous information which can be omitted once they are senior. In theory trainees could write an exam version and another version to be sent out to the clinicians, but in practice there is not time to fiddle with that”.*

This comment does not provide insight into the ranking exercise but suggests expectations change during training

*“As above-there are other considerations as to overall professional behaviour in addition to competence in diagnosis. I have found this exercise difficult and almost all of them could be discussed from "well if you look at it this way" point of view.”*

This comment supports items relating to professionalism (item 8. “Trainee is motivated and has a good attitude”) and reflection (items 16. “Trainee is aware of limitations” and 18. “Trainee shows evidence of improvement following reflection”) which were rated or ranked highly (see Table 5.12). In contrast, although the comment suggests competence in diagnosis may be important, item 9. “The diagnosis is correct”, received the third lowest mean score (Table 5.13).

*“It is essential to consider the stage of training. Trainees need to receive positive constructive feedback, so even if diagnosis is wrong, improvement happens.”*

The second sentence of this comment supports item 18. “Trainee shows evidence of improvement following reflection” which received the joint highest mean score (Table 5.12).

*“Motivation and self-awareness are really important.”*

This comment supports item 8. “Trainee is motivated and has a good attitude” which was ranked the third most important (Table 5.12).

*“My least important selection is a bit arbitrary as all are important, so would have preferred an N/A choice”.*

Given that 93% of items had mean scores of 7 or more, the comment does provide support that many statements are important.

### 5.4.5 Round 2 ranking of items

Table 5.14 shows those items which had the highest mean score and also the items which were ranked as being most important when determining diagnostic competence after round 2. Item 16 (Trainee is aware of their own limitations) under the heading “Reflection” was the only item to appear on both sides of the table (i.e. a high mean score and rank). However, the other items which received the highest mean score differed from those which were ranked highly. For example, item 26 (Report ensures the clinician/surgeon receives the appropriate message) under “The patient’s report” and item 18 (“Trainee shows evidence of improvement following reflection”) under “Reflection” received the highest and third highest mean score, respectively. However, they did not rank as highly, with item 13 (“Trainee follows sounds pathological principles to reach a diagnosis”) under “Forming a diagnosis” and item 22 (“Trainee asks for help/second opinion when needed”) under “Trust” being the 2nd and 3rd highest ranked items, respectively.

|  |  |  |  |
| --- | --- | --- | --- |
| **Highest mean score** | **Item** | **Highest rank (rank 1 = most important)** | **Item** |
| 9.4 | **The patient’s report**  26. Report ensures the clinician/surgeon receives the appropriate message | 1 | **Reflection**  16. Trainee is aware of their own limitations |
| 9.3 | **Reflection**  16. Trainee is aware of their own limitations | 2 | **Forming a diagnosis**  13. Trainee follows sound pathological principles to reach a diagnosis |
| 9.3 | **Reflection**  18. Trainee shows evidence of improvement following reflection | 3 | **Trust**  22. Trainee asks for help/second opinion when needed |

**Table 5.14. Items which had the highest mean score or were ranked as being most important in round 2.** Note, some items have been shortened for reasons of clarity.

Table 5.15 shows those items which had the lowest mean score and also the items which were ranked as being least important when determining diagnostic competence after round 2. Items 2 (“RCPath curricula and training guidelines”) under “Stage of training”, 4 (“Individual cases in an examination setting’) under “Timing” and 23 (“Your opinion of a trainee’s diagnostic competence is related to an overall impression”) under “Trust” received both the lowest mean scores and were ranked least important. In addition item 23 received the lowest mean score and was ranked jointly with item 4 as the least important.

|  |  |  |  |
| --- | --- | --- | --- |
| **Lowest mean score** | **Item** | **Lowest rank (rank 1 = least important)** | **Item** |
| 6.2 | **Trust**  23. Your opinion of a trainee’s diagnostic competence is related to an overall impression | =1 | **Trust**  23. Your opinion of a trainee’s diagnostic competence is related to an overall impression |
| 6.8 | **Timing**  4. Individual cases in an examination setting | =1 | **Timing**  4. Individual cases in an examination setting |
| 7 | **Stage of training**  2. RCPath curricula and training guidelines | 3 | **Stage of training**  2. RCPath curricula and training guidelines |

**Table 5.15. Items which had the lowest mean score or were ranked as being least important in round 2.** Note, some items have been shortened for reasons of clarity.

### 5.4.6 Round 2 comments relating to ranking items

Where applicable, these comments have been compared to Tables 5.14 (highest mean score or ranked items) and 5.15 (lowest mean scores and ranked items) to see their relationship to the results.

*“Repeat scoring is interesting – is there anchoring bias here – I do like the concept that one can change their mind having seen the bigger picture though. Ranking the 3 least important more difficult than most important. If a trainee completes with sound principles in box 13 then their training has been successful in my opinion. If they are good at 13 with the aid of 10 then that is even better – they will be good consultants”.*

This comment supports item 13. “Trainee follows sound pathological principles to reach a diagnosis” which was ranked as the second most important item (Table 5.14). However, in contrast, item 10 “Trainee consistently produces a correct diagnosis commensurate with the stage of training” did not score highly (mean score 7.7) nor reach “consensus in” after 2 rounds (Table 5.8).

*“Snapshots and adherence to tick boxes (ticking off the curriculum etc.) is of limited benefit and can be gamed. Accuracy, the basis of which is understood and reflected on is what is required.”*

This comment supports item 16 “Trainee is aware of their own limitations” and item 18. “Trainee shows evidence of improvement following reflection” which were among the three highest rated and ranked items – see Table 5.14.

The following comments do not help to explain the ranking but suggest the exercise was not straight-forward.

*“I made a reading mistake on 23 first time that I rated it”.*

*“I have steered away from least important. College criteria are likely to change with new curriculum. Is competence the same as exam-based proof of ability? I’d argue not – but this doesn’t devalue the role of the exam”!*

*“24 really important but is a summary so I haven’t put it in the first ranking”.*

*“As before, difficult to decide on least important items.”*

### 5.4.7 Comparison of round 1 and round 2 ranking of items

Tables 5.16 and 5.17 compare the highest and lowest ranked items respectively between rounds 1 and 2. In Table 5.16, only item 16 (‘Trainee is aware of their own limitations’) under “Reflection” was ranked within the top three for both rounds. The other items differed. In round 1 items 12 (‘Trainee shows evidence of clinico-pathological correlation’) under “Forming a diagnosis” and 8 (‘Trainee is motivated and has a good attitude) under “Professionalism” were rated highly whereas in round 2, items 13 “Trainee follows sound pathological principles to reach a diagnosis” (under “Forming a diagnosis”) and 22 (‘Trainee asks for help/second opinion when needed’) under “Trust” were considered more important.

|  |  |  |  |
| --- | --- | --- | --- |
| **Round 1** | **Rank** | **Round 2** | **Rank** |
| **Forming a diagnosis**  12. Trainee shows evidence of clinico-pathological correlation | 1 | **Reflection**  16. Trainee is aware of their own limitations | 1 |
| **Reflection**  16. Trainee is aware of their own limitations | 2 | **Forming a diagnosis**  13. Trainee follows sound pathological principles to reach a diagnosis | 2 |
| **Professionalism**  8. Trainee is motivated and has a good attitude | 3 | **Trust**  22. Trainee asks for help/second opinion when needed | 3 |

**Table 5.16. Comparison of highest ranked items in rounds 1 and 2.** Note, some items have been shortened for reasons of clarity.

In Table 5.17, items 2 (‘RCPath curricula and training guidelines’) under “Stage of training”, 4 (‘Individual cases in an examination setting’) under “Timing” and 23 (‘Your opinion of a trainee’s diagnostic competence is related to an overall impression’) under “Trust” all ranked within the bottom three for both rounds. However, item 4 became relatively more important in round 2 as it was ranked joint first with item 23 whereas it was ranked second in round 1.

|  |  |  |  |
| --- | --- | --- | --- |
| **Round 1** | **Rank** | **Round 2** | **Rank** |
| **Trust**  23. Your opinion of a trainee’s diagnostic competence is related to an overall impression | 1 | **Trust**  23. Your opinion of a trainee’s diagnostic competence is related to an overall impression | =1 |
| **Timing**  4. Individual cases in an examination setting | 2 | **Timing**  4. Individual cases in an examination setting | =1 |
| **Stage of training**  2. RCPath curricula and training guidelines | 3 | **Stage of training**  2. RCPath curricula and training guidelines | 3 |

**Table 5.17. Comparison of lowest ranked items in rounds 1 and 2**.Note, some items have been shortened for reasons of clarity.

### 5.4.8 Comparison of round 1 and round 2 highest rated by mean score

Tables 5.18 and 5.19 compare the highest rated items by mean score for rounds 1 and 2. In Table 5.18, it can be seen that items 16 (‘Trainee is aware of limitations’), 18 (‘Trainee shows evidence of improvement following reflection’) and 26 (‘Report ensures the clinician/surgeon receives the appropriate message’) had the highest scores in both rounds 1 and 2. In round 1, items 16 and 18 (under “Reflection”) appeared in first and second positions, respectively. However, in round 2, item 26 (under “The patient’s report”) became more important and had the highest mean score (9.4). Overall, the mean score for all three items fell from round 1 to round 2.

|  |  |  |  |
| --- | --- | --- | --- |
| **Round 1** | **Mean score** | **Round 2** | **Mean score** |
| **Reflection**  16. Trainee is aware of limitations | 9.6 | **The patient’s report**  26. Report ensures the clinician/surgeon receives the appropriate message | 9.4 |
| **Reflection**  18. Trainee shows evidence of improvement following reflection | 9.6 | **Reflection**  16. Trainee is aware of their own limitations | 9.3 |
| **The patient’s report**  26. Report ensures the clinician/surgeon receives the appropriate message | 9.5 | **Reflection**  18. Trainee shows evidence of improvement following reflection | 9.3 |

**Table 5.18. Comparison of highest rated items by mean score in rounds 1 and 2.** Note, some items have been shortened for reasons of clarity.

In Table 5.19, it can be seen that item 23 (‘Your opinion of a trainee’s diagnostic competence is related to an overall impression’) under “Trust” received the lowest mean score in both rounds 1 and 2. However, its score did rise by 0.4 points between rounds. Item 2 (‘RCPath curricula and training guidelines’) under “Stage of training” received a low score in rounds 1 (second lowest) and 2 (third lowest). Item 9 (‘The diagnosis is correct ‘) under “Forming a diagnosis” appeared as the third lowest scored item in round 1 (7.1) but received a higher mean score of 7.8 in round 2 so was no longer one of the lowest scored items (Table 5.10). Item 4 (‘Individual cases in an examination setting’) under “Timing” did not appear in the bottom three in round 1 but did in round 2 (its mean score fell from 6.9 to 6.8).

|  |  |  |  |
| --- | --- | --- | --- |
| **Round 1** | **Mean score** | **Round 2** | **Mean score** |
| **Trust**  23. Your opinion of a trainee’s diagnostic competence is related to an overall impression | 5.8 | **Trust**  23. Your opinion of a trainee’s diagnostic competence is related to an overall impression | 6.2 |
| **Stage of Training**  2. RCPath curricula and training guidelines | 7.1 | **Timing**  4. Individual cases in an examination setting | 6.8 |
| **Forming a diagnosis**  9. The diagnosis is correct | 7.1 | **Stage of training**  2. RCPath curricula and training guidelines | 7.0 |

**Table 5.19. Comparison of lowest rated items by mean score in rounds 1 and 2.** Note, some items have been shortened for reasons of clarity.

## 5.5 Discussion

The results indicate that the majority of items identified in the qualitative interviews were also valued by the Delphi panel with twenty-two of the twenty-seven items reaching “consensus in” after two rounds. This helps to triangulate and validate our previous findings. Only five items failed to reach “consensus in” and of those five, none were deemed unimportant enough to be scored “consensus out”. Of the five items which were “no consensus”, items 2, 9, 10 still had mean scores of 7 or above, suggesting they were closer to “definitely important” than “definitely not important” even though the criteria for consensus were not met. The remaining “no consensus” items were 4 and 23 which received the lowest mean score, with mean scores of 6.75 and 6.2 respectively.

There appears to be consistency between the items which had the lowest mean score and also those items which were separately ranked as being least important after two rounds (Table 5.15). These items also didn’t reach consensus “in or “out”. These were items 2, 4 and 23 which are shown below:

Item 2. Under ‘Stage of training’ RCPath curricula and training guidelines (mean score 7, rank 3)

Item 4. Under ‘Timing’ Individual cases in an examination setting (mean score 6.8, rank =1)

Item 23. Under ‘Trust’ Your opinion of a trainee’s diagnostic competence is related to an overall impression you have of them, rather than items of evidence e.g. gut-feelings/impressions/ your perceptions about trainees (mean score 6.2, rank=1)

The qualitative interviews identified that curricula and training guidelines were used in part to make judgements on trainee diagnostic competence (item 2). Even though it was one of the lowest valued items, it still appeared relatively important because it had a mean score of 7 after round 2. However, there may be an element of a consultant being able to judge diagnostic competence without the use of curricula and guidelines. Indeed, the panel members were all considered “experts” so their extra experience may mean they are very familiar with training pathways and don’t need to refer to guidelines or curricula as frequently as others. Given the comment “*College criteria are likely to change with new curriculum”* it may be that several participants were aware that the curriculum is due to change soon and this influenced how they scored the item. Whether these views represents the wider view (e.g. consultants who are less experienced) is unclear.

Of note, the qualitative interview data suggested that competence is best judged longitudinally and not through “individual cases in an examination setting” as stated in item 4. So the fact that it received a low score is actually in line with our previous findings. In support of this, item 3 “Placements/cases assessed longitudinally over a period of time” had a mean score of 8.5 and reached “consensus in”.

Item 23 stemmed from qualitative interview data which suggested “feelings” are also taken into account when determining if trainees are diagnostically competent, such as: *“Quite often people do have a feeling but they cannot pin it down to something”* and *“So if you've got a nagging, you've always got a sort of nagging uncertainty that they would ask you for help when they are unsure, then it's difficult to let go of the reins”.* It appears that the Delphi panel were rather equivocal with regard to whether this was important when judging trainee diagnostic competence. It is possible that their experience (s) made them unconsciously able to measure competence and they didn’t recognize this ability or it being important. In contrast to item 23, item 24 had a mean score of 9.2 which related to trainee diagnostic competence being “related to bringing all the evidence together and triangulating findings from exams, workplace-based assessments and day-to-day work.” It therefore appears that assimilating the evidence is more important than “feelings” even though “feelings” do sometimes exist. In addition, given that many of the Delphi panelists were TPDs it is understandable that they rated item 24 highly, as it is the exact process they have to follow at the annual ARCP. However, it could also be argued that although “feelings” do not directly dictate decisions on competency, they might indirectly influence what evidence is collected by trainers as a means to justify decisions regarding competency.

It is worth noting that items 2, 4 and 23 (discussed above) received similarly low scores in round 1 (Table 5.13). However, in round 1, item 9 “The diagnosis is correct” appeared in Table 5.13 as it had the third lowest mean score of 7.1. In round 2, this rose to 7.83 and therefore no longer remained in the bottom three by mean score. This is likely due to an amendment made to item 9 based on participant feedback, whereby item 9 was changed to “The diagnosis is correct commensurate with the stage of training”. This shows that the expectations change during training and is in accordance with the findings from the qualitative interviews.

Focusing on those items which had the highest mean scores and rank after two rounds, these are shown in Table 5.14 and again, below.

‘Forming a Diagnosis’

13. Trainee follows sound pathological principles to reach a diagnosis e.g. the trainee understands how they reached the diagnosis, has a consistent, accurate way of examining slides and knows the steps involved when making a diagnosis. The diagnosis has not stemmed from guessing. (Ranked as second most important item)

‘Reflection’

16. Trainee is aware of their own limitations. (Received joint second highest mean score of 9.3 and highest ranked item)

18. Trainee shows evidence of improvement following reflection. (Received joint second highest mean score of 9.3)

‘Trust’

22. Trainee asks for help/second opinion when needed. (Ranked as third most important item)

‘The patients report’

26. Report ensures the clinician/surgeon receives the appropriate message e.g. when diagnoses are not clear-cut and a differential diagnosis or other message needs to be conveyed. This might include a discussion about the limitations of the biopsy or if further tests or information are needed. (Received highest mean score of 9.4)

One can see from Table 5.14 that only item 16 received both one of the highest mean scores and ranks. Items 26, 18, 13 and 22 only received a top 3 highest mean score or top 3 rank, but did not appear at the top of both lists unlike item 16. However, it is not surprising that several items appear to be clustered at the top for most important, as nine items all had mean scores of 9 or above so were considered “definitely important”. In reality, this makes no significant difference to how the results are interpreted but suggests that there are numerous items highly valued by the participants which helps to triangulate the qualitative interview data. It also demonstrates the complexity of assessing diagnostic competence.

Two of these items, 16, 18, relate to the theme of “Reflection” and item 22 relates to the theme of “Trust”. However, it could be argued that item 22 (Trainee asks for help/second opinion when needed) is a function of item 16 (Trainee is aware of their own limitations) as it is this insight which should ensure trainee’s ask for help when they need it. Consultants appear to highly value trainees who know their own limitations, do not act beyond their competency and know when to ask for help. As mentioned previously, ‘following sound pathological principles’ was also ranked highly (item 13), reflecting that understanding of cases and the underpinning elements that support a diagnosis are highly valued. Simply getting a diagnosis correct on its own without the understanding was not valued highly by consultants as evidenced by item 9 (“The diagnosis is correct commensurate with the stage of training”) which failed to reach “consensus in”. These views also have links to item 26 which “ensured the clinician/surgeon received the appropriate message” when diagnoses were not clear-cut. Hence, one’s understanding of the case allows one to convey uncertainty in diagnosis. This is also mirrored in the interview data from the previous chapter with comments such as “*But also having the insight to be able to recognize when I can't make a diagnosis or when there's a grey area and I have to be able to express that ambiguity”.*

Items, 16, 18, 26 were consistently either scored or ranked highly for both rounds 1 and 2 but others were not. For example, item 12. “Trainee shows evidence of clinio-pathological correlation” ranked highly in round 1 but did not rank as highly in round 2. Item 12 is closely related to item 13 “following sound pathological principles” which actually ranked second highest. Similarly, item 8. “Trainee is motivated and has a good attitude” ranked highly in round 1 but did not rank as highly in round 2. It is not clear why these differences occurred as the wording of items 12 and 8 were unchanged between rounds. The mean scores for items 12 and 8 dropped slightly (both fell from 9.17 to 9.0) between rounds but as the ranking exercise was independent of mean score, it cannot be used to fully explain the difference. It may reflect that opinion will fluctuate between rounds but these differences (at least when comparing median scores) were not statistically significant. Another possible explanation is that participants were influenced by the opinions of others when they looked at the results of the first round.

To this end, although stability testing showed that there was no significant difference in the rating of individual items between rounds the results show that scores for items did change between rounds. If a measure of stability hadn’t been used, perhaps a round 3 would have been utilized. Whether or not different results would have been obtained is impossible to know, but given that the mean scores between rounds 1 and 2 (see Table 5.10) were so similar, the IQR and SD reduced for nearly all items (Tables 5.10) and the amount of written feedback reduced between rounds, it is probable a round 3 would not have changed the results substantially.

However, if one examines those items which changed from “no consensus” to “consensus in” in more detail, it is possible to understand the affect the fluctuations did have on the outcomes from the Delphi. For example, after round 2, a further five items moved from “no consensus” to “consensus in” (Table 5.11). Examining Table 5.5, it can be seen that in round 1 these items did not reach consensus either because 70% of participants did not score the items 8-10 or because the IQR was >2. However, after round 2 these items did reach “consensus in” as the participants rated these items more favourably. This suggests participants reflected on their scoring from round 1 and/or other participant’s scores. Indeed, the IQR was either the same or reduced for 24 of the 27 items and the SD was either the same or reduced for 23 of the 27 items (Table 5.10). This suggests there was increasing agreement between rounds 1 and 2. Interestingly, of the items which did not show a reduction in IQR and/or SD, all but one (item 15) had reached “consensus in” after round 1, with item 15 also reaching “consensus in” after round 2. This is because the fluctuations in IQR were not affecting the definitions that were stipulated for “consensus in” or “consensus out”. For example, item 15 “Trainee has appropriate level of up-to-date knowledge e.g. latest papers/datasets” did not reach “consensus in” after round 1 because 70% of participants did not score the items 8-10 and not because of the IQR as this was 0.91 (so was less than the stipulated ≤2). In round 2, 70% of participants did score the item 8-10 and even though the IQR rose to 1.25 it was still ≤2 so did not affect the item reaching “consensus in”. The criteria that was agreed to determine consensus was therefore sensitive enough to detect fluctuations in opinion. However, it could be argued that a different measure of consensus might have given a different list of items reaching consensus. Although this may be true, if the items which reached consensus are ignored, it is still possible to appreciate that many items were rated highly by the participants. Indeed, determining the importance of the items was one of the primary objectives of this chapter, whether they reached consensus or not was useful to know but does not necessarily mean that those items that did not reach consensus are not important. It might just reflect that they are either less important than other items or that the definition for consensus was too strict for these items to reach “consensus in”. For example, for all the items which did not reach consensus after round 2 (Table 5.8) this was because 70% of participants did not score the items 8-10. If the definition for “consensus in” was made more lenient, such that 70% of participants had to score the item 7-10, then a further two items would have reached “consensus in”. This illustrates that the outcomes from a Delphi must be interpreted in light of the definition of consensus. For example, if, as in this study, the Delphi is triangulating results from previous research, perhaps the definition need not be too problematic. However, if a high stakes decision is to be based on the Delphi outcomes, it is important to be certain that the definition chosen for consensus is appropriate for that context.

Many of the items in this study have significant overlap and impact on one another, which is reflected in a Delphi participant commenting that “*Some of these are complexly interrelated and whilst on their own are not important, are functions of other traits”.* Therefore, scoring and ranking items, in some ways, oversimplifies how complex the judgement task is. It is also reflected in some participants struggling to rank the least important items. For example, one participant stated *“Ranking the 3 least important more difficult than most important”* and another commented *“As before, difficult to decide on least important items”.* Although this suggests that there were no items which definitely stood out as not being relevant to diagnostic competence, it shows that many of the items are important to some degree and in the right context, which might explain the aforementioned inconsistencies for mean score and rank between rounds. This is reflected in another comment which explained: “*I have found this exercise difficult and almost all of them could be discussed from "well if you look at it this way" point of view.”* Hence, our conceptual model is also complex, incorporating both “process” and “person” qualities (which in themselves contain rich data) while taking into account stage of training, time and the development of trust.

Feedback on the list of 27 items by participants after round 1 mostly indicated that “Stage of training” was relevant to make a judgement on the importance of certain items. When creating the initial set of 27 items, it was reasoned that some items could be qualified further by adding the stage of training. However, it was decided not to do this to see if Delphi participants would identify issues themselves and not be too led by the research team. It is therefore useful to see that the Delphi panelists identified stage of training as being integral to the judgement process. The stage of training is also present in our conceptual model of diagnostic competence so these comments helped to corroborate our findings.

As previously noted, getting “the diagnosis correct commensurate with the stage of training” (item 9, round 2 mean score 7.8) or even consistently correct (item 10, round 2 mean score 7.7), did not reach “consensus in”. Items 9 and 10 still received relatively high scores so are still somewhat important to diagnostic competence, but not as highly valued as other items, such as reflection and having sound pathological principles to follow. It is not clear why this is, but it may be that getting diagnoses correct only suggests your competency in that moment (or a correct “guess”), whereas reflection ensures you will continue to act within your competency and maintain that competency. The qualitative interview data suggest that getting the correct diagnosis becomes more important in the latter stages of training, but even with the amendment to items 9 and 10 after round 1 to incorporate “commensurate with the stage of training”, these items still did not reach “consensus in”. However, the qualitative interview data did suggest that the approach to diagnosis and understanding how the diagnosis is reached, are actually more important than the end product; the diagnosis. This is borne out in the Delphi results where item 13 “trainee follows sound pathological principles” was the second highest ranked item and has previously been discussed.

Importantly, no new items were provided by participants which suggests the qualitative interviews explored diagnostic competence to the correct depth. However, some participants did suggest alternative ways of phrasing items or examples that could be provided to add detail, for example: *“I think you have covered all the important aspects. Under general professionalism, you could have specifically considered Stress Management, Conflict Management, responding to Own Errors (i.e. reflection and not reflex over-correction like calling then next xxx cases malignant)”.* Creating the 27 items was always a balance between clarity and wordiness, as it was feared that if the items were overly complicated, the Delphi might have become onerous to complete. Also, stress management, conflict management and responding to own errors were thought to be incorporated in items 17. “Trainee reflects on their own limitations/performance”, Item 6. ‘Trainee communicates with all staff appropriately e.g. effectively, timely, politely’ and Item 7. ‘Trainee works as part of a team e.g. works effectively with colleagues and does not create conflict’. However, the comment does indicate that specific examples can help assessors to interpret forms correctly and any new assessment tools which stem from this research will need to consider this.

It can be seen from Table 5.7 that all the items under “The patient’s report”, “Reflection” and “Professionalism” reached “consensus in” suggesting that these categories were particularly important to participants. In contrast, “Stage of training”, “Timing” and “Trust” each had one item which did not reach ‘consensus in’ and “Forming a diagnosis” had two. The reasons for these differences are discussed below.

As mentioned previously, item 4 “individual cases in an examination setting” under “Timing” did not reach “consensus in”. This is not surprising as our qualitative data suggested competence should be assessed longitudinally rather than on a single episode. In support of this item 3 ‘Placements/cases assessed longitudinally over a period of time’ did reach consensus in. Thus it would not be expected that item 4 would reach consensus.

Other categories had items that did not reach “consensus in” and there are a number of explanations why this could be. Firstly, the simplest explanation is that these items are not considered as important as the other items that did reach “consensus in”. However, if they were completely irrelevant items, then it may be assumed they would have reached “consensus out” which they did not. Indeed, the conceptual model did not aim to ascribe relative importance values to the themes and sub-themes which emerged from the interview data, so it is probable that some items will be valued more than others. Secondly, perhaps the scores reflect that scoring items in a Delphi does not exactly match what individuals do in their real practice (from which the qualitative data and these items stemmed from). Thirdly, perhaps those items not reaching “consensus in” reflects that the Delphi participants value the wording of items differently to those who were interviewed. This is unlikely though, as TPDs were represented in the interviews and the Delphi and saturation of data had been reached. This is further demonstrated by the fact that no new items were suggested through the Delphi and no items reached “consensus out”.

Finally, although most participants scored items similarly, items 21 and 22 were scored quite differently by one participant. These items were:

Item 21. Senior trainees can be trusted to report independently (the consultant does not check the report and the trainee authorises it).

Item 22. Trainee asks for help/second opinion when needed.

Item 21 had a median score of 8 but was given a score of 3 by a single participant. The participant did not offer any feedback regarding the low score after rounds 1 or 2. It is unclear why they took this stance. A possible explanation is that they feel independent reporting is not used routinely during training so is not something they use to determine diagnostic competence.

Item 22 had a median score of 10 but was given a score 1 by a single participant. The participant did not offer any feedback following round 1 regarding why they had given such a low score. However, having seen the mean panel score in round 2, and maintaining a score of 1, the participant explained *“I strongly believe trainees should be discouraged to do this as the trainer then is not aware of the trainees’ gap in knowledge and a very useful opportunity for teaching is missed”.* This opinion was distinctly different to the other participants. It probably reflects that some trainees may talk to other trainees about cases before getting them checked by the consultant. While this does mean the trainer might not be as aware of “gaps in knowledge”, it could be argued that the trainee is actively learning by discussing cases and may feel more comfortable asking questions to their peers in the first instance. It is also very difficult to stop trainees from talking to one another.

## 5.6 Limitations

A consensus is only a pre-defined level of agreement which may not reflect the complexity of the judgement or that full agreement has actually been reached. However, the decreasing IQR and SD for most statements as well as the reduced number of comments between the first and second rounds support that agreement had been reached and further rounds were unlikely to add further information as the results were also statistically stable.

The results from the Delphi only indicate agreement with items and further work is needed to determine the pedagogic approach required to deliver these components within assessments.

All participants for the Delphi study were purposefully sought but eight of them required personal emails in order for them to agree to be part of the study. It could be argued that those who do not volunteer may have lower motivation to accurately score and rate items within the Delphi study. Although the response rate was 100%, multiple reminder emails had to be sent and many participants provided little feedback to justify their scoring. A deeper insight into the Delphi items may have been gained with more feedback from participants, but as the main purpose of the Delphi was to triangulate our findings, and a great deal of insight into the issues had been explored through the interviews, the Delphi nevertheless achieved its aims.

## 5.7 Conclusion

The Delphi has triangulated our findings from the qualitative interviews, in that the items presented appear to be highly valued by the Delphi participants as well. No new items were suggested by participants, suggesting the qualitative interviews explored diagnostic competence in sufficient depth. Three factors were highlighted by the Delphi panel which relate to the judgement ecology and which were also help to corroborate the theoretical model of diagnostic competence provided in the previous chapter. Firstly, the stage of training is very important to the judgement of diagnostic competence as the majority of comments from participants mentioned this factor. The stage of training reflects that competence develops with experience and expectations increase with time. Secondly, several participants stated that it was difficult to rate items or that they correlated with other items in some way. This reflects the complex nature of the judgement ecology which we have previously described. It also helps to justify why case vignettes would not have been appropriate methodology to judge the importance of the items as manipulating individual qualities would have oversimplified the judgement. Thirdly, participants gave the lowest score to item 4 which read “individual cases in an examination setting”. This further highlights that the judgement ecology is longitudinal in nature and competency is best measured over time.

# **Chapter 6**

# **Final discussion**

**Chapter 6: Final discussion**

## 6.1 Discussion

Within the medical profession and its affiliated specialties, there is a need to train doctors and dentists of the future who are competent to treat and care for patients. Hence, assessments purported to fulfil this objective must be fit for purpose and also provide a quality training experience for trainees (Brightwell and Grant 2013).It appears logical to therefore presume that these assessments are created from sound educational research and are acceptable to both the stakeholders and trainees.

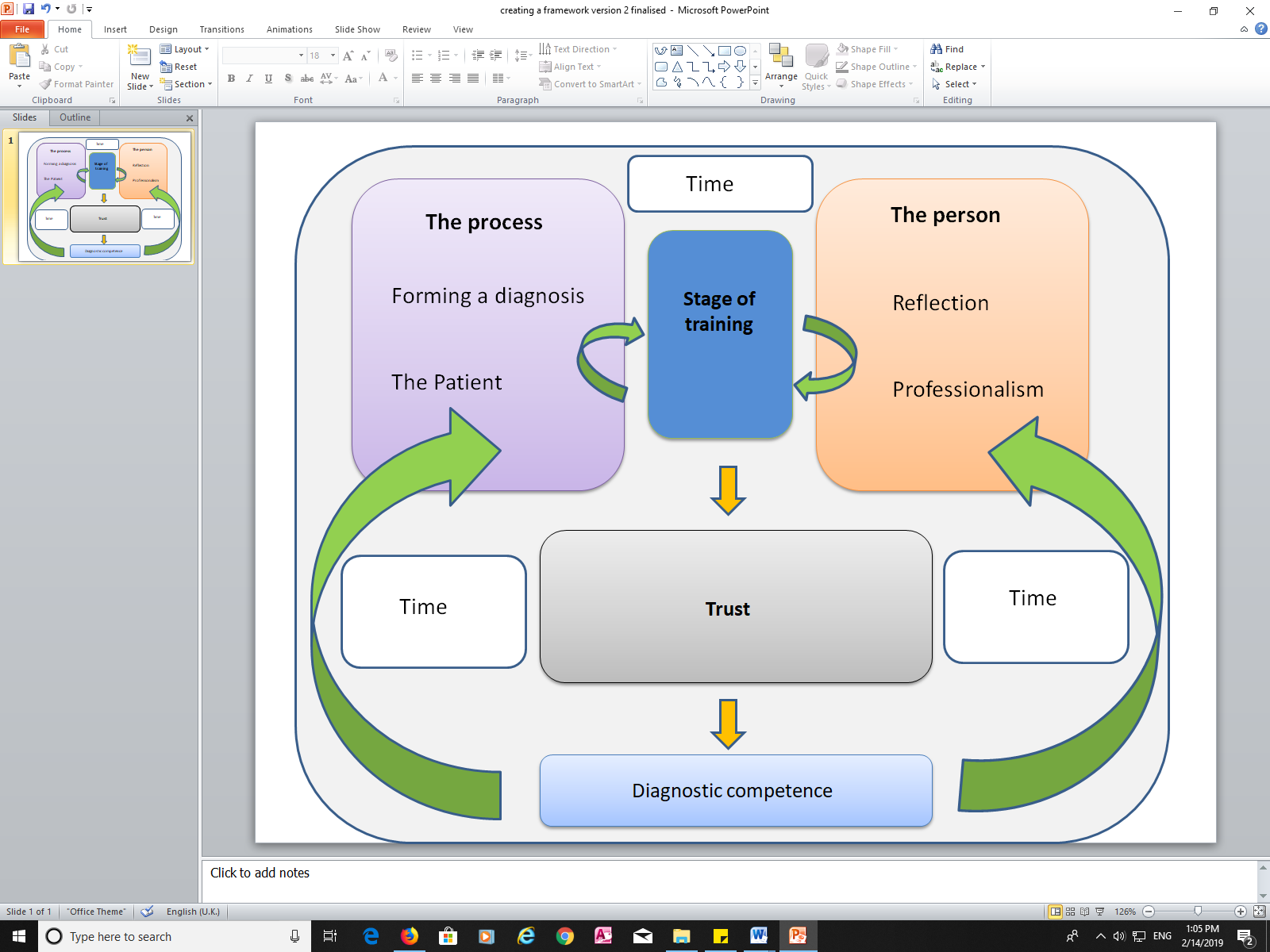
However, despite WBAs being a familiar part of specialist training in histopathology the healthcare literature suggests there are numerous issues related to these assessment tools, with very little evidence justifying their use in histopathology. Coupled with this, it was also not clear from the literature what aspects of knowledge, skill and behaviour constitute competence in histopathology. Without a clear understanding of what competence is, it seemed reasonable to assume the current WBAs would not be able to measure this “unknown” entity.

Thus the aims and objectives of this thesis were therefore to:

1. To explore the perceptions of WBAs in histopathology by assessors and trainees using a questionnaire to:
2. understand how WBAs are perceived in histopathology including a comparison to the themes identified in the healthcare literature: their perceived usefulness, validity and acceptance.
3. understand if WBAs in histopathology are perceived to measure diagnostic competence
4. To explore diagnostic competence in histopathology by means of a qualitative interview study using Framework Method to understand:
5. what diagnostic competence encompasses in histopathology
6. the cues consultants look for in their trainees when determining diagnostic competence
7. the judgement ecology which pertains to determining diagnostic competency in trainees
8. To use a Delphi study to triangulate the findings from the qualitative interviews in assessing diagnostic competence to:
9. determine if the cues identified from the qualitative interviews are valued by experts.
10. determine the relative importance of individual cues
11. identify any additional cues

These objectives were achieved. Firstly, the questionnaire found multiple perceived issues with the use of WBAs in histopathology training. In particular, they were perceived by trainees and consultants not to measure diagnostic competence. These findings add to the literature, where previously, there was not a single study examining the perception of all WBAs in histopathology. It further justified the need to explore and understand exactly what diagnostic competence entails so the issues raised in the questionnaire outcomes could be understood fully. For instance, in Chapter 1, the literature demonstrated that competence was a complex entity to understand with many different meanings and contexts to consider. Hence, the underlying constructs related to histopathological diagnostic competence were unknown and needed to be identified to form a valid basis from which any future assessment tools could be fashioned.

The qualitative interviews which followed were able to generate data which described the qualities most relevant to histopathology as well as the nature and context in which judgements about trainees are made. Diagnostic competence emerged as a complex construct that should be assessed longitudinally. Elements of both “process” and “person” are taken into account with expectations rising through training until sufficient time has passed for the consultant to trust that the trainee is diagnostically competent to report independently and act autonomously. Previously, there were no studies in the literature which clearly described these qualities or how consultants judge diagnostic competence in their trainees. While one study (Finall and Allery 2016) examined the learning needs required for independent reporting to occur in histopathology, it suffered several limitations (see Chapter 1, section 1.2.10.4). Furthermore, independent reporting is only one indicator of being competent and the perceived learning needs may not necessarily be the same as the qualities consultants look for in their trainees. To this end, while the aforementioned study shared some of the qualities which were also identified in our data (such as accumulation of experience, recognizing normal, effects of pathology results on patients and being aware of diagnostic pitfalls), others such as professionalism and reflection were not. The relationships between these different factors was also not explored. The resulting conceptual model that has been described in this thesis illustrates the relationship between these qualities and the judgement ecology, adding a significant amount of educational research to help address how competency is actually perceived in histopathology and therefore how it should be assessed (Figure 6.1).



**Figure 6.1. Conceptual model of diagnostic competence**

Finally, our findings were triangulated with a Delphi study which also aided our understanding of the judgement ecology. The majority of the qualities identified in the interview data reached “consensus in” while the remaining items generally scored highly. This suggests that our findings were in agreement with the perceptions of TPDs, representatives from RCPath and members from the postgraduate deanery. The ability to reach consensus with these stakeholders suggests the data to be valid and useful to histopathology training.

The research clearly highlights that a longitudinal approach is needed for consultants to determine diagnostic competence in trainees, signifying that it is important that future assessment strategies do not examine competence only on the basis of single events or cases. RCPath currently employ a competency-based curriculum similar to many other educational programmes (see Chapter 1, section 1.29), but identifying individual competencies is at odds with our data. Consultants appear interested in the “bigger picture” and what they can gauge from a trainee over a longer period of time. For example, participants from the interviews and Delphi did not favour exams as a measure of competency, but instead valued repeated demonstrations of “process” and “person” to develop trust. This echoes the sentiments of Oerlemans and colleagues who found that clinical supervisors appreciate consistent behaviours when assessment of a trainee is based on a series of observations (2017). However, it is important to stress that examinations still have a role in training. While they may not be good at measuring qualities such as professionalism or reflection, they can provide an external, quality assured assessment to determine if trainees are able to apply their knowledge in their field of practice. Indeed, the RCPath acknowledge that a number of different assessments should be used during training as they each serve different purposes (Royal College of Pathologists 2015).

It has been recognised that implementing outcome-based medical education can be challenging in terms of translating the complexities of medical practice into meaningful assessment strategies and curricula (Frank et al. 2010a). These complexities are clearly reflected in the data supporting our conceptual model of diagnostic competence, where the various factors involved are all linked to one another. By separating these into competencies, the complexities of performance are unlikely to be captured. This is in agreement with Durning et al who suggest that the whole is greater than the sum of its parts (2010). Assessing diagnostic competency longitudinally mirrors another aspect of our conceptual model, which is that it is the consistency of competency which is particularly important, and not “one-off” demonstrations of competency. This is because the judgement ecology is not consistent, where stage of training, environment, attitudes and emotions can all affect the outcome (Essers et al. 2011; Ginsburg et al. 2012). In line with the literature, diagnostic competence involves the ability to manage ambiguous problems, tolerate uncertainty and make decisions with limited information (Schon 1983). Our findings suggest there is a degree of resentment towards measuring outcomes through WBAs and “ticking boxes”, echoing others who have criticized competency based education as being politically motivated, impersonal or rigid (Carraccio et al. 2002; Leung 2002). Hence, the reductionist approach to assessment has risked measuring what is measurable but not what is important (Snadden 1999).

The step-wise, logical transition from novice to expert in Miller’s pyramid (Miller 1990) does seem to oversimplify how diagnostic competence is achieved. For instance, the data supporting our conceptual model reflects that there are cycles of reflection and learning which happen throughout training, where trainees may need to revisit and augment competencies. Secondly, the levels in Miller’s pyramid suggest a trainee can move up and down these levels freely. However, the ability of trainees to do certain things is also controlled by a consultant delegating work when they deem it appropriate for the trainee to move to the next level of the pyramid. Thirdly, at the top of Miller’s pyramid, being able to “do” something suggests one is an expert, yet in reality, our data show that consultants aren’t just interested in trainees making diagnoses unless they are also reflective practitioners, aware of their limitations, act professionally and refer appropriately. It is also not a “final destination” and must be maintained through one’s career.

Our findings are akin to the Cambridge model of competency which shows how numerous other factors can influence performance (Rethans et al. 2002). The Cambridge model indicates that system based influences can affect performance such as guidelines and regulatory bodies, which have also appeared in our data. Personal relationships are also relevant and relate to aspects of professionalism which are central to our model of diagnostic competency. Therefore, those studies identified in the literature review which only examined accuracy of diagnoses (Cross et al. 2000; Cross, Dennis, Start 2002; Parham et al*.* 2010) would only be examining one aspect of becoming a sound diagnostic pathologist. For example, these studies compared trainee and consultant diagnoses as a means to determine accuracy and development of competence in trainees over time. However, getting the diagnosis correct is only a small part of being competent. Our research has shown that the underpinning processes which support getting to the correct diagnosis are actually more valued by consultants than the diagnosis itself. Furthermore, aspects of reflection and professionalism also impact on this process.

The importance of trusting trainees, especially those trainees near the end of their training appeared to be an extremely important aspect of diagnostic competence. Currently, consultants appear to want to review “all the evidence” and spend time with trainees before they feel completely happy for them to report independently. The “time” spent with a trainee to assess competence and delegation of tasks is in line with the work of Dijksterhuis and colleagues who concluded that the depth of acquaintance with a trainee is the most important factor affecting when to delegate work in post-graduate training (2009). Our research suggests a role for EPAs in the assessment of diagnostic competence, but the practicalities of determining what and when certain activities can be delegated to trainees requires further work. There is also the wider issue of working with the current ”risk adverse” culture within the medical profession and creating detailed guidelines to help inform exactly how independent reporting and similar forms of delegation can be put into practice safely. For example, despite two studies citing the possible positive contribution that increased responsibility can bring, these studies also concluded that trainees are currently rarely exposed to it (Pascal 1993; Davey et al. 1996). This issue has not been resolved and has been echoed more recently by Allen (2013).

The response rate for the questionnaire and the fact that few people responded to the invite for the Delphi study means that the reliability of our findings could be questioned. For instance, the questionnaire may have captured biased opinions given the small sample who responded. However, the findings from the subsequent qualitative study further highlighted that there are multiple issues with the current assessment strategies, suggesting the conclusions drawn from the questionnaire are accurate. This is further supported by the Delphi study as the majority of items reached “consensus in”. It could be argued that this Delphi study only asked participants to rate and rank items in terms of their importance, so only measured what consultants valued when determining diagnostic competence. While this is very important and helps to support the validity of our findings, validity, nonetheless, is just one part of the utility of an assessment, where reliability, educational value, acceptability and cost must also be considered (van der Vleuten and Schuwirth 2005). Further research is needed to determine how this body of work would translate into useful assessment strategies and tools. For example, just because an item in the Delphi reached “consensus in”, it might not be easy to measure!

Furthermore, development of this work requires a close working relationship with RCPath to ensure the research conclusions and recommendations are interpreted appropriately. To this end, the RCPath are intending to update their entire training curricula by the end of 2020. Until this process has begun, it is not apposite to design a new assessment tool as 1) the assessment tool must meet the objectives of the new curriculum and the curriculum must be in line with what the assessment tests (constructive alignment) 2) the findings from this research may affect the design of the curriculum and not just the assessment strategies and 3) the curriculum is likely to be wider than just recognising diagnostic competence. Therefore, thorough dialogue is required between the relevant stakeholders to utilize this research most effectively. It is only then that the “utility” of any new assessment tools can be addressed.

It is important to stress that participants were asked for their views on diagnostic competence. As stated in the literature review, competence is somewhat different to “performance” which takes into account how one performs in less than ideal situations and the ability of the individual to learn and adapt to his or her environment. Nonetheless, the research participants amalgamated both aspects of competency and performance when discussing diagnostic competency, such that it appears that the terms competence and performance are inter-changeable in the pathology workplace. Furthermore, some participants incorporated many different elements into their judgement of diagnostic competence, whereas others were more careful to suggest only those that were relevant to “diagnosis”. This reflects the nature of assessment in medical education, how people interpret competency in different ways and that there are differing expectations of what it means (Eraut 1998).In reality, the interviewees may have been describing what makes a “sound pathologist” because diagnostic competence is a function of the trainee as a whole. This finding links to aspects of the literature review which reported how subjectivity can affect how judges rate individuals because they have poor insight and consistency into how they make and apply judgements to scenarios (Unsworth, Harries and Davies 2015).

Finally, Cooksey’s framework guided the direction of this study and allowed the research objectives to be achieved. However, it was not possible to follow the classical application of this theory by using case vignettes and regression analysis to determine the importance of cues. In that regard, the Delphi study that was utilized may not completely be in line with the classical application of the theory, yet it was a more pragmatic approach to the problem and triangulated the findings and determined the importance of the cues.

# **Chapter 7**

# **Conclusions and recommendations**

**Chapter 7: Conclusions, recommendations and further work**

## 7.1 Summary

In short, the purpose of this research was to understand and explore how consultants determine diagnostic competence in trainees and if the current WBAs are perceived to measure it. These objectives were achieved. The current WBAs appear to have multiple issues, but are not perceived to measure diagnostic competence. The outcomes from other questions within the questionnaire and written feedback suggested this might be because the WBA forms do not reflect the knowledge, skills or behaviours most relevant to diagnostic competence or are applied in the appropriate context. This is reflected in our qualitative interview data which shows diagnostic competence to be a complex entity, requiring trainees to possess both “process” and “person” characteristics which were modified by stage of training and developed over time to develop trust between the trainee and consultant. These qualities are not fully described in the current WBAs and the longitudinal nature of how consultants determine diagnostic competence was at odds with the current WBA utilizing single cases to determine competency. Consistency of both “process” and “person” was more important to consultants and led to accurate diagnoses based on sound pathological approach and an awareness of one’s own diagnostic ability. This study is novel in that it is the first to report the key components which consultants look for when determining diagnostic competence. The resulting conceptual model provides an overview of these parts and the importance of context when determining competence, which is particularly relevant to how future assessment tools should be utilized. The finding that diagnostic competence should be assessed longitudinally is of particular importance and has implications for when the curriculum is re-written for specialist training in histopathology. Secondly, further research could be undertaken to see if a longitudinal approach to assessing competence is appropriate for other specialties in healthcare.

## 7.2 Conclusions for practice

Changing the current assessment strategies based on the findings of this study should:

- ensure the qualities underpinning diagnostic competence are clearly described in assessment tools

-ensure the method of delivery of these assessments is in line with the judgement ecology we have described. In particular, competency should be measured over time and not via single encounters examining one-off, individual competencies.

-demystify diagnostic competence, making it more explicit to trainees and consultants so that expectations and the ability to achieve competency become transparent to all parties involved

## 7.3 Conclusions for research

This research has addressed a significant void in the medical literature. Firstly, by reporting the current WBAs are not perceived to measure diagnostic competence and suffer from many other perceived issues. Secondly, the key qualities that consultant look for when determining diagnostic competence in trainees, as well as the context, were described through qualitative interviews and triangulated with a Delphi study. The resulting conceptual model suggests that the current methods for measuring diagnostic competence are not assessing all that matters nor assessing it in the right way, suggesting modifications to the format and delivery of assessment tools in histopathology are required.

## 7.4 Recommendations for practice

The findings from this research alongside collaboration with the RCPath could be used to:

-modify the current WBA tools

-inform future assessment strategies in histopathology

-inform the new pathology curriculum which is due to be re-written in the next few years

## 7.5 Recommendations for research

The conceptual model could be used to develop assessment strategies for diagnostic histopathology:

-in the UK

-in other countries e.g. outside the UK

-in other specialties

The approach and methodology used in this thesis could be used to determine competence in:

-other medical or dental specialties

-undergraduate medical and dental education

## 7.6 Further work

A new assessment form should be designed to measure diagnostic competence in histopathology trainees. The assessment should be longitudinal in nature and administered over placements or blocks of cases. The reliability, validity and feasibility of this tool will need testing.

# **Appendices**

## Appendix 1: Questionnaire invitation emails

Dear assessor,

There is a scarcity of evidence in the literature regarding the experience of using workplace-based assessments in histopathology. The histopathology environment is quite different from many of the surgical specialties so it is important to determine if the current workplace-based assessments are appropriate.

The link below will take you to a short online questionnaire that consists of ten questions. It should take only a few minutes to complete. The questionnaire will not collect any of your personal information and is completely voluntary.

**Please complete the questionnaire only if you complete workplace-based assessments for trainees**

<https://www.surveymonkey.co.uk/r/257MXBY>

Your views are very important and may help to shape future assessment tools in histopathology.

If you have any questions, please email these to [d.j.brierley@sheffield.ac.uk](mailto:d.j.brierley@sheffield.ac.uk)

Many thanks,

The Royal College of Pathologists

Dear trainee,

There is a scarcity of evidence in the literature regarding the experience of using workplace-based assessments in histopathology. The histopathology environment is quite different from many of the surgical specialties so it is important to determine if the current workplace-based assessments are appropriate.

The link below will take you to a short online questionnaire that consists of ten questions. It should take only a few minutes to complete. The questionnaire will not collect any of your personal information and is completely voluntary.

<https://www.surveymonkey.co.uk/r/26RYXMC>

Your views are very important and may help to shape future assessment tools in histopathology.

If you have any questions, please email these to [d.j.brierley@sheffield.ac.uk](mailto:d.j.brierley@sheffield.ac.uk)

Many thanks,

The Royal College of Pathologists

## Appendix 2: Participant Information Sheet

You have been given this information sheet as we would like you to participate a questionnaire.

**1. Research Project Title**

Determination of diagnostic competence in histopathology trainees.

**2. Invitation paragraph**

You are being invited to take part in a research project. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. Thank you for reading this.

**3. What is the project’s purpose?**

There is a scarcity of evidence in the literature detailing how consultants determine trainee diagnostic competency in histopathology. Workplace-based assessments are often used in part to document trainee diagnostic competency but the full judgement ecology is not completely understood.

This questionnaire will gather data on the general experience of using workplace-based assessments in histopathology with assessors and trainees.

This project is part of a PhD project by Daniel Brierley (StR in Oral and Maxillofacial Pathology).

**4. Why have I been chosen?**

You have been chosen because you are a histopathology assessor or trainee and have experience of using workplace based assessments.

**5. Do I have to take part?**

It is up to you to decide whether or not you would like to take part or not. If you do decide to take part you can still withdraw at any time without it affecting any benefits that you are entitled to in any way. You do not have to give a reason.

**6. What will happen to me if I take part?**

The information from the questionnaires will be anonymous and stored/used for 10 years. The questionnaire will be conducted via an online link and may be completed by the participants in their own time.

The questionnaire results will be analysed to explore the opinions, experiences and attitudes of the participants.

**7. What do I have to do?**

The questionnaire can be completed online and submitted when the participant has answered each question. There is no time limit.

**8. What are the possible disadvantages and risks of taking part?**

There are no foreseen disadvantages or risks of taking part in this study. Participants will, however, have to dedicate roughly 20 minutes of their time to complete the questionnaire.

**9. What are the possible benefits of taking part?**

Whilst there are no immediate benefits for those people participating in the project, it is hoped that this work will help to understand how consultants determine trainee diagnostic competency. This may help with the training of future trainees and determine if current training programmes adequately assess diagnostic competence.

**10. What happens if the research study stops earlier than expected?**

It is not anticipated that the research study will be stopped earlier than expected, but if it is, participants will be informed of the reason for this. Any data collected to this point will be destroyed.

**11. What if something goes wrong?**

If any participant would like to make a complaint about the study, participants should contact the researcher (Daniel Brierley) or the lead supervisor (Professor Paula Farthing). If you feel like your complaint has not been handled to your satisfaction, you may contact the Head of the Academic Unit of Oral and Maxillofacial Medicine, Pathology and Surgery, Professor Keith Hunter, who will escalate the complaint through the appropriate channels. Contact numbers are given on item 17 (see below).

**12. Will my taking part in this project be kept confidential?**

You are not required to provide any personal information before completing the questionnaire. You will not be able to be identified in any reports or publications.

Data will be stored on a password protected computer or USB device.

**13. What type of information will be sought from me and why is the collection of this information relevant for achieving the research project’s objectives?**

The type of information which is sought from you is regarding your opinions relating to your experience of workplace-based assessments in histopathology. This information is relevant so we can determine the impact of workplace-based assessments and if they are a good measure of competence. This may lead to changes to trainee assessment in the future.

**14. What will happen to the results of the research project?**

The results of this research may be published in histopathology or medical education journals. However, participants will not be identified in the report or publication. If results are published, participants will be informed and told where they may access the published material. The results may also be presented at conferences and due to the nature of this research it is very likely that other researchers may find the data collected useful in answering future research questions. We will ask for your explicit consent for your data to be shared in this way and if you agree, we will ensure that the data collected about you is untraceable back to you before allowing others to use it.

**15. Who is organising and funding the research?**

The research is funded through Daniel Brierley’s training as part of his 8 year programme as a specialist trainee in oral and maxillofacial pathology. Additional funding may be requested from other societies or institutions.

**16. Who has ethically reviewed the project?**

This project has been ethically approved via the Clinical Dentistry ethics review procedure, University of Sheffield

**17. Contacts for further information**

Researcher

Daniel Brierley

Room E36

Academic Unit of Oral and Maxillofacial Medicine, Pathology and Surgery

The School of Clinical Dentistry

19 Claremont Crescent

Sheffield

S10 2TA

Tel: 01142265477

Email: d.j.brierley@sheffield.ac.uk

Lead supervisor

Emeritus Professor Paula Farthing

Academic Unit of Oral and Maxillofacial Medicine, Pathology and Surgery

The School of Clinical Dentistry

19 Claremont Crescent

Sheffield

S10 2TA

Tel: 0114 2717969

Email: p.farthing@sheffield.ac.uk

Head of Department

Professor Keith Hunter

Room E32

Academic Unit of Oral and Maxillofacial Medicine, Pathology and Surgery

The School of Clinical Dentistry

19 Claremont Crescent

Sheffield

S10 2TA

Tel: 0114 2717960

Email: k.hunter@sheffield.ac.uk

Each participant will be given a copy of this information sheet and consent form to keep.

We would like to thank all participants for taking part in this study

**Appendix 3: Interview invitation email**

Dear colleagues,

I am currently undertaking a PhD as part of my histopathology training. I completed my year in general pathology two years ago, so I have met a number of you already.

I am interested in:

-How consultants define diagnostic competence

-Whether you think workplace-based assessments help with determining trainee diagnostic competence

-And what cues you use when determining if a trainee is diagnostically competent or not.

I would like to explore these ideas by interviewing consultants in histopathology.

The interviews are likely to last about an hour, depending on the amount of discussion generated. Ethical approval from the University of Sheffield has been granted and all data collected will be strictly confidential and only accessible to members of the research team.

I have attached a participant information sheet to this email, which gives further information about the study.

If you would be happy to be interviewed, please could you let me know by email: [Daniel.brierley@sth.nhs.uk](mailto:Daniel.brierley@sth.nhs.uk)

If you would like to know further information about the study, or you have any questions, please don’t hesitate to contact me on the above email address.

Best wishes

Daniel Brierley

## Appendix 4: Information sheet for qualitative interviews

**Participant Information Sheet for interview**

You have been given this information sheet as we would like you to participate in an interview

**1. Research Project Title**

Determination of diagnostic competence in histopathology trainees

**2. Invitation paragraph**

You are being invited to take part in a research project. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. Thank you for reading this.

**3. What is the project’s purpose?**

There is a scarcity of evidence in the literature detailing how consultants determine trainee diagnostic competency in histopathology. Workplace-based assessments are often used in part to document trainee diagnostic competency but the full judgement ecology is not completely understood. An initial questionnaire has gathered data on the general experience of using workplace-based assessments in histopathology with consultants and trainees.

Following this, we would like to conduct interviews to explore what cues consultants use when making a judgement of trainee diagnostic competency. As part of this, you may be asked to define diagnostic competency and your experience of using workplace-based assessments in making this judgement.

These projects are part of a PhD project by Daniel Brierley (StR in Oral and Maxillofacial Pathology).

**4. Why have I been chosen?**

You have been chosen because you are a histopathology consultant involved with training of histopathology trainees and use workplace-based assessments. The anticipated number of participants that will be interviewed will be dictated by the point at which no new information would come to light by interviewing further participants (approximately 10-20).

**5. Do I have to take part?**

It is up to you to decide whether or not you would like to take part or not. If you do decide to take part you will be given this information sheet to keep (and be asked to sign a consent form) and you can still withdraw at any time without it affecting any benefits that you are entitled to in any way. You do not have to give a reason.

**6. What will happen to me if I take part?**

Interviews with each participant may take about an hour to complete depending on the discussion that is generated.

The information from the interviews be anonymous and stored/used for 10 years. As the interviewer will visit each participant for the interviews, travel expenses are not available.

The interview will be conducted in a comfortable environment and audio recorded. The questioning in the interviews will be mostly open-ended and participants are encouraged to speak freely and openly. Sometimes the interviewer may ask more specific questions if this is necessary.

The recordings from the interviews will be transcribed and analysed to explore the opinions, experiences and attitudes of the participants.

**7. What do I have to do?**

For the interview, you need only turn up to your interview and communicate with the interviewer. No prior preparation is required. You do not need to bring anything with you.

**8. What are the possible disadvantages and risks of taking part?**

There are no foreseen disadvantages or risks of taking part in this study. Participants will, however, have to dedicate about an hour of their time to do the interview.

**9. What are the possible benefits of taking part?**

Whilst there are no immediate benefits for those people participating in the project, it is hoped that this work will help to understand how consultants determine trainee diagnostic competency. This may help with the training of future trainees and determine if current training programmes adequately assess diagnostic competence.

**10. What happens if the research study stops earlier than expected?**

It is not anticipated that the research study will be stopped earlier than expected, but if it is, participants will be informed of the reason for this. Any data collected to this point will be destroyed.

**11. What if something goes wrong?**

If any participant would like to make a complaint about the study, participants should contact the researcher (Daniel Brierley) or the lead supervisor (Professor Paula Farthing). If you feel like your complaint has not been handled to your satisfaction, you may contact the Head of the Academic Unit of Oral and Maxillofacial Medicine, Pathology and Surgery, Professor Keith Hunter, who will escalate the complaint through the appropriate channels. Contact numbers are given on item 18 (see below).

**12. Will my taking part in this project be kept confidential?**

All the information that we collect about you during the course of the research will be kept strictly confidential. You will not be able to be identified in any reports or publications.

Each participant will be given a reference number and not be identified by name. The reference numbers will only be known by the researcher (Daniel Brierley). All data will be anonymised (all identifying information will be removed from all presentations of the data).Data will be stored on a password protected computer or USB device.

**13. What type of information will be sought from me and why is the collection of this information relevant for achieving the research project’s objectives?**

The type of information which is sought from you is regarding your opinions relating to your experience of workplace-based assessments in histopathology. We are also interested in your opinion of competence in histopathology and how this can be measured. This information is relevant so we can determine the impact of workplace-based assessments and if they are a good measure of competence. This may lead to changes to trainee assessment in the future.

**14. What will happen to the results of the research project?**

The results of this research may be published in histopathology or medical education journals. However, participants will not be identified in the report or publication. If results are published, participants will be informed and told where they may access the published material. The results may also be presented at conferences and due to the nature of this research it is very likely that other researchers may find the data collected useful in answering future research questions. We will ask for your explicit consent for your data to be shared in this way and if you agree, we will ensure that the data collected about you is untraceable back to you before allowing others to use it.

**15. Who is organising and funding the research?**

The research is funded through Daniel Brierley’s training as part of his 8 year programme as a specialist trainee in oral and maxillofacial pathology. Additional funding may be requested from other societies or institutions.

**16. Who has ethically reviewed the project?**

This project has been ethically approved via the Clinical Dentistry ethics review procedure, University of Sheffield

**17**. **Will I be recorded and how will the recorded media be used?**

Yes. The recordings will be transcribed by the researcher (Daniel Brierley). The information collected during this research will be used only for analysis and for illustration in conferences, presentations and lectures. No other use will be made of them without your written permission, and no one outside the project will be allowed access to the original recordings. No other use will be made of them without your written permission, and no one outside the project will be allowed access to the original recordings. Data will be stored for 10 years and will be anonymised (all identifying information will be removed from all presentations of the data).

**18. Contacts for further information**

Researcher

Daniel Brierley

Room E36

Academic Unit of Oral and Maxillofacial Medicine, Pathology and Surgery

The School of Clinical Dentistry

19 Claremont Crescent

Sheffield

S10 2TA

Tel: 01142265477

Email: d.j.brierley@sheffield.ac.uk

Lead supervisor

Emeritus Professor Paula Farthing

Academic Unit of Oral and Maxillofacial Medicine, Pathology and Surgery

The School of Clinical Dentistry

19 Claremont Crescent

Sheffield

S10 2TA

Tel: 0114 2717969

Email: p.farthing@sheffield.ac.uk

Head of Department

Professor Keith Hunter

Room E32

Academic Unit of Oral and Maxillofacial Medicine, Pathology and Surgery

The School of Clinical Dentistry

19 Claremont Crescent

Sheffield

S10 2TA

Tel: 0114 2717960

Email: k.hunter@sheffield.ac.uk

Each participant will be given a copy of this information sheet and consent form to keep.

We would like to thank all participants for taking part in this study

## Appendix 5: Interview guide

**“Determination of diagnostic competence in histopathology trainees”**

Dr Daniel J Brierley (Researcher – Specialist Registrar, Academic Unit of Oral and Maxillofacial Medicine, Pathology and Surgery)

Professor Paula M Farthing (Supervisor and Principal Investigator**,** Emeritus Professor in Oral and Maxillofacial Pathology, Academic Unit of Oral and Maxillofacial Medicine, Pathology and Surgery)

Dr Sandra Zijlstra-Shaw (Co-supervisor and Senior Clinical Teacher, Deputy Director of Learning and Teaching (Professionalism and Social Accountability) Academic Unit of Primary Dental Care

**1. Introduction**

**Research objectives:**

To explore and understand:

1. What diagnostic competence encompasses in histopathology
2. The cues consultants look for in their trainees when determining diagnostic competence
3. The judgement ecology which pertains to determining diagnostic competency in trainees

1.1 Introduction and information sheet

The researcher introduces himself and states the aim of the project which is to explore the judgement of diagnostic competency in histopathology. He then ensures that the participant has a copy of the information sheet and understands the scope of the research.

1.2 Consent

The researcher then ensures that the ethics protocol is followed and that consent has been given and the consent form signed.

The researcher then reminds the participant that they can withdraw at any time and confirms the permission to record the interview.

**2. Background**

2.1 Participant identification

The participant is identified using a pre-arranged code and the date and time of recording is noted.

2.2 Participants position with respect to histopathology

Background in histopathology

Length of involvement in histopathology and trainee supervisory roles within histopathology

**3. What is diagnostic competence in histopathology?**

How would you define diagnostic competence?

How does a trainee show they are diagnostically competent?

How do you currently determine trainee diagnostic competence?

**4. What cues are used by consultants to measure diagnostic competence?**

What qualities do you look for in trainees and why?

Are some qualities more important than others? If so, why?

**5. The judgement ecology which pertains to diagnostic competence**

Do you think that current WBAs are a valid tool to measure diagnostic competence? If so, why? If not, why not?

Do you know of a better way to measure trainee diagnostic competence? If so, what?

How difficult/easy are these judgement to make?

What else do you take into account and why?

**6. Further material**

Are there any further aspects which you feel should have been covered?

**7. Closure**

The researcher thanks the participant for their time and help and explains that they may have a copy of the report on the construct when it becomes available. He then ascertains if the participant would like a copy of this report and confirms the details of where this can be sent.

Note: Questions were modified for trainees who were interviewed.

## Appendix 6: Consent form for interview

#### Consent Form

|  |
| --- |
| **Title of Research Project**: Determination of diagnostic competence in histopathology trainees  **Name of Researcher**: Daniel Brierley  Participant Identification Number for this project: Please initial box   1. I confirm that I have read and understand the information sheet/letter (delete as applicable) dated *[ ]* explaining the above research   project and I have had the opportunity to ask questions about the project.   1. I understand that my participation is voluntary and that I am free to   withdraw at any time without giving any reason and without there being  any negative consequences. In addition, should I not wish to answer any particular  question or questions, I am free to decline. *Lead researcher contact telephone*  *number: 0114 2265477*   1. I understand that my responses will be kept strictly confidential. I give permission for members of the research team to have access   to my anonymised responses. I understand that my name will not be  linked with the research materials, and I will not be identified or identifiable in the  report or reports that result from the research.   1. I agree for the data collected from me to be used in future research.   Recordings will be stored securely for 10 years.  The recorded media will be used to determine what cues are relevant to the judgement of diagnostic competence in histopathology from the perspective of Social Judgement Theory   1. I agree to take part in the above research project.   \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  Name of Participant Date Signature  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  Name of person taking consent Date Signature  (*if different from lead researcher*)  *To be signed and dated in presence of the participant*  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  Lead Researcher Date Signature  *To be signed and dated in presence of the participant*  Copies:  *Once this has been signed by all parties the participant should receive a copy of the signed and dated participant consent form, the letter/pre-written script/information sheet and any other written information provided to the participants. A copy of the signed and dated consent form should be placed in the project’s main record (e.g. a site file), which must be kept in a secure location.* |
|  |

## Appendix 7: Invitation email for Delphi study

**A Delphi Study exploring diagnostic competency in histopathology**

I am currently undertaking a PhD exploring what diagnostic competence incorporates in histopathology. I am particularly interested in what trainers look for in their trainees when determining diagnostic competency and the judgement ecology which supports these decisions.

I have undertaken qualitative interviews to investigate this, but wish to perform a Delphi study to understand this further. Specifically, I would like at least 12 consultants (who are training programme directors or have significant experience of training) to help refine the items of competency which my research has identified.

I hope that understanding these concepts will help provide an evidence base which could be used to inform future assessments in histopathology, for the benefit of trainees and the profession.

I would be extremely grateful if you could take part and anticipate each round will take approximately 25 minutes to complete. There will up to three rounds. I anticipate starting the Delphi by the end of January/beginning of February 2019. More details are available on the attached information sheet. If you wish to take part, I will also send you a consent form.

The results from this study will be shared with the Royal College of Pathologists who have helped circulate this email.

It is important to note that those who take part in the Delphi will only be known to me and won’t be identifiable to the other members taking part or anyone external to the study. The study has ethical approval from the University of Sheffield.

If you are happy to be part of the Delphi Study, I would be grateful if you would email me on [d.j.brierley@sheffield.ac.uk](mailto:d.j.brierley@sheffield.ac.uk). To ensure a representative panel, I would be grateful if you could provide the following information when you contact me. This will not be linked to the information generated by the Delphi study.

**Name:**

**Gender:**

**Educational supervisor:** (Yes/No). If yes, for how many years?

**Training Programme Director:** (Yes/No) If yes, for how many years?

**Other deanery or College roles (other than roles above)?**

**Number of years you have been a trainer:**

Many thanks,

Dr Daniel Brierley (Consultant in Oral and Maxillofacial Pathology, Sheffield)

**Appendix 8: Participant Information Sheet for Delphi study**

You have been given this information sheet as we would like you to participate in a type of study called a Delphi.

**1. Research Project Title**

Determination of diagnostic competence in histopathology trainees

**2. Invitation paragraph**

You are being invited to take part in a research project. Before you decide to take part it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. Thank you for reading this.

**3. What is the project’s purpose?**

There is a scarcity of evidence in the literature detailing how consultants determine trainee diagnostic competency in histopathology. Workplace-based assessments are often used in part to document trainee diagnostic competency but the full judgement ecology is not completely understood. An initial questionnaire has gathered data on the general experience of using workplace-based assessments in histopathology with consultants and trainees.

Following this, interviews have been conducted to explore what cues consultants use when making a judgement of trainee diagnostic competency. As part of this, the consultants were asked to talk about how they judge diagnostic competency and the process this entails.

The final part of this project involves using a Delphi study. A Delphi study involves using a panel of experts in a given field to come to an agreement on a particular topic or set of items. The participants are presented with information on the topic and asked to determine, individually, how they rate each item (usually using a Likert scale or similar scoring method). Participants are also encouraged to provide feedback on items so they can be refined in future rounds of the Delphi. The researcher collates the responses to the items and is able to see which items participants rated highly (or agreed on) and which ones did not reach agreement and need refinement. In subsequent rounds, the participants are provided with data from the previous round where they are able to continue to refine and reach agreement on the remaining items. Each participant’s responses are anonymised so each participant is able to express their views and is not coerced into agreeing with a dominant view.

Specifically, the research team would like to use the Delphi technique to further understand the importance of the cues of diagnostic competency we have identified from the previous qualitative interviews

These projects are part of a PhD project by Daniel Brierley (Consultant and Clinical Teacher in Oral and Maxillofacial Pathology).

**4. Why have I been chosen?**

You have been chosen because you are an experienced histopathology consultant (or trainee) involved with the training of histopathology trainees and using workplace-based assessments. The anticipated number of participants that this Delphi will use is approximately 10.

**5. Do I have to take part?**

It is up to you to decide whether or not you would like to take part or not. If you do decide to take part you will be given this information sheet to keep (and be asked to sign a consent form) and you can still withdraw at any time without it affecting any benefits that you are entitled to in any way. You do not have to give a reason.

**6. What will happen to me if I take part?**

The Delphi will comprise 2-3 rounds.

In each round, you will be asked to express your agreement on the information provided. Each round could take between 15-25 minutes to complete.

The information from the Delphi will be anonymous and stored/used for 10 years. The Delphi will be conducted via email (with an attachment) and may be completed by the participants in their own time. However, we would encourage each round of the Delphi to be returned within a 2-3 week period.

The information taken from the Delphi may be used to help design future assessment tools in histopathology.

**7. What do I have to do?**

For the Delphi, you have to examine the information presented to you. This is likely to be a list of statement which we would like you to judge. No prior preparation is required. In subsequent rounds you will be able to see anonymised responses from the other participants and have an opportunity to express your judgements again in light of the new information.

**8. What are the possible disadvantages and risks of taking part?**

There are no foreseen disadvantages or risks of taking part in this study. Participants will, however, have to dedicate some of their time to do the Delphi.

**9. What are the possible benefits of taking part?**

Whilst there are no immediate benefits for those people participating in the project, it is hoped that this work may form an evidence base to help design assessment tools in histopathology. This may help with the training of future trainees.

**10. What happens if the research study stops earlier than expected?**

It is not anticipated that the research study will be stopped earlier than expected, but if it is, participants will be informed of the reason for this. Any data collected to this point will be destroyed.

**11. What if something goes wrong?**

If any participant would like to make a complaint about the study, participants should contact the researcher (Daniel Brierley) or the lead supervisor (Professor Paula Farthing). If you feel your complaint has not been handled to your satisfaction, you may contact the Head of the Academic Unit of Oral and Maxillofacial Medicine, Pathology and Surgery, Professor Keith Hunter, who will escalate the complaint through the appropriate channels. Contact numbers are given on item 18 (see below).

**12. Will my taking part in this project be kept confidential?**

All the information that we collect about you during the course of the research will be kept strictly confidential. You will not be able to be identified in any reports or publications.

Each participant will be given a reference number and not be identified by name. The reference numbers will only be known by the researcher (Daniel Brierley). All data will be anonymised (all identifying information will be removed from all presentations of the data).Data will be stored on a password protected computer or USB device.

However, please see item 14.

**13. What type of information will be sought from me and why is the collection of this information relevant for achieving the research project’s objectives?**

The type of information which is sought from you is regarding your opinion on diagnostic competency and the cues or characteristics which consultants use when judging it in their trainees. This may help to inform changes to trainee assessments in the future.

**14. What will happen to the results of the research project?**

The results of this research may be published in histopathology or medical education journals. However, participants will not be identified in the report or publication. If results are published, participants will be informed and told where they may access the published material. The results may also be presented at conferences.

The data will also be shared with external bodies such as the Royal College of Pathologists and GMC/GDC. We will ensure that the data collected about you is untraceable back to you before allowing others to use it.

**15. Who is organising and funding the research?**

The research study has not required funding to date. However, the project is being conducted by Daniel Brierley in his role as a clinical teacher for the School of Clinical Dentistry, Sheffield.

**16. Who has ethically reviewed the project?**

This project has been ethically approved via the Clinical Dentistry ethics review procedure, University of Sheffield

**18. Contacts for further information**

Researcher

Daniel Brierley

Room E06

Academic Unit of Oral and Maxillofacial Medicine, Pathology and Surgery

The School of Clinical Dentistry

19 Claremont Crescent

Sheffield

S10 2TA

Tel: 0114 2159346

Email: d.j.brierley@sheffield.ac.uk

Lead supervisor

Emeritus Professor Paula Farthing

Academic Unit of Oral and Maxillofacial Medicine, Pathology and Surgery

The School of Clinical Dentistry

19 Claremont Crescent

Sheffield

S10 2TA

Tel: 0114 2717969

Email: p.farthing@sheffield.ac.uk

Head of Department

Professor Keith Hunter

Room E32

Academic Unit of Oral and Maxillofacial Medicine, Pathology and Surgery

The School of Clinical Dentistry

19 Claremont Crescent

Sheffield

S10 2TA

Tel: 0114 2717960

Email: k.hunter@sheffield.ac.uk

Each participant will be given a copy of this information sheet to keep. We would like to thank all participants for taking part in this study

## Appendices 9 and 10: Instructions for Delphi and consent form

**“Determination of diagnostic competence in histopathology trainees”**

**Delphi Study**

**Instructions for round 1**

You have been invited to take part in a Delphi study as part of a research project, the aim of which is to understand what constitutes diagnostic competency in histopathology trainees.

A Delphi study involves using a panel of experts in a given field to come to an agreement on a particular topic or set of items. The participants are presented with information on the topic and asked to determine, individually, how they rate each item (usually using a Likert-type scale or similar scoring method). Participants are also encouraged to provide feedback on items which can then be refined in future rounds of the Delphi. The researcher collates the responses to the items and is able to see which items are rated highly (or agreed on) by participants, and which ones did not reach agreement and need refinement. In subsequent rounds, the participants are provided with data from the previous round so they are able to continue to refine and reach agreement on the remaining items. Further information can be found on the attached information sheet you have previously been provided.

As part of this study, we should be grateful if you would:

-complete the background information about yourself if you have not already provided this (Part1). The background information helps to ensure that a representative Delphi panel has been selected.

-sign the consent form (Part 2). Please note that the consent form signature will not be linked to the results

-rate the items against a scale (Part 3)

-rank the three items which you think are the “most important” and “least important” (Part 4)

**Part 1: Background information**

Please would you provide:

Your gender:

The number of years you have been an educational supervisor (if appropriate):

The number of years you have been a Training Programme Director (TPD) (if appropriate):

The number of years you have been a trainer:

**Part 2: Consent**

#### Consent Form

|  |
| --- |
| **Title of Research Project**: Determination of diagnostic competence in histopathology trainees  **Name of Researcher**: Daniel Brierley  Please initial or place an “X” in the box  I confirm that I have read and understand the information sheet explaining  the above research project and I have had the opportunity to ask questions  about the project.  I understand that my participation is voluntary and that I am free to withdraw  at any time without giving any reason and without any negative  consequences. In addition, should I not wish to answer any particular question  or questions, I am free to decline.  I understand that my responses will be kept strictly confidential. I give permission  for members of the research team to have access to my anonymised responses.  I understand that my name will not be linked with the research materials, and I will  not be identified or identifiable in the report or reports that result from the research.  I agree for the data collected from me to be used in future research and for it to be  shared with external bodies such as the Royal College of Pathologists. Data will be  stored securely for 10 years.  I agree to take part in the above research project.  \_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_  Name of Participant Date Signature |

**Part 3: Rating of items**

Our research has suggested that the following items/qualities may be important to diagnostic competence in histopathology. Please rate each item according to how important you think it is when determining a trainee’s diagnostic competency. The scale consists of 10 points, where “10” indicates that you think the item is definitely important and “1” indicates you think the item is definitely not important. Please only select one option for each item by placing a cross “x” in the relevant box. The items have been divided into 7 sections.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Definitely not important Definitely important** | | | | | | | | | | | | | | |
| **Stage of training** | **1** | **2** | **3** | | **4** | **5** | **6** | | **7** | **8** | | | **9** | | **10** |
| 1. The stage of training e.g. ST1 versus ST5 |  |  |  | |  |  |  | |  |  | | |  | |  |
| 2. RCPath curricula and training guidelines |  |  |  | |  |  |  | |  |  | | |  | |  |
|  | **Definitely not important Definitely important** | | | | | | | | | | | | | | | |
| **Timing** | **1** | **2** | **3** | **4** | | **5** | **6** | **7** | | | **8** | **9** | | **10** | | |
| 3. Placements/cases assessed longitudinally over a period of time |  |  |  |  | |  |  |  | | |  |  | |  | | |
| 4. Individual cases in an examination setting |  |  |  |  | |  |  |  | | |  |  | |  | | |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Definitely not important Definitely important** | | | | | | | | | |
| **Professionalism** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
| 5. Trainee is organised and timely when conducting themselves in the department e.g. does not lose cases, is aware of turnaround times, triages urgent cases |  |  |  |  |  |  |  |  |  |  |
| 6. Trainee communicates with all staff appropriately e.g. effectively, timely, politely |  |  |  |  |  |  |  |  |  |  |
| 7. Trainee works as part of a team e.g. works effectively with colleagues and does not create conflict |  |  |  |  |  |  |  |  |  |  |
| 8. Trainee is motivated and has a good attitude |  |  |  |  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Definitely not important Definitely important** | | | | | | | | | |
| **Forming a diagnosis** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
| 9. The diagnosis is correct |  |  |  |  |  |  |  |  |  |  |
| 10. Trainee consistently produces a correct diagnosis |  |  |  |  |  |  |  |  |  |  |
| 11. The histopathology report is accurate & does not contain factual errors or omit important information |  |  |  |  |  |  |  |  |  |  |
| 12. Trainee shows evidence of clinico-pathological correlation: clinical information/macroscopic work is used to support/refute possible diagnoses/features. Examples might be correlating radiographic with histological findings; referring to previous pathology reports that are relevant to the current biopsy; recognising that clinical information provided may affect what is seen histologically, e.g. previous radiotherapy |  |  |  |  |  |  |  |  |  |  |
| 13. Trainee follows sound pathological principles to reach a diagnosis e.g. the trainee understands how they reached the diagnosis, has a consistent, accurate way of examining slides and knows the steps involved when making a diagnosis. The diagnosis has not stemmed from guessing. |  |  |  |  |  |  |  |  |  |  |
| 14. Trainee has appropriate level of basic knowledge. e.g. recognises normal histology for different organs/tissues, knows basic pathological processes, can compare normal to abnormal |  |  |  |  |  |  |  |  |  |  |
| 15. Trainee has appropriate level of up-to-date knowledge. e.g. latest papers/datasets |  |  |  |  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Definitely not important Definitely important** | | | | | | | | | |
| **Reflection** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
| 16. Trainee is aware of their own limitations |  |  |  |  |  |  |  |  |  |  |
| 17. Trainee reflects on their own limitations/performance |  |  |  |  |  |  |  |  |  |  |
| 18. Trainee shows evidence of improvement following reflection |  |  |  |  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Definitely not important Definitely important** | | | | | | | | | |
| **Trust** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
| 19. Trainee can be trusted to carry out macroscopic examination and “cut-up” independently |  |  |  |  |  |  |  |  |  |  |
| 20. Trainee can be trusted to report cases (however, the consultant will check the reports and authorise them) |  |  |  |  |  |  |  |  |  |  |
| 21. Trainee can be trusted to report independently (the consultant does not check the report and the trainee authorises it) |  |  |  |  |  |  |  |  |  |  |
| 22. Trainee asks for help/second opinion when needed |  |  |  |  |  |  |  |  |  |  |
| 23. Your opinion of a trainee’s diagnostic competence is related to an overall impression you have of them, rather than items of evidence e.g. gut-feelings/impressions/ your perceptions about trainees |  |  |  |  |  |  |  |  |  |  |
| 24. Your opinion of a trainee’s diagnostic competence is related to bringing all the evidence together and triangulating findings from exams, workplace-based assessments, day-to-day work |  |  |  |  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Definitely not important Definitely important** | | | | | | | | | |
| **The patient’s report** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
| 25. Report is useful to the clinician/surgeon e.g. it does not contain unnecessary information or detail; it shows the surgeon’s clinical differential diagnoses have been considered |  |  |  |  |  |  |  |  |  |  |
| 26. Report ensures the clinician/surgeon receives the appropriate message e.g. when diagnoses are not clear-cut and a differential diagnosis or other message needs to be conveyed. This might include a discussion about the limitations of the biopsy or if further tests or information are needed |  |  |  |  |  |  |  |  |  |  |
| 27. Report is organised appropriately and well-written e.g. it is logically laid out, is not confusing and uses appropriate language |  |  |  |  |  |  |  |  |  |  |

If you would like to comment on any of the items in the tables above, please write your suggestions in the text box below. In addition, if there are there any qualities/traits which you think are part of diagnostic competence which are not included in the table, please also write these in the same box.

**Part 4: Ranking of items**

Please list the 3 items from the tables which you feel are most important when judging diagnostic competence in trainees. The items are number 1-27. **Please put the most important item first.**

1.

2.

3.

Please list the 3 items from the tables which you feel are least important when judging diagnostic competence in trainees. **Please put the least important item first.**

1.

2.

3.

If you wish to explain your ranking, please write in the box below:

Thank you.

**Delphi Study**

**Instructions for round 2**

Thank you for submitting your responses to round 1. I have pleasure in enclosing the results from round 1 including:

-the original set of statements, some of which have now been amended following your feedback (modifications are in green)

-the score you gave each statement

-the mean score from the panel

-the range of scores

-and whether the statement reached consensus

**For round 2, I would be grateful if you could:**

- RE-SCORE ALL STATEMENTS regardless of whether they have reached consensus or not (Part 1)

-rank the three items which you think are the “most important” and “least important” (Part 2)

# **References**

Ajjawi, R. *et al.* (2017). Developing a national dental education research strategy: Priorities, barriers and enablers. *BMJ Open*, **7,** e013129.

Albanese, M. A. (2000). Challenges in using rater judgements in medical education. *Journal of Evaluation in Clinical Practice*. **6**(3), 305–319.

Alexander, H.A. (1996). Physiotherapy student clinical education: The influence of subjective judgements on observational assessment. *Assessment & Evaluation in Higher Education*. **21**(4), 357–366.

Ali, J. and Goh, A. (2017). Student perceptions of workplace-based assessment. *Clinical Teacher*. **14**(5), 319–324.

Allen, T. C. (2013). Graduated responsibility for pathology residents: No time for half measures. *Archives of Pathology and Laboratory Medicine*. **137**(4):457-61.

Allsbrook, W. C. *et al.* (2001). Interobserver reproducibility of Gleason grading of prostatic carcinoma: Urologic pathologists. *Human Pathology*. **32**(1), 74–80.

Anderson, L. and Krathwohl, D. (2001). *A taxonomy for learning, teaching, and assessing : A Revision of Bloom’s Taxonomy of Educational Objectives*. 1st ed. New York: Longman.

Arnold, L. (2002). Assessing professional behavior: yesterday, today, and tomorrow. *Academic medicine : journal of the Association of American Medical Colleges*. **77**(6), 502–515.

Arnold, L. *et al.* (2007). Can there be a single system for peer assessment of professionalism among medical students? A multi-institutional study. *Academic medicine : Journal of the Association of American Medical Colleges*. **82**(6), 578–86.

Artino, A. R. *et al.* (2014). Developing questionnaires for educational research: AMEE Guide No. 87. *Medical Teacher*. **36**(6), 463–474.

Asch, E and Saltzberg, D. K. S. (1998). Reinforcement of self-directed learning and the development of professional attitudes through peer- and self-assessment. *Acad Med*. **73** (5), 575.

Atkins, S. and Murphy, K. (1993). Reflection: a review of the literature. *Journal of Advanced Nursing*. **18**(8), 1188–1192.

Barbour, R. S. (2001). Checklists for improving rigour in qualitative research: a case of the tail wagging the dog? *British Medical Journal*. **322**(7294),1115–1117.

Bargagliotti, T. (1999). Reducing Threats to the Implementation of a Competency-Based Performance Assessment System. *The* *Online Journal of Issues in Nursing.* **4**(2), 1–10.

# Barrett, A. *et al.* (2016). A BEME (Best Evidence in Medical Education) review of the use of workplace-based assessment in identifying and remediating underperformance among postgraduate medical trainees: BEME Guide No. 43. *Medical Teacher*. **38**(12):1188-1198

Beard, J. D. (2008). Assessment of surgical skills of trainees in the UK. *Annals of the Royal College of Surgeons of England*. **90**(4), 282–5.

Beeston, S. and Higgs, J. (2001). Professional practice: Artistry and connoisseurship. In: H. Titchen and J. Higgs, eds. *Practice Knowledge and Expertise in the Health Professions.* Oxford: Butterworth-Heinemann. pp. 108–121.

Benner, P. (1982). From novice to expert. *The American Journal of Nursing*. **82**(3), 402–407.

Benner, P., Tanner, C., Chesla, C. (1996). *Expertise in Nursing Practice: Caring, Clinical Judgment, and Ethics.* 2nd ed*.* New York: Springer.

Bindal, T., Wall, D. and Goodyear, H. M. (2011). Trainee doctors’ views on workplace-based assessments: Are they just a tick box exercise? *Medical Teacher*. **33**(11), 919–927.

Black, D. (2016). An end to box ticking: an overhaul of competency based education. *British Medical Journal*. **353**(3020).

Black, D. and Welch, J. (2009). The under-preforming trainee - concerns and challenges for medical educators. *The clinical teacher*. **6**(2), 79–82.

Bloom, B. S. *et al.* (1956). *Taxonomy of educational objectives: The classification of educational goals: Handbook 1, cognitive domain*. London: Longmans

Bloor, M. *et al* (2015). Useful but no Oracle: Reflections on the use of a Delphi Group in a multi-methods policy research study. *Qualitative Research*, **15**(1), 57–70.

Boenink, A. D. *et al.* (2005). The effects of teaching medical professionalism by means of vignettes: an exploratory study. *Medical Teacher*. **27**(5), 429–432.

Boshuizen, H. P. A. and Schmidt, H. G. (2000). The development of clinical reasoning expertise. In: J Higgs and M Jones, eds. *Clinical Reasoning in the Health Professions*. 2nd ed. Oxford: Butterworth-Heinemann. pp. 15–22.

Boud, D. (2000). Sustaineable Assessment: Rethinking assessment for the learning society. *Studies* in *Continuing Education*. **22**(2), 151–167.

Boulkedid, R. *et al.* (2011). Using and reporting the Delphi method for selecting healthcare quality indicators: A systematic review. *PLoS ONE*, **6**(6):e20476

Boursicot, K. A. M., Roberts, T. E. and Burdick, W. P. (2011). Structured Assessments of Clinical Competence. In: T. Swanwick, ed. *Understanding Medical Education: Evidence, Theory and Practice*. London: Wiley-Blackwell. pp. 246–258.

Braun, V. and Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*. **3**(2), 77–101.

Brehmer, A. and Brehmer, B. (1988). Chapter 3 What Have we Learned about Human Judgment from Thirty Years of Policy Capturing? *Advances in Psychology*. **54**, 75–114.

Brennan, R. L. (2001). An essay on the history and future of reliability from the perspective of replications. *Journal of Educational Measurement*. **38**(4), 295–317.

Brightwell, A. and Grant, J. (2013). Competency-based training: Who benefits? *Postgraduate Medical Journal*. **89**(1048), 107–110.

Brookhart, S. M. (2001). Successful Students’ Formative and Summative Uses of Assessment Information. *Assessment in Education*. **8**(2), 153-168.

Brown, G., Manogue, M. and Rohlin, M. (2002). Assessing attitudes in dental education: Is it worthwhile? *British Dental Journal*. **193**(12), 703–707.

Brualdi, A. (1999). Traditional and Modern Concepts of Validity. *ERIC/AE Digest*. 1–6. ERIC identifier: ED435714

Brunswik, E. (1952). *The conceptual framework of psychology.* Chicago: University of Chicago Press.

Brunyé, T. T. *et al.* (2014). Eye movements as an index of pathologist visual expertise: A pilot study. *PLoS ONE*. **9**(8): e103447.

Bryman, A. (2012) ‘Qualitative data analysis. In: A. Bryman , ed. *Social Research Methods*. 4th ed. Oxford: Oxford University Press. pp. 564–5588.

Bullock, A. D. *et al.* (2009). How ratings vary by staff group in multi-source feedback assessment of junior doctors. *Medical Education*. **43**(6), 516–520.

Bullock, A. D. (2010). Guest Editorial. *Eur J Dent Educ*. **14**(2), 65–68.

Burg, F. D., Lloyd, J. S. and Templeton, B. (1982). Competence in medicine. *Medical Teacher*. **4**(2), 60–64.

Burr, B., Wade, W. and Watts, H. (2012). Improving workplace based assessments. *British Medical Journal*. **345**:e4986

Bussolati, G. (2006). Dissecting the pathologist’s brain: mental processes that lead to pathological diagnoses. *Virchows Archiv*. **448**(6), 739–743.

van de Camp, K. *et al.* (2004). How to conceptualize professionalism: A qualitative study. *Medical Teacher*. **26**(8), 696–702.

Carifio, J. and Perla, R. J. (2007). Ten Common Misunderstandings, Misconceptions, Persistent Myths and Urban Legends about Likert Scales and Likert Response Formats and their Antidotes. *Journal of Social Sciences*. **3**(3), 106–116.

Carr, S. J. (2004). Assessing clinical competency in medical senior house officers: Now and why should we do it? *Postgraduate Medical Journal*. **80**(940), 63–66.

Carraccio, C. and Burke, A. E. (2010). Beyond Competencies and Milestones: Adding Meaning Through Context. *Journal of Graduate Medical Education*. **2**(3), 419–422.

Carraccio, C. *et al.* (2002). Shifting paradigms: From flexner to competencies. *Academic Medicine*. **77**(5), 361–367.

Carter, R. (1985). A taxonomy of objectives for professional education. *Studies in Higher Education*. **10**(2), 135–149.

Cash, K. (1995). Benner and expertise in nursing: a critique. *International Journal of Nursing Studies*. **32**(6), 527–534.

ten Cate, O. (2005). Entrustability of professional activities and competency-based training. *Medical Education*. **39**(12), 1176–1177.

ten Cate, O. (2013). Nuts and Bolts of Entrustable Professional Activitites. *Journal of Graduate Medical Education*, **5**(1), 157–158.

ten Cate, O and Billett, S. (2014). Competency-based medical education: Origins, perspectives and potentialities. *Medical Education*. **48**(3), 325–332.

ten Cate, O., Snell, L. and Carraccio, C. (2010). Medical competence: the interplay between individual ability and the health care environment. *Medical Teacher*. **32**(8), 669–675.

ten Cate, O., Kusurkar, R. and Williams, G. (2011). How self-determination theory can assist our understanding of the teaching and learning processes in medical education. AMEE guide No. 59. *Medical Teacher*. **33**(12), 961–973.

ten Cate, O. *et al* (2015). Curriculum development for the workplace using Entrustable Professional Activities (EPAs): AMEE Guide No. 99. *Medical Teach*er. **37**(11), 983–1002.

Chang, A. *et al.* (2013). Transforming Primary Care Training—Patient-Centered Medical Home Entrustable Professional Activities for Internal Medicine Residents. *Journal of General Internal Medicine*. **28**(6), 801-809.

Chapman, J. (1998). Agonising about assessment. In: D. Fish and C. Coles, eds. *Developing Professional Judgement in Health Care: Learning through the critical appreciation of practice*. Oxford: Butterworth-Heinemann. pp. 157–181.

Chelliah, K. K. and Arumugam, Z. (2012). Does Reflective Practice Enhance Clinical Competency in Medical Imaging Undergraduates? *Procedia - Social and Behavioral Sciences*. **60**, 73–77.

Cohen, S. J., Halvorson, H. W. and Gosselink, C. A. (1994). Changing physician behavior to improve disease prevention. *Preventive Medicine*. **23**(3), 284–291.

Cook, C., Heath, F. and Thompson, R. L. (2000). A meta-analysis of response rates in Web- or internet-based surveys. *Educational and Psychological Measurement*. **60**(6), 821–836.

Cook, D. A. *et al.* (2009). Effect of rater training on reliability and accuracy of mini-CEX scores: A randomized, controlled trial. *Journal of General Internal Medicine*. **24**(1), 74–79.

Cooksey, R.W. (1996a). *Judgement Analysis: Theory, Methods and Applications.* Bingley: Emerald Group Publishing.

Cooksey, R. W. (1996b). The Methodology of Social Judgement Theory. *Thinking & Reasoning*. **2**(2-3), 141–173.

Cowpe, J. *et al.* (2010). Profile and competences for the graduating European dentist - update 2009. *European Journal of Dental Education*. **14**(4), 193–202.

Cresswell, J. W. *et al.* (2003). Advanced mixed methods research designs. In:A. Tashakkori and C.Teddlie, eds. *Handbook of Mixed Methods in Social and Behavioral Research*. Thousand Oaks, California: Sage Publications. pp. 209–240.

Creswell, J. W. (2014). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. 4th ed. Thousand Oaks, California: Sage publications

Crooks, T. J., Kane, M. T. and Cohen, A. S. (1996). Threats to the Valid Use of Assessments. *Assessment in Education: Principles, Policy & Practice*. **3**(3), 265–286.

Cross, S. S. (2005). Aptitude testing and assessment of training progress in histopathology. *Current Diagnostic Pathology*. **11**(5), 299–307.

Cross, S.S., Dennis, T and Start, R.D. (2002). The area under the receiver operating characteristic (ROC) curve can be used as a sensitive measure of progress during the early stages of histopathology training. *Histopathology*, **41**(1 Suppl), 206.

Cross, S. S. *et al.* (2000). What levels of agreement can be expected between histopathologists assigning cases to discrete nominal categories? A study of the diagnosis of hyperplastic and adenomatous colorectal polyps. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc*. **13**(9), 941–944.

Cross, V., Hicks, C. and Barwell, F. (2001). Assessment and Evaluation in Higher Education Exploring the Gap Between Evidence and Judgement : Using video vignettes for practice-based assessment of physiotherapy undergraduates. *Assessment & Evaluation in Higher Education*. **26**(3), 189–212.

Crossley, J. (2013). Validity and truth in assessment. *Medical Education*. **47**(12), 1152–1154.

Crossley, J and Jolly, B. (2012). Making sense of work-based assessment: Ask the right questions, in the right way, about the right things, of the right people. *Medical Education*. **46**(1), 28–37.

Crossley, J. *et al.* (2002). Generalisability: A key to unlock professional assessment. *Medical Education*. **36**(10), 972–978.

Crossley, J. *et al.* (2011). Good questions, good answers: construct alignment improves the performance of workplace-based assessment scales. *Medical Education*. **45**(6), 560-569.

Cruess, S. R. and Cruess, R. L. (1997). Professionalism must be taught. *British Medical Journal*. **315**(7123), 1674–1677.

Cruess, S. R. *et al.* (2006). The Professionalism Mini-evaluation Exercise: a preliminary investigation. *Academic medicine : Journal of the Association of American Medical Colleges*. **81**(10 Suppl), S74-78.

Cusimano, M. D. (1996). Standard setting in medical education. *Academic Medicine*. 71(**10**), 112-20.

Dalton, L. W., Page, D. L. and Dupont, W. D. (1994). Histologic grading of breast carcinoma. A reproducibility study. *Cancer*. **73**(11), 2765–2770.

Damodaran, A., Shulruf, B. and Jones, P. (2017). “Trust” versus “competency” in the workplace. *Medical Education*. **51**(3), 338-338.

Daniel, W. (1999). *Biostatistics: A Foundation for Analysis in the Health Sciences.* 7th Ed. New York, NY: John Wiley & Sons, Ltd.

Dauphinee, W. D. (1995). Assessing clinical performance: Where do we stand and what might we expect?. *The Journal of the American Medical Association*. **274**(9), 741–743.

Davey, D. D. *et al.* (1996). Cytopathology and the pathology resident: A survey of residency program directors. *Archives of Pathology and Laboratory Medicine*. **120**(1), 101–104.

Davies, H. *et al.* (2009). Initial evaluation of the first year of the Foundation Assessment Programme. *Medical Education*. **43**(1), 74–81.

Dean, B. J. F. and Duggleby, P. M. (2013). Foundation doctors’ experience of their training: a questionnaire study. *JRSM Short Reports*. **4**(1), 1–7.

Denzin, N. K. and Lincoln, Y. S. (2005). Introduction: The Discipline and Practice of Qualitative Research. In: N. K. Denzin and Y.S. Lincoln, eds. *Handbook of Qualitative Research*. Thousand Oaks, California: Sage publishing Ltd. pp. 1–32.

Department of Health. (2007). *A Reference Guide to Postgraduate Specialty Training in the UK: The Gold Guide.* London: Department of Health.

Diamond, I. R. *et al.* (2014). Defining consensus: A systematic review recommends methodologic criteria for reporting of Delphi studies. *Journal of Clinical Epidemiology*. **67**(4), 401–409.

Diamond, J. *et al.* (2002). A computer-based training system for breast fine needle aspiration cytology. *The Journal of pathology*. **196**(1), 113–121.

Dijksterhuis, M.G. *et al.* (2009) ‘Assessment of competence and progressive independence in postgraduate clinical training. *Medical Education*. 43(12), 1156-1165

Donato, A. A. *et al.* (2008). Evaluation of a novel assessment form for observing medical residents: A randomised, controlled trial. *Medical Education*. **42**(12), 1234–1242.

Dowding, D. and Thompson, C. (2003). Measuring the quality of judgement and decision-making in nursing. *Journal of Advanced Nursing*. **44**(1), 49–57.

Downing, S. M. (2002). Assessment of Knowledge with Written Test Forms. In: G. Norman, C. van der Vleuten and D. Newble, eds. *International Handbook of Research in Medical Education*. Dordrecht: Kluwer Academic Publishers. pp. 647–672.

Downing, S. M. (2003). Validity: On the meaningful interpretation of assessment data. *Medical Education*. **37**(9), 830–837.

Dreyfus, H.L and Dreyfus, S.E. (1980). A five-stage model of the mental activities involved in directed skill acquisition. University of California, Berkeley: Operations Research Centre. pp. 1-18.

Dreyfus, H.L and Dreyfus, S. E. (1986). *Mind Over Machine: The Power of Human Intuition and Expertise in the Age of the Computer.* Oxford: Basil Blackwell.

Dreyfus, H. L and Dreyfus, S. E. (1996). The relationship of theory and practice in the acquisition of skill. In: P. Benner, C. Tanner and C. Chelsa, eds. *Expertise in Nursing Practice: Caring, clinical judgment, and ethics.* New York: Springer Publishing Company. pp. 29–47.

Duke, M. (1996). Clinical evaluation--difficulties experienced by sessional clinical teachers of nursing: a qualitative study. *Journal of advanced nursing*. **23**(2), 408–414.

Dunphy, B. C. *et al.* (2010). Cognitive elements in clinical decision-making. *Advances in Health Sciences Education.* **15**(2), 229–250.

Durning, S. J. *et al.* (2010). Perspective: Redefining context in the clinical encounter: Implications for research and training in medical education. *Academic Medicine*. **85**(5), 894–901.

Einhorn, H. J. (1974). Expert judgment: Some necessary conditions and an example. *Journal of Applied Psychology*. **59**(5), 562–571.

Epstein, R. M. and Hundert, E. M. (2002). Defining and assessing professional competence. *Journal of the American Medical Association*. **287**(2), 226–235.

Eraut, M. (1994). *Developing Professional Knowledge and Competence*. London: Falmer Press.

Eraut, M. (1998). Concepts of competence. *Journal of Interprofessional Care*. **12**(2), 127–139.

Essers, G. *et al.* (2011). Identifying context factors explaining physician’s low performance in communication assessment: An explorative study in general practice. *BMC Family Practice*. **12** (138), 1–8.

Eva, K. W. and Regehr, G. (2008). I’ll never play professional football and other fallacies of self-assessment. *Journal of Continuing Education in the Health Professions*. **28**(1), 14–19.

Fandel, T. M. *et al.* (2008). Do we truly see what we think we see? The role of cognitive bias in pathological interpretation. *Journal of Pathology*. **216**(2), 193–200.

Feldt, L. S. and Brennan, R. L. (1989). Reliability. In: R. L. Linn, ed. The American Council on Education/Macmillan series on higher education*. Educational measurement,* 3rd ed. New York, NY, England: Macmillan Publishing Co, Inc; American Council on Education. pp. 105–146.

Ferguson, R. P. (2014). Professionalism: hard to measure but you know it when you see it. *Journal of Community Hospital Internal Medicine Perspectives*. **4**(2), 24226.

Field, J. C., Cowpe, J. G. and Walmsley, A. D. (2017). The Graduating European Dentist: A New Undergraduate Curriculum Framework. *European Journal of Dental Education*. **21**(1 suppl), 2–10.

Finall, A. (2012). Trainers' perceptions of the direct observation of practical skills assessment in histopathology training: a qualitative pilot study. *Journal of Clinical Pathology*. **65**(6), 538–540.

Finall, A. and Allery, L. (2016). Ready, steady, go! What do histopathology trainees think they need from training to enable them to develop autonomy in surgical pathology reporting? *Journal of Clinical Pathology*. **69**(1), 42–46.

Fincham, J. E. (2008). Response rates and responsiveness for surveys, standards, and the Journal. *American journal of pharmaceutical education*. **72**(2), 43.

Frank, J. R. (Ed.) (2005). *The CanMEDS 2005 physician competency framework. Better standsards. Better physicians. Better care.* Ottawa, Canada: The Royal College of Physicians and Surgeons of Canada.

Frank, J. R. *et al.* (2010a). Competency-based medical education: theory to practice. *Medical Teacher*. **32**(8), 638–45.

Frank, J. R. *et al.* (2010b). Toward a definition of competency-based education in medicine: A systematic review of published definitions. *Medical Teacher*. **32**(8), 631–637.

Friedman, M. and Mennin, S. P. (1991). Rethinking critical issues in performance assessment. *Academic Medicine.* **66**(7), 390–395.

Gardiner, J. (Ed.) (2006). *Assessment and Learning*. London, UK: Sage publications.

Garland, R. (1991). The Mid-Point on a Rating Scale: Is it Desirable? *Marketing Bulletin*. **2**(3), 66–70.

General Dental Council. (2013). Standards for the Dental Team. (Online). London: General Dental Council. (Viewed 20 August 2019). Available from: https://standards.gdc-uk.org/Assets/pdf/Standards%20for%20the%20Dental%20Team.pdf

General Medical Council. (2009). *Tomorrow’s Doctors - Outcomes and standards for undergraduate medical education*. London: General Medical Council

General Medical Council. (2010). *Workplace-Based Assessment: A guideline for implementation*. London: General Medical Council.

General Medical Council. (2011). *Learning and assessment in the clinical environment: the way forward*. London: The General Medical Council.

General Medical Council. (2015). *Outcomes for graduates* (*Tomorrow’s Doctors*). London: General Medical Council.

Gigerenzer, G. and Gaissmaier, W. (2011). Heuristic Decision Making. *Annual Review of Psychology*. **62**(1), 451–482.

Ginsburg, S. *et al.* (2012). “It Dependsa”: Results of a qualitative study investigating how practicing internists approach professional dilemmas. *Academic Medicine*. **87**(12), 1685–1693.

Glaser, B. G., Strauss, A. L. T. (1967). *The discovery of grounded theory: Strategies for qualitative research.* New York, NY: Aldine De Gruyter.

Goldstein, N. S. (1999). Diagnostic errors in surgical pathology. *Clinics in Laboratory Medicine*. **19**(4), 743–756.

Gomez-Mejia, G. (1998). Evaluating employee performance: Does the appraisal instrument make a difference? *Journal of Organizational Behavior Management*. **9**(2), 155–172.

Govaerts, M. J. B. *et al.* (2007). Broadening perspectives on clinical performance assessment: Rethinking the nature of in-training assessment. *Advances in Health Sciences Education*. **12**(2), 239–260.

Govaerts, M.J.B. and Van der Vleuten, C. P. (2013). Validity in work-based assessment: Expanding our horizons. *Medical Education*. **47**(12), 1164–1174.

Govaerts, M. J. B., van der Vleuten, C. P. M. and Schuwirth, L. W. T. (2002) . Optimising the reproducibility of a performance-based assessment test in midwifery education. *Advances in Health Sciences Education*. **7**(2), 133–145.

Graber, M., Gordon, R. and Franklin, N. (2002). Reducing diagnostic errors in medicine: What’s the goal? *Academic Medicine*. **77**(10), 981–992.

von der Gracht, H. A. (2012). Consensus measurement in Delphi studies. Review and implications for future quality assurance. *Technological Forecasting and Social Change*. **79**(8), 1525–1536.

von der Gracht, H. A. and Darkow, I. L. (2010). Scenarios for the logistics services industry: A Delphi-based analysis for 2025. *International Journal of Production Economics*. **127**(1), 46–59.

Graham, M. J. *et al.* (2009). Systems-Based Practice Defined: Taxonomy Development and Role Identification for Competency Assessment of Residents. *Journal of Graduate Medical Education*. **1**(1), 49–60.

Grant. J. (1999). The incapacitating effects of competence: A critique. *Sciences Education*. **4**(3), 271–277.

Gray, J. D. (1996). Global rating scales in residency education. *Academic Medicine.* **71**(1 Suppl), S55–S63.

Greatorex, J. and Dexter, T. (2000). An accessible analytical approach for investigating what happens between the rounds of a Delphi study. *Journal of Advanced Nursing*. **32**(4), 1016–1024.

Greenhalgh, T. and Taylor, R. (1997). How to read a paper: Papers that go beyond numbers (qualitative research). *British Medical Journal*. **315**(7110), 740–743.

Grieveson, B. *et al.* (2011). Evaluation of workplace based assessment tools in dental foundation training. *British Dental Journal*. **211**(4), E8.

Hambleton, R.K. and Powell, P. (1983). A Framework for Viewing the Process of Standard Setting. *Evaluation and the Health Professions.* **6**(1), 3-24.

Hamilton, P. W. *et al.* (2009). Do we see what we think we see. The complexities of morphological assessment. *Journal of Pathology*. **218**(3), 285–291.

Hammond, K. R., Hursch, C. J. and Todd, F. J. (1964). Analyzing The Components Of Clinical Inference. *Psycholigcal Review*. **71**(6), 438–456.

Harden, R. M. and Crosby, J. R. (1999). AMEE Guide No.14: Outcome-based education: Part 5-From competency to meta-competency: a model for the specification of learning outcomes. *Medical Teacher*. **21**(6), 546-552.

Hargreaves, K. (2016). Reflection in Medical Education. *Journal of University Teaching and Learning*. **13**(2), 6.

Harlen, W. (2007). Criteria for evaluating systems for student assessment. *Studies in Educational Evaluation*. **33**(1), 15–28.

Harman, N. L. *et al.* (2015). The importance of integration of stakeholder views in core outcome set development: Otitis Media with Effusion in children with cleft palate. *PLoS One.* **10**(6), e0129514

Harris, I. (1993). New expectations for professional competence. In: L. Curry and J. F. Wergin, eds. *Educating Professionals: Responding to new expectations for competence and accountability*. San Francisco: Jossey-Bass. pp. 17–52.

Harris, I. B. and Simpson, D. (2005). Christine McGuire: At the heart of the maverick measurement maven. *Advances in Health Sciences Education*. **10**(1), 65–80.

Hodges, B.D. (2010). A tea-steeping or i-Doc model for medical education? *Academic Medicine*. **85**(9 Suppl), S34-44.

Hauer, K. E. *et al.* (2013). Identifying entrustable professional activities in internal medicine training. *Journal of Graduate Medical Education*. **5**(1), 54–9.

Hauer, K. E., Holmboe, E. S. and Kogan, J. R. (2011). Twelve tips for implementing tools for direct observation of medical trainees’ clinical skills during patient encounters. *Medical Teacher*. **33**(1), 27–33.

Hawkins, R. E. *et al.* (2010). Constructing a validity argument for the mini-clinical evaluation exercise: A review of the research. *Academic Medicine*. **85**(9),1453–1461.

Hejri, S. M. *et al.* (2013). Assessing the reliability of the borderline regression method as a standard setting procedure for objective structured clinical examination. *Journal of Research in Medical Sciences.***18**(10), 887–91.

Higgs, J. and Hunt, A. (1999). Rethinking the beginning practitioner: introducing the “Interactional Professional”. In: J. Higgs and H. Edwards, eds. *Educating Beginning Practitioners. Challenges for health professional education*. Oxford: Butterworth-Heinemann. pp. 10–18.

Hofer, B. K. and Pintrich, P. R. (1997). The Development of Epistemological Theories: Beliefs About Knowledge and Knowing and Their Relation to Learning. *Review of Educational Research*. **67**(1), 88–140.

Hojat, M. *et al.* (2007). Components of postgraduate competence: Analyses of thirty years of longitudinal data. *Medical Education*. **41**(10), 982–989.

Holey, E. A. *et al.* (2007). An exploration of the use of simple statistics to measure consensus and stability in Delphi studies. *BMC Medical Research Methodology*. **29**(7), 52.

Holmboe, E. S., Hawkins, R. E. and Huot, S. J. (2004). Effects of training in direct observation of medical residents’ clinical competence: a randomized trial. *Ann Intern Med*. **140**(11), 874–881.

Holmboe, E. S. *et al.* (2003). Construct validity of the miniclinical evaluation exercise (miniCEX). *Academic Medicine.* **78**(8), 826–830.

Humphrey-Murto, S. *et al.* (2017). Using consensus group methods such as Delphi and Nominal Group in medical education research. *Medical Teacher*. **39**(1), 14–19.

# Hunt, A., Adamson, B. and Harris, L. (1999). Community and workplace expectations of health science graduates. In: J. Higgs and H. Edwards, eds. *Educating Beginning Practitioners*. Challenges for health professional education. Oxford: Butterworth-Heinemann. pp. 38–45.

Iglesias, C. P. (2002). Increasing response rates to postal questionnaires. *British Medical Journal*. **325**(7361), 444–444.

Ilott, I. and Murphy, R. (1997). Feelings and Failing in Professional Training: the assessor’s dilemma. *Assessment and Evaluation in Higher Education*. **22**(3), 307–316.

Irvine, D. (1997). The performance of doctors. 1: Professionalism and self regulation in a changing world. *British Medical Journal.* **314**(7093), 1540–1542.

Jones, A. (1999). The place of judgement in competency-based assessment. *Journal of Vocational Education and Training*. **51**(1), 145–160.

Jones, J. and Hunter, D. (1995). Qualitative Research: Consensus methods for medical and health services research. *British Medical Journal*. **311**(7001), 376-380.

Jones, M. D. *et al.* (2011). Perspective: Competencies, outcomes, and controversy-linking professional activities to competencies to improve resident education and practice. *Academic Medicine.* **86**(2), 161–165.

Jonsson, A. and Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*. **2**(2), 130–144.

Jünger, S. *et al.* (2017). Guidance on Conducting and REporting DElphi Studies (CREDES) in palliative care: Recommendations based on a methodological systematic review. *Palliative Medicine*. **31**(8), 684–706.

Kak, N., Burkhalter, B. and Cooper, M. (2001). Measuring the Competence of Healthcare Providers. *Operations Research.* **2**(1), 1-28.

Kane, M. (2006). Validation. In: R. Brennan, ed. *Educational Measurement*. 4th ed. Westport, CT: Praeger. pp. 17–64.

Kaufman, D. M. *et al.* (2000). A comparison of standard-setting procedures for an OSCE in undergraduate medical education. *Academic Medicine*. **75**(3), 267–271.

Kaufman, D. M. and Mann, K. (2010). Teaching and learning in medical education: How theory can inform practice. In: T. Swanwick, ed. *Understanding Medical Education: Evidence, Theory and Practice*. Oxford, UK: Wiley-Blackwell. pp. 16–36.

Keen, A. J. A., Klein, S. and Alexander, D. A. (2003). Assessing the communication skills of doctors in training: Reliability and sources of error. *Advances in Health Sciences Education*. **8**(1), 5–16.

Kember, D. *et al.* (2000). Development of a questionnaire to measure the level of reflective thinking. *Assessment and Evaluation in Higher Education*. **25**(4) 381–395.

Kennedy, T. J. T. *et al.* (2005). Progressive independence in clinical training: a tradition worth defending? *Academic Medicine.* **80**(10 Suppl), S106-11.

Kilminster, S. *et al.* (2007). AMEE Guide No. 27: Effective educational and clinical supervision. *Medical Teacher*. **29(**1), 2–19.

Kirton, J. *et al.* (2013). A national evaluation of workplace-based assessment tools (WPBAs) in foundation dental training: a UK study. Effective and useful but do they provide an equitable training experience? *British Dental Journal*. **214**(6), 305–309.

Knight, P. T. (2000). The Value of a Programme-wide Approach to Assessment. *Assessment & Evaluation in Higher Education*. **25**(3), 237–251.

Kogan, J. R., Holmboe, E. S. and Hauer, K. E. (2009). Tools for direct observation and assessment of clinical skills of medical trainees: a systematic review. *JAMA*. **302**(12), 1316–1326.

Kogan, J. R. *et al.* (2011). Opening the black box of clinical skills assessment via observation: A conceptual model. *Medical Education*. **45**(10), 1048–1060.

Kogan, J. R. *et al.* (2012). Faculty staff perceptions of feedback to residents after direct observation of clinical skills. *Medical Education*. **46(**2), 201–215.

Kolb.D (1984). *Experiential learning*. Englewood Cliffs, NJ: Prentice-Hall.

Krupinski, E. A., Graham, A. R. and Weinstein, R. S. (2013). Characterizing the development of visual search expertise in pathology residents viewing whole slide images. *Human Pathology*. **44**(3), 357–364.

Krupinski, E. A. *et al.* (2006). Eye-movement study and human performance using telepathology virtual slides. Implications for medical education and differences with experience. *Human Pathology*. **37**(12), 1543–1556.

Kusurkar, R. A. *et al.* (2012). Have motivation theories guided the development and reform of medical education curricula? A review of the literature. *Academic Medicine*. **87**(6), 735–743.

Lave, J. and Wenger, E. (1991). *Situated Learning. Legitimate peripheral participation*. Cambridge, UK: University of Cambridge Press.

Leach, L., Neutze, G. and Zepke, N. (2001). Assessment and empowerment: Some critical questions. *Assessment & Evaluation in Higher Education*. **26**(4), 293–305.

Lee, J., Chan, A. C. M. and Phillips, D. R. (2006). Diagnostic practise in nursing: A critical review of the literature. *Nursing and Health Sciences*. **8**(1),57–65.

Lee, V., Brain, K. and Martin, J. (2017). Factors Influencing Mini-CEX Rater Judgments and Their Practical Implications: A Systematic Literature Review. *Academic Medicine*. 92(6), 880–887.

Leung, W. C. (2002). Competency based medical training: review. *British Medical Journal.* **325**(7366), 693-696

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*. **22**(140), 1–55.

Lincoln, Y. S. and Guba, E. G. (1985). *Naturalistic Inquiry.* Beverly Hills, California: Sage Publications Inc

Lind, A. C. *et al.* (1995). Prospective Peer-Review in Surgical Pathology. *American Journal of Clinical Pathology*. **104**(5), 560–566.

Linn, R. L., Baker, E. L. and Dunbar, S. B. (1991). Complex, Performance-Based Assessment: Expectations and Validation Criteria. *Educational Researcher*. **20**(8), 15–21.

Lockyer, J. (2003). Multisource feedback in the assessment of physician competencies. *The Journal of Continuing Education in the Health Professions*. **23**(1), 4–12.

Long, D. M. (2000). Competency-based residency training: The next advance in graduate medical education. *Academic Medicine*. **75**(12), 1178–1183.

Lowry, S. (1993). *Medical Education*. London, UK: BMJ books. pp. 46–47.

Luijk, S. J. Van *et al.* (2000). Assessing professional behaviour and the role of academic advice at the Maastricht Medical School. *Medical Teacher*. **22**(2), 168–172.

MacDougall, M. (2015). Variation in assessment and standard setting practices across UK undergraduate medicine and the need for a benchmark. *International Journal of Medical Education*. 31(6), 125–135.

MacLellan, E. (2001). Assessment for Learning: the differing perceptions of tutors and students. *Assessment and Evaluation in Higher Education*. **26**(4), 307–318.

Mann, K., Gordon, J. and MacLeod, A. (2009). Reflection and reflective practice in health professions education: A systematic review. *Advances in Health Sciences Education*. **14**(4), 595–621.

Marriott, J. *et al.* (2011). Evaluation of procedure-based assessment for assessing trainees skills in the operating theatre. *British Journal of Surgery*. **98**(3), 450–457.

Massie, J. and Ali, J. M. (2015). Workplace-based assessment: a review of user perceptions and strategies to address the identified shortcomings. *Advances in Health Sciences Education*.**21**(2), 455-73.

Masters, G. N., Adams, R. J. and Wilson, M. (1999). Charting of student progress. In: G. N. Masters and J. P. Keeves, eds. *Advances in Measurement in Educational Research and Assessment*. Bingley, UK: Emerald Group Publishing. pp. 254–267.

Matveevskii, A., Moore, D. L. and Samuels, P. J. (2012). Competency and professionalism in medicine. *Clinical Teacher*. **9**(2), 75–79.

Mayer, R. C., Davis, J. H. and Schoorman, F. D. (1995). An integrative model of organisational trust. *The Academy of Management Review*. **20**(3), 709-734.

McCloskey, C. B., Dormen, R.E., Conran, R.M. (2017). Entrustable Professional Activities for Pathology: recommendations from the College of American Pathologists Graduate Medical Education Committee. *Academic Pathology*. **4**

McKavanagh, P., Smyth, A. and Carragher, A. (2012). Hospital consultants and workplace based assessments: How foundation doctors view these educational interactions? *Postgraduate Medical Journal*. **88**(1037), 119–124.

McKee, R. F. (2008). The Intercollegiate Surgical Curriculum Programme (ISCP). *Surgery*. **26**, 411–416.

Merlin, L. R. *et al.* (2014). A competency-based longitudinal core curriculum in medical neuroscience. *Neurology*. **83**(5), 456–462.

Messick, S. (1989). Validity. In: R. Linn, ed. *Educational Measurement*. 3rd ed. New York, NY: American Council on Education and Macmillan. pp. 13–103.

Messick, S. (1994). The Interplay of Evidence and Consequences in the Validation of Performance Assessments. *Educational Researcher*. **23**(2), 13–23.

Miller, G. E. (1990). The assessment of clinical skills/competence/performance. *Academic Medicine.* **65**(9 Suppl), S63–S67.

Miller, A. and Archer, J. (2010). Impact of workplace based assessment on doctors’ education and performance: a systematic review. *British Medical Journal.* **341**, c5064.

Miller, A. H., Imrie, B. W. and Cox, K. (1998). *Student assessment in higher education: a handbook for assessing performance.* London: Kogan Page Limited.

Mislevy, R. J. (1996). Test Theory Reconceived. *Journal of Educational Measurement*. **33**(4), 379–416.

Mitchell, C. *et al.* (2011). Workplace-based assessments of junior doctors: Do scores predict training difficulties? *Medical Education*. **45**(12), 1190–1198.

Mitchell, C. *et al.* (2013). Workplace-based assessments in Foundation Programme training: Do trainees in difficulty use them differently? *Medical Education*. 47(3), 292–300.

van Mook, W. N. K. A. *et al.* (2007). Factors inhibiting assessment of students’ professional behaviour in the tutorial group during problem-based learning. *Medical Education*. **41**(9), 849–856.

van Mook, W. N. K. A. *et al.* (2009). Professionalism: Evolution of the concept. *European Journal of Internal Medicine*. **20**(4), 81-84.

Morrison, M. L. *et al.* (2002). Expert system support using a Bayesian belief network for the classification of endometrial hyperplasia. *The Journal of Pathology*. **197**(3), 403–14.

Moskal, B. M. and Leydens, J. A. (2000). Scoring rubric development: validity and reliability. *Practical Assessment, Research & Evaluation*. **7**(10), 1–10.

Mullin, M. H., Brierley, D. J. and Speight, P. M. (2015). Second opinion reporting in head and neck pathology: the pattern of referrals and impact on final diagnosis. *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*. **119**(6), 656–660.

Murphy DJ *et al.* (2009). The reliability of workplace-based assessment in postgraduate medical education and training: A national evaluation in general practice in the United Kingdom. *Advances in Health Sciences Education*. **14**(2), 219–232.

Murphy, E. *et al.* (1998a) ‘Qualitative research methods in health technology assessment: A review of the literature. *Health Technology Assessment*. **2**(16), iii-ix, 1-274.

Murphy, M. K. *et al.* (1998b) ‘Consensus development methods and their use in clinical guideline development. *Health Technology Assessment.* **2**(3), i–iv, 1–88.

Murray, E. *et al.* (2000) ‘The accountability of clinical education: Its definition and assessment. *Medical Education*, **34**(10), 871–879.

Naritoku, W. Y. *et al.* (2014) ‘The pathology milestones and the next accreditation system. *Archives of Pathology and Laboratory Medicine*. **138**(3), 307–315.

National Council On Measurement In Education (1999). *Standards for educational and psychological testing*. Washington DC: American Psychological Association, American Educational Research Association, American Council on Measurement in Education.

Neary, M. (2000). Supporting students’ learning and professional development through the process of continuous assessment and mentorship. *Nurse Educ Today*. **20**(6), 463–474.

Nicol, D. and Macfarlane‐Dick, D. (2006). Formative assessment and self-regulated learning : A model and seven principles of good feedback practice. *Studies in Higher Education*. **31**(2), 199–218.

Norcini, J. J. (2003). Setting standards on educational tests. *Medical Education*. **37**(5), 464–469.

Norcini, J. J. *et al.* (1997). Examiner differences in the mini-CEX. *Advances in Health Sciences Education: Theory and Practice*. **2**(1), 27–33.

Norcini, J. J. *et al.* (2003). The Mini-CEX: A Method for Assessing Clinical Skills. *Annals of Internal Medicine*. **138**(6), 476–481.

O’Cathain, A. and Thomas, K. J. (2004). "Any other comments?" Open questions on questionnaires – a bane or a bonus to research? *BMC Medical Research Methodology*. **4**(1), 25.

Oerlemans, M. *et al.* (2017). Should we assess clinical performance in single patient encounters or consistent behaviors of clinical performance over a series of encounters? A qualitative exploration of narrative trainee profiles. *Medical Teacher*. **39**(3), 300–307.

Oliver C.W. (Ed.) (2007). Developing and maintaining an assessment system: A PMETB guide to good practice. London, UK: PMETB. 50 p.

Owen, D. (2007). The Five Elements of Negligence. *Hofstra Law Review*. **35**(4), 1671-1686.

Parham, D. M. *et al.* (2010). Experts and performance in histopathology-a study in breast pathology. *Pathology, Research and Practice*. **206**(11), 749–52.

Pascal, R. (1993). Graded responsibility of residents in anatomic pathology: a survey and commentry. *Am J Clin Pathol*. **100**(4 Suppl), S41-3.

Patton, M. Q. (1990). Designing qualitative studies. In: M. Q. Patton, ed. *Qualitative evaluation and research methods*. Thousand Oaks, California: Sage Publications. pp. 169–186.

Pelgrim, E. A. M. *et al.* (2011). In-training assessment using direct observation of single-patient encounters: A literature review. *Advances in Health Sciences Education: Theory and Practice.* **16**(1), 131–142.

Pelgrim, E. A. M. *et al.* (2012). The process of feedback in workplace-based assessment: Organisation, delivery, continuity. *Medical Education*. **46**(6), 604–612.

Pena, G. P. and De Andrade-Filho, J. S. (2009). How does a pathologist make a diagnosis? *Archives of Pathology and Laboratory Medicine*. **133**(1), 124–132.

Pereira, E. A. and Dean, B. J. (2009). British surgeons’ experiences of mandatory online workplace-based assessment-reply to authors. *Journal of the Royal Society of Medicine*. **102**(9), 357–358.

Pittenger, A. L. *et al.* (2016). Entrustable professional activities for pharmacy practice. *American Journal of Pharmaceutical Education*. **80**(4), 57.

Pope, C., Ziebland, S. and Mays, N. (2000). Analysing qualitative data. *British Medical Journal*. **320**, 114–6.

Powell, H., DiMarco, A. and Saeed, S. (2014). Trainee and trainer opinion on increasing numbers of workplace-based assessments. *Bulletin of The Royal College of Surgeons of England*. **96**(5), 160–162.

Pratt, J. W. (1959). Remarks on Zeros and Ties in the Wilcoxon Signed Rank Procedures. *Journal of the American Statistical Association*. **54**(287), 655–667.

Prescott, L. E. *et al.* (2002). Facing the challenges of competency-based assessment of postgraduate dental training: Longitudinal Evaluation of Performance (LEP). *Medical Education*. **36**(1), 92–7.

Prescott-Clements, L. E. *et al.* (2011). Measuring the development of insight by dental health professionals in training using workplace-based assessment. *European Journal of Dental Education.* **15**(3), 159–64.

Ramsay, A.D. (1999). Errors in histopathology reporting: Detection and avoidance. *Histopathology*. **34**(6), 481–490.

Rasmussen, J. (1983). Skills, Rules, and Knowledge; Signals, Signs, and Symbols, and Other Distinctions in Human Performance Models. *IEEE Transactions on Systems, Man and Cybernetics*. **13**(3), 257–266.

Rees, C. and Shepherd, M. (2005). Students’ and assessors’ attitudes towards students’ self-assessment of their personal and professional behaviours. *Medical Education*. **39**(1), 30–39.

Rees, C. E. *et al.* (2014). Supervised learning events in the foundation programme: a UK-wide narrative interview study. *British Medical Journal Open*. **4**(10), e005980.

Regehr, G. *et al.* (1998). Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. *Academic Medicine.* **73**(9), 993–997.

Rekman, J. *et al.* (2016). A New Instrument for Assessing Resident Competence in Surgical Clinic: The Ottawa Clinic Assessment Tool. *Journal of Surgical Education*. **73**(4), 575-582.

Rest, J. R. (1984). Research on Moral Development: Implications for Training Counseling Psychologists. *The Counseling Psychologist*. **12**(3), 19–29.

Rethans, J. J. *et al.* (2002). The relationship between competence and performance: Implications for assessing practice performance. *Medical Education*. **36**(10), 901–909.

Rimmer, A. (2016). Junior doctors’ low morale is putting patients at risk, Royal College of Physicians warns. *British Medical Journal.* ***355****,* i6493.

Ritchie, J. S. L. (1994). Qualitative data analysis for applied policy research. In: A. Bryman an R.G. Burgess, eds. *Analyzing qualitative data*. London: Routledge. pp. 173–194.

Ritchie, J. and Lewis, J. (2003). *Qualitative Research Practice: A Guide for Social Science Students and Researchers*. London: Sage Publications.

Roa-Peña, L., Gómez, F. and Romero, E. (2010). An experimental study of pathologist’s navigation patterns in virtual microscopy. *Diagnostic Pathology*. **5**, 71.

Robertson, S., Rosenthal, J. and Dawson, V. (1997). Using assessment to promote student learning, facilitating Learning in Clinical Settings. In: L. McAllister, S. McLeod, M. Lincoln and D. Maloney, eds. Facilitating Learning in Clinical Settings. London: Stanley Thornes.

Romagnano, L. (2001). The myth of objectivity in mathematics assessment. *Mathematics Teacher*. **94**(1), 31–37.

Rose, S. H. *et al.* (2009). A historical perspective on resident evaluation, the accreditation council for graduate medical education outcome project and accreditation council for graduate medical education duty hour requirement. *Anesthesia and Analgesia*. **109**(1), 190–193.

Ryland, I. *et al.* (2006). The portfolio: how was it for you? Views of F2 doctors from the Mersey Deanery Foundation Pilot. *Clinical Medicine.* **6**(4), 378–80.

Sachdev, S. B and Verma, H. V. (2004). Relative importance of service quality dimensions: a multisectorial study. *Journal of Services Research*. **4**(1), 93-116.

Safrin, R. E. and Bark, C. J. (1993). Surgical Pathology Signout - Routine Review of Every Case by a 2nd Pathologist. *American Journal of Surgical Pathology*. **17**(11), 1190–1192.

Sale, J. E. M., Lohfeld, L. H. and Brazil, K. (2002). Revisiting the quantitative-qualitative debate: Implications for mixed-methods research. *Quality and Quantity*. **36**(1), 43–53.

Sandars, J. (2009). The use of reflection in medical education: AMEE Guide No. 44. *Medical Teacher*. **31**(8), 685–695.

Sandars, J. and Cleary, T. J. (2011). Self-regulation theory: Applications to medical education: AMEE Guide No. 58. *Medical Teacher*. **33**(11), 875–886.

Sargeant, J. M. *et al.* (2003). Responses of rural family physicians and their colleague and coworker raters to a multi-source feedback process: a pilot study. *Academic Medicine*. **78**(10 Suppl), S42-4.

## Scheibe, M., Skutch, M. and Schofer, J. (1975). Experiments in Delphi Methodology. In: H. A. Linstone and M. Turoff, eds. *The Delphi Method Techniques and Application*. Newark, NJ: Murray Turoff and Harold A. Linstone. pp. 257-281

Schmittdiel, J. *et al.* (2000). Effect of physician and patient gender concordance on patient satisfaction and preventive care practices. *Journal of General Internal Medicine*. **15**(11), 761–769.

Schon, D.A. (1983). *The reflective practitioner: how professionals think in action*. New York: Basic Books.

Schuwirth, L. and Ash, J. (2013). Assessing tomorrow’s learners: in competency-based education only a radically different holistic method of assessment will work. Six things we could forget. *Medical Teacher*. **35**(7),555-559.

Schuwirth L. W. and van der Vleuten, C. P. M. (2013). How to Design a Useful Test: The Principles of Assessment. In: T. Swanwick, ed. *Understanding Medical Education*. Oxford, UK: John Wiley & Sons, Ltd. pp. 241–254.

Schuwirth, L. W. *et al.* (2002). When enough is enough: A conceptual basis for fair and defensible practice performance assessment. *Medical Education*. **36**(10), 925-30.

Shalhoub, J., Marshall, D. C. and Ippolito, K. (2017). Perspectives on procedure-based assessments: A thematic analysis of semistructured interviews with 10 UK surgical trainees. *British Medical Journal Open*. **7**, e013417.

Shanteau, J. *et al.* (2002). Performance-based assessment of expertise: How to decide if someone is an expert or not. *European Journal of Operational Research*. **136**(2), 253–263.

Shaughnessy, A. F. *et al.* (2013). Entrustable Professional Activities in Family Medicine. *Journal of Graduate Medical Education*. **5**(1), 112–118.

Shepard, L. A. (2000). The Role of Assessment in a Learning Culture. *Educational Researcher*. **29**(7), 4–14.

Sia, JY, Harrison, J. (2015). Trainers’ and trainees’ perceptions of workplace-based assessment in orthodontics. *Journal of Orthodontics.***42**(4), 284–300.

Silber, C. G. *et al.* (2004). Do global rating forms enable program directors to assess the ACGME competencies? *Academic Medicine*. **79**(6), 549–556.

Silverman, D. (2006). What is Qualitative Research? In: D. Silverman, ed. *Interpreting Qualitative Data.* 3rd ed. London: Sage. pp. 33-61.

Sloan, D. A. *et al.* (1993). Use of an Objective Structured Clinical Examination (OSCE) to measure improvement in clinical competence during the surgical internship. *Surgery*. **114**(2), 343–350.

Smith, S. R. and Dollase, R. (1999). AMEE guide No. 14: Outcome-based education: Part 2 - Planning, implementing and evaluating a competency-based curriculum. *Medical Teacher*. **21**(1), 15–22.

Snadden, D. (1999). Portfolios–attempting to measure the unmeasurable. *Medical Education*. **33**(7), 478–479.

Southgate, L. (1999). Professional competence in medicine. *Hospital.Medicine.* **60**(3), 203–205.

Studdert, D. M. *et al.* (2006). Claims, Errors, and Compensation Payments in Medical Malpractice Litigation. *New England Journal of Medicine*. **354**(19), 2024–2033.

Swaffield, S. (2011). Getting to the heart of authentic Assessment for Learning. *Assessment in Education: Principles, Policy & Practice*. **18**(4), 433–449.

Swing, S. R. (2007). The ACGME outcome project: Retrospective and prospective. *Medical Teacher*. 29(7), 648–654.

Tammela, O. (2013). Applications of consensus methods in the improvement of care of paediatric patients: A step forward from a “good guess”. *Acta Paediatrica, International Journal of Paediatrics*. **102**(2), 111–115.

Teasdale, G. M. (2002). Learning from Bristol: report of the public inquiry into children’s heart surgery at Bristol Royal Infirmary 1984-1995. *British Journal of Neurosurgery*, **16**(3), 211–216.

Teman, N. R. *et al.* (2014). Entrustment of General Surgery Residents in the Operating Room: Factors Contributing to Provision of Resident Autonomy. *Journal of the American College of Surgeons*. **219**(4), 778–787.

The Royal College of Pathologists. (2009). *A Competency Based Framework for Graded Responsibility for Specialist Registrar and Specialty Registrars in Histopathology and Cytopathology*. (Online). The Royal College of Pathologists. (Viewed 20th August 2019). Available from: https://www.rcpath.org/uploads/assets/b0908b3c-aa01-479c-b4db45358cd8da25/competency-based-framework-for-graded-responsibility-arranged-by-level-of-competence.pdf

The Royal College of Pathologists. (2010). *Blueprints for the pathology assessment system: Histopathology*. (Online). The Royal College of Pathologists. (Viewed 20th August 2019). Available from: https://www.rcpath.org/uploads/assets/c6b83853-df18-4374-b5b6728731e24efa/histopathology-blueprint.pdf

The Royal College of Pathologists. (2015). *Curriculum for specialty training in histopathology*. (Online). The Royal College of Pathologists. (Viewed 20th August 2019). Available from:https://www.rcpath.org/uploads/assets/uploaded/f8695439-ce47-4046-b13c1f6202fd8468.pdf

The Royal College of Pathologists. (2018). *Workplace-based Assessment*. The Royal College of Pathologists. (Viewed 20th August 2019). Available from: https://www.rcpath.org/trainees/assessment/workplace-based-assessment-wpba.html

The Royal College of Physicians. (2005). Doctors in society: medical professionalism in a changing world.  *Report of a Working Party of the Royal College of Physicians of London*. London: Royal College of Physicians.

The Royal College of Physicians.(2018). *Never too busy to learn.* London: Royal College of Physicians.

Thompson, C. A. *et al.* (2005). Using social judgement theory to model nurses’ use of clinical information in critical care education. *Nurse Education Today*. **25**(1), 68–77.

Tiersma, E. S. M. *et al.* (2003). Visualising scanning patterns of pathologists in the grading of cervical intraepithelial neoplasia. *Journal of Clinical Pathology*. **56**(9), 677–680.

Tooke, J. (2008). Aspiring to excellence: final report of the independent inquiry into Modernising Medical Careers. London: MMC Inquiry.

Tormey, W. (2014). Education, learning and assessment: current trends and best practice for medical educators. *Irish Journal of Medical Science*. **84**(1), 1–12.

Torsney, K. M., Cocker, D. M. and Slesser, A. A. P. (2015). The modern surgeon and competency assessment: Are the workplace-based assessments evidence-based? *World Journal of Surgery*. **39**(3), 623–633.

Trathen, A and Gallagher, J. E. (2009). Dental professionalism: Definitions and debate. *British Dental Journal*. **206**(5), 249–253.

Tsichlaki, A. *et al.* (2017). Development of a core outcome set for orthodontic trials using a mixed-methods approach: Protocol for a multicentre study. *Trials*. **18**(1), 366.

Tyson, L. (2006). *Critical Theory Today: A User-Friendly Guide*. London: Routledge.

Unsworth, C., Harries, P. and Davies, M. (2015). Using social judgment theory method to examine how experienced occupational therapy driver assessors use information to make fitness-to-drive recommendations. *British Journal of Occupational Therapy*. **78**(2), 109–120.

Veloski J. J. (2006). Measuring specific elements of professionalism: empathy, teamwork and lifelong learning. In: D.T. Stern, ed. *Measurement of Professionalism in Medicine.* Oxford, UK: Oxford University Press. pp. 117–45.

Verkerk, M. A., de Bree, M. J. and Mourits, M. J. E. (2007). Reflective professionalism: Interpreting CanMEDS’ “professionalism”. *Journal of Medical Ethics*. **33**(11), 663–666.

de Villiers, M. R., de Villiers, P. J. T. and Kent, A. P. (2005). The Delphi technique in health sciences education research. *Medical Teacher*. **27**(7), 639–643.

Vitek, C. R. *et al.* (2014). Development and initial validation of a project-based rubric to assess the systems-based practice competency of residents in the clinical chemistry rotation of a pathology residency. *Archives of pathology & laboratory medicine*. **138**(6), 809-813.

van der Vleuten, C. P. M. (1996). The assessment of professional competence: Developments, research and practical implications. *Advances in Health Sciences Education*. **1**(1), 41–67.

van der Vleuten, C. P. M. and Schuwirth, L. W. T. (2005). Assessing professional competence: From methods to programmes. *Medical Education*. **39**(3), 309–317.

Wakeling, J. *et al.* (2011). Is foundation training producing competent doctors? What do foundation trainees, educational supervisors and nurses in scotland have to say? *Scottish Medical Journal*. **56(**2), 87–93.

Walsh, A. (2007). An exploration of Biggs’ constructive alignment in the context of work‐based learning. *Assessment & Evaluation in Higher Education*. **32**(1), 79–87.

Wass, V. *et al.* (2001). Assessment of clinical competence. *Lancet*. **357**(9260), 945–9.

Wigton, R. S., Hoellerich, V. L. and Patil, K. D. (1986). How Physicians Use Clinical Information in Diagnosing Pulmonary Embolism: An Application of Conjoint Analysis. *Medical Decision Making*. **6**(1), 2–11.

Wilkinson, J. R. *et al.* (2008). Implementing workplace-based assessment across the medical specialties in the United Kingdom. *Medical education*. **42**(4), 364–73.

Winter, G. (2000). A comparative discussions of the Notion of “validity” in Qualitative and Quantitative Research. *The Qualitative Report*. **4(**3), 1–14.

de Winter, J. C. F. and Dodou, D. (2010). Five-Point Likert Items: t test versus Mann-Whitney-Wilcoxon. *Practical Assessment, Research & Evaluation*. **15**(11), 1–16.

Wood, D. F. (2010). Formative Assessment. In: T. Swanwick, ed. *Understanding Medical Education: Evidence, Theory and Practice*. Oxford, UK: Wiley-Blackwell. pp. 259–270.

Wood, T. J. (2014). Exploring the role of first impressions in rater-based assessments. *Advances in Health Sciences Education*. **3**, 409-427.

Wragg, A. *et al.* (2003). Assessing the performance of specialist registrars. *Clinical Medicine.***3**(2), 131–134.

Yorke, M., Bridges, P. and Woolf, H. (2000). Mark Distributions and Marking Practices in UK Higher Education: Some Challenging Issues. *Active Learning in Higher Education*. **1**(1), 7–27.

Zijlstra-Shaw, S., Roberts, T. E. and Robinson, P. G. (2013). Perceptions of professionalism in dentistry-a qualitative study. *British Dental Journal*. **215**(9), E18-E18.

Zijlstra-Shaw, S., Robinson, P. G. and Roberts, T. (2012). Assessing professionalism within dental education; the need for a definition. *European Journal of Dental Education*. **16**(1), e128-36.