Secondary school performance measures in England – their effectiveness, their effects and an exploration of alternative approaches

Ian Widdows

A thesis submitted in partial fulfilment of the requirements

for the degree of Doctor of Philosophy

The University of Sheffield

Faculty of Social Sciences

Information, Journalism and Communication School

January 2026

Acknowledgements

This research was funded by the Economic and Social Research Council through the White Rose Doctoral Training Programme. I am very grateful for the opportunities this support provided, including an overseas institutional visit to the Autonomous University of Barcelona and an internship in the Department for Education's Strategy Unit, as part of the Behavioural Insights Unit.

I would like to express my gratitude to my supervisors, Professor Jo Bates and Dr Denis Newman-Griffis, for their guidance, insight and support throughout this PhD. Their intellectual challenge, encouragement and generosity with their time have been invaluable. I simply could not have asked for better supervisors – thank you very, very much.

I am grateful to the Department for Education for providing access to key datasets, including the National Pupil Database, the School Workforce Census and the Individualised Learner Record.

My thanks also go to colleagues and staff at the University of Sheffield. In particular, I would like to acknowledge the teaching staff in the Sheffield Methods Institute; the year spent studying with them provided an incredibly strong foundation for this research, and their teaching was consistently inspiring. I am also grateful to the PGR support staff, IT services and library staff for their professionalism and for helping to smooth the many inevitable bumps in the road.

I would like to thank all focus group participants who so generously gave their time and shared their insights. In total, this involved over 27 hours of discussion, with each participant contributing close to five hours. Their openness, creativity and willingness to engage deeply with complex issues were central to this research.

To my gorgeous children, Ella and Gabe - a large part of my motivation for undertaking this PhD was you. Alongside proving something important to myself, I hope this work makes you as proud of me as I am, always, of you.

Finally, but far, far from least, I want to thank the absolute love of my life, my wife Anita. Without her, none of this would have happened. Her inspiration, support, encouragement and love were the initial spark for even considering embarking on a PhD, and they have sustained me throughout. I was once asked by a journalist, when writing a profile of me, which one person I most admire. Without hesitation, I answered "my wife, Anita Fox, the strongest person I know". Nothing has changed - she continues to inspire and amaze me. Thank you for putting up with me "banging on about data", and for everything else. You are such a very special person, and I love you very, very much.

Abstract

Following their introduction over three decades ago, school performance measures (SPMs) now feel like a central component of the education system in England. Generations of teachers, school leaders and school governors have little or no experience of a school system which did not include such performance metrics. For some, they play an essential role in holding schools to account for the important work that they do. For others, the existence of such high-stakes performance measures shapes, or even distorts, many aspects of the work in and around schools.

This project investigates the use, effectiveness and effects of secondary school performance measures in England. Drawing upon quantitative and qualitative methods, including gathering rich data in the form of the opinions of stakeholders, thematic analysis of government documentation and the use of advanced quantitative approaches, it uses theoretical frameworks which draw on Critical Data Studies (CDS), Fairness, Accountability, Transparency and Ethics (FATE) and academic fields related to policy logic. This research examines and evaluates the complex functions of SPMs. Based on the findings of this work, new approaches to the SPM system are proposed.

While others have conducted research in this field, this has been dominated by quantitative and statistical critiques of the specific measures used and their calculation. This work adopts a far more holistic approach, seeking to gain a thorough understanding of the entire data assemblage associated with school performance and accountability.

The outcomes of this project are twofold. First, the production of a PhD thesis will contribute to the understanding of school performance measures in England. Second, new approaches to school accountability are suggested. The considerable experience of the researcher in the education system, the unique, holistic and critical approach of this project and innovative work which seeks to harness new, advanced quantitative approaches mean that the findings of this research have the potential to inform and influence policy in this important area.

Contents

1. Introduction 1

1.1 School Performance Measures 1

1.2 Secondary System in England 3

1.3 Policy Context for School Performance Measures 4

1.4 Research Questions, Aims and Objectives 7

1.5 Researcher Positionality 9

2. Literature Review 12

2.1 Introduction and Scope 12

2.2 Historical and socio-political context of the introduction and evolution of SPMs 13

2.2.1 Neoliberalism 13
2.2.2 The Legacy of the 1988 Education Reform Act 15
2.2.3 New Public Management (NPM) Practices and Audit Culture 16
2.2.4 Policy implications and governance of SPMs 17
2.2.5 Accountability regimes 18
2.2.6 Summary of historical and socio-political context of SPMs 18

2.3 Design and Dissemination of SPMs 19

2.3.1 Types of SPMs and their evolution in England 20
2.3.2 Technical design considerations 21
2.3.3 Data infrastructures and data quality 22
2.3.4 Communication, visualisation and dissemination 22
2.3.5 Summary of SPM design and dissemination 23

2.4 Critiques and Alternative Approaches 24

2.4.1 Implications of design and context for the effectiveness of SPMs 24
2.4.2 Bias in model design 27
2.4.3 Alternative models and reform proposals 28
2.4.4 International comparisons 29
2.4.5 Consequences and unintended effects 30
2.4.6 Sociological critiques of performativity 31
2.4.7 Summary of Critiques and Alternative Approaches 31

2.5 Methodological Approaches in the Literature 33

2.5.1 Quantitative methodological approaches 33
2.5.2 Qualitative methodological approaches 34
2.5.3 Mixed-methods approaches 35
2.5.4 Data ethics methodologies 35
2.5.5 Summary of Methodological Approaches in the Literature 37

2.6 Synthesis of Findings and Identification of Gaps 38

2.6.1 Positioning This Research 40

3. Theoretical-Analytical Frameworks 42

3.1 Critical Data Studies 43

3.2 Fairness, Accountability, Transparency and Ethics (FATE) 50

3.3 Policy Logic Framework 54

3.4 Conclusion to Theoretical-Analytical Frameworks 63

4. Methodology 64

4.1 Introduction to Methodology 64

4.2 Research Philosophy 64
4.2.2 Interpretivism 66
4.2.3 Positivism vs Interpretivism? The 'Paradigm Wars' 67
4.2.4 Moving Beyond the Paradigm Wars 68
4.2.5 Discussion of Research Philosophy 69
4.2.6 Conclusion to Research Philosophy 70

4.3 Research Design 71
4.3.1 Phase 1: Policy Documentation Analysis 72
4.3.2 Phase 2: Quantitative Statistical Modelling 72
4.3.3 Phase 3: Qualitative Focus Groups with Stakeholders 73
4.3.4 Sequential Design 74

4.4 Context of the Study 75

4.5 Data Sampling 76
4.5.1 Phase 1: Policy Documentation Analysis 76
4.5.2 Phase 2: Quantitative Statistical Modelling 80
4.5.3 Phase 3: Qualitative Focus Groups with Stakeholders 83

4.6 Data Collection Methods 87
4.6.1 Phase 1: Policy Documentation Analysis 87
4.6.2 Phase 2: Quantitative Statistical Modelling 87
4.6.3 Phase 3: Qualitative Focus Groups with Stakeholders 87

4.7 Analytical Methods 88
4.7.1 Phase 1: Policy Documentation Analysis 88
4.7.2 Phase 2: Quantitative Statistical Modelling 89
4.7.3 Phase 3: Qualitative Focus Groups with Stakeholders 96
4.7.4 Mixed-Methods Integration 97

4.8 Ethical Considerations 98
4.8.1 Phase 1: Policy Documentation Analysis 98
4.8.2 Phase 2: Quantitative Statistical Modelling 98
4.8.3 Phase 3: Qualitative Focus Groups with Stakeholders 98

4.9 Reliability, Validity and Trustworthiness 100
4.9.1 Reflections on Research Quality and Rigour 100
4.9.2 Phase 1: Policy Documentation Analysis 101
4.9.3 Phase 2: Quantitative Statistical Modelling 101
4.9.4 Phase 3: Qualitative Focus Groups with Stakeholders 102
4.9.5 Mixed-Methods Integration 102

4.10 Limitations of the Methodology 103
4.10.1 Phase 1: Policy Documentation Analysis 103
4.10.2 Phase 2: Quantitative Statistical Modelling 103
4.10.3 Phase 3: Qualitative Focus Groups with Stakeholders 103
4.10.4 Mixed-Methods Integration 104

4.11 Conclusion to Methodology 104

5. Findings from Phase 1: Thematic Analysis of Documents Related to Secondary School Performance Measures in England 105

5.1 Socio-political context (Research Question 1) 107
5.1.1 The continuous influence of 'neoliberalism' 107
5.1.2 The influence of traditional neo-conservative thinking. 117
5.1.3 Increased focus on Social Democracy 120

5.2 Aims and purposes of SPMs (Research Question 2) 123
5.2.1 Purpose of SPMs 123
5.2.2 Methodology of SPMs 126
5.2.3 Limitations of SPMs 127
5.2.4 Interpretation of SPMs 129

5.3 Conclusion to Thematic Analysis of Documents Related to Secondary School Performance Measures in England 130

6. Findings from Phase 2: Analysis of the Structure of SPMs using Quantitative Analysis 133

6.1 Number of factors underlying headline SPMs 135

6.2 Factors underlying SPMs 139
6.2.1 SPM factors ('Perform', 'Attain', 'Progress', 'Curriculum' and 'Destinations') 139

6.3 Correlation analysis 142

6.4 Relationship between the concept of 'performance' and factors underlying wider school information 144
6.4.1 School Information constructs ('School Context' and 'School Character') 144
6.4.2 SEM: Relationships between latent and observed school factors 147

6.5 Multiple regression analysis of school context and school characteristics variables 151
6.5.1 School Context effects 151
6.5.2 School character effects 152
6.5.3 Patterns across headline SPMs 152
6.5.4 Residual structure among headline SPMs 152

6.6 Cluster Analysis 154
6.6.1 Agglomerative hierarchical clustering 154
6.6.2 Summarising the features of the clusters. 156
6.6.3 Visualising the cluster features 159
6.6.4 Evaluating the clusters 162

6.7 Conclusion to Analysing the Structure of SPMs using Quantitative Analysis 164

7. Findings from Phase 3: Thematic analysis of focus group discussions on the subject of school performance measures 166

7.1 Overarching themes from thematic analysis of focus group transcripts 167
7.1.1 Overarching Theme 1: Purposes, Principles and Priorities of SPMs 167
7.1.2 Overarching Theme 2: The Fundamental Flaws of Current Performance Measurement 176
7.1.3 Overarching Theme 3: The Detrimental Impact on Equity and Practice 186
7.1.4 Overarching Theme 4: Towards Holistic, Supportive and Contextualised Accountability 195

7.2 Conclusion to Thematic analysis of focus group discussions on the subject of school performance measures 219

8. Discussion 221

8.1 Introduction ... *221*

8.2 SPM Underpinning Policy Logic ... *222*
8.2.1 Conceptual and Ethical Foundations of SPMs ... 223
8.2.2 Aims and Rationales of SPMs ... 228
8.2.3 Values and Assumptions of SPMs ... 230
8.2.4 Conditions and Mechanisms of the SPM System ... 233
8.2.5 Conclusion to SPM Underpinning Policy Logic ... 237

8.3 SPM Effectiveness ... *238*
8.3.1 SPMs Lack Breadth and Do Not Reflect Whole School Effectiveness ... 238
8.3.2 Questioning SPM Fairness and Bias ... 245
8.3.3 Communication and Interpretation of SPMs ... 251
8.3.4 Wider Implications for the School System ... 252
8.3.5 Conclusion to SPM effectiveness ... 253

8.4 SPM effects ... *257*
8.4.1 Micro level (typically Professional/Behavioural) ... 258
8.4.2 Meso level (typically Institutional/Organisational) ... 262
8.4.3 Macro level (typically Systemic/Political) ... 267
8.4.4 Conclusion to SPM effects ... 270

8.5 Alternative Approaches ... *273*
8.5.1 Moving Beyond Academic Metrics to Provide a More Holistic Picture of School Quality ... 274
8.5.2 Contextualised Data to Allow Fairer Comparisons and to Address Bias ... 280
8.5.3 Reconceptualise SPMs as Communicative Tools to Meet Different User Needs and Support Interpretation and Understanding ... 282
8.5.4 Reconfiguring Accountability to a More Supportive System, Built on Improved Trust ... 284

8.6 Future directions for SPMs ... *285*

8.7 Conclusion to Discussion ... *285*

9. Conclusion ... 288

9.1 Overview of the Research ... *288*

9.2 Empirical Contribution ... *289*

9.3 Theoretical Contribution ... *291*

9.4 Recommendations for Policy Makers ... *292*

9.5 Future Research ... *294*

9.6 Concluding Remarks ... *294*

10. References ... 295

Appendix 1. Coding frequency within documents ... 334

Socio-political context (Research Question 1) ... *334*
Neoliberalism ... 334
Neoconservatism ... 335
Social Democracy ... 335

Aims and purposes of SPMs (Research Question 2) ... *336*

Appendix 2. Timeline to show political context of production of documents. ... 337

Appendix 3. Details of data elements used in quantitative analyses.....340

Headline Secondary School Performance Measures 340

Wider School Context and Intake Measures..... 341

Appendix 4. SEM models 342

Appendix 5. Details of School Clusters.....344

Cluster 1 (n=1415) 344

Cluster 2 (n = 1292) 345

Cluster 3 (n = 255) 346

Cluster 4 (n = 163) 347

Appendix 6. Focus Group 3 information (alternative approaches to SPMs).....348

Appendix 7. Advertisement for participants 378

Appendix 8. Ethics Approval Letter 379

List of figures

Figure 1. Overview of research design 71
Figure 2. Parallel Analysis Scree Plot for dataset of 13 variables (6 headline SPMs and 7 school information metrics) for 3125 mainstream state secondary schools in England, academic year 2021-2022 90
Figure 3. Frequency per 10000 words of the terms 'freedom' or 'autonomy' 113
Figure 4. Parallel Analysis Scree Plot for dataset of 6 headline SPMs for 3125 mainstream state secondary schools in England, academic year 2021-2022. 135
Figure 5. Factor Analysis Diagram showing factor loadings 137
Figure 6. Model diagram for Initial SPM model 139
Figure 7. Model diagram for Final SPM model 141
Figure 8. Correlation matrix of all current headline SPMs. 142
Figure 9. Model diagram for Initial School Information Model. 144
Figure 10. Model diagram for Final School Information Model. 145
Figure 11. Initial SPM Model – initial SEM model incorporating all four SPM factors as responding to both 'School Character' and 'School Context' factors and 'School Context' responding to 'School Character'. 147
Figure 12. Refined SPM Model – SEM model incorporating all four SPM factors as responding to 'School Context' and only 'Destinations', 'Curriculum' and 'School Context' responding to 'School Character'. 148
Figure 13. Final SPM Model for all SPM latent variables - incorporating latent variables of 'School Character' and 'School Context'' and all four SPM latent variables ('Attainment', 'Destinations', 'Curriculum' and 'Progress'). 149
Figure 14. Final SPM Model for only the 'Attainment' SPM latent variable - incorporating latent variables of 'School Character' and 'School Context' 150
Figure 15. Parallel scree plot. 155
Figure 16. Percentage of variance explained by each dimension. 155
Figure 17. Average silhouette width plot - indicating optimal number of clusters to be 4. 156
Figure 18. Boxplots to Show Distribution of Variables by Cluster. 159
Figure 19. Stacked Bar Charts to Show Distribution of Variables by Cluster. 160
Figure 20. Boxplots of the distributions of mean scaled values of the 6 SPM variables for clusters 1 to 4. 160

Figure 21. Radar plot of the mean scaled values for each of the thirteen variables for Cluster 1 -4 formed using hierarchical clustering of the seven school information variables. 161
Figure 22. Silhouette Plot for agglomerative hierarchical clustering using Ward's minimum variance method for all English state schools and based on six variables of 'wider school information'. 162
Figure 23. t_SNE Visualisation for agglomerative hierarchical clustering using Ward's minimum variance method for all English state schools and based on 6 variables of 'wider school information'. 163
Figure 24. Model diagram for Initial School Information Model. 342
Figure 25. Model diagram for first Refined School Information Model..... 342
Figure 26. Model diagram for second Refined School Information Model. 343
Figure 27. Model diagram for Final School Information Model..... 343

List of tables

Table 1. Summary of Themes in Critical Data Studies Framework 47
Table 2. FATE Framework for School Performance Measures (SPMs)..... 52
Table 3. Policy Logic Framework for School Performance Measures (SPMs) 57
Table 4. Sub-framework: Pragmatic Conditions for Neoliberal Policy Operation 61
Table 5. Documents related to SPM Methodology (n = 7)..... 76
Table 6. Documents related to SPM Advice/Information (n = 4)..... 77
Table 7. Documents related to SPM Legislation (n = 8) 77
Table 8. Government Education White Papers, Green Papers and Reports relating to SPMs (n = 20) 78
Table 9. Headline performance measures 80
Table 10. Wider school information metrics..... 80
Table 11. Description of focus group participants 84
Table 12. Participants included in each focus group..... 86
Table 13. Eigenvalues of the principal components arising from Principal Component Analysis (PCA) of the study dataset, indicating 3 principal components with eigenvalues greater than 1 (in green) 91
Table 14. Factor loadings for 4 factor EFA using 6 headline SPM variables..... 92
Table 15. Model fit indices used in CFA/SEM analysis and development of models..... 93
Table 16. Eigenvalues of the factor arising from Exploratory Factor Analysis of the study dataset, indicating 3 factors with eigenvalues greater than 1 (in green) 136
Table 17. Factor loadings for the 3-factor EFA using the 6 headline SPM variables..... 137
Table 18. Model fit indices for Initial SPM Model..... 139
Table 19. Model fit indices for Final SPM Model 141
Table 20. Correlation between Attainment 8, EBacc APS and 'Basics' SPMs. 143
Table 21. Model fit indices for Initial School Information Model 145
Table 22. Model fit indices for Final School Information Model..... 145
Table 23. Model fit indices for initial SPM Model 1. 147
Table 24. Model fit indices for refined SPM Model. 148
Table 25. Model fit indices for final SPM Model for all SPM latent variables..... 149
Table 26. Model fit indices for final SPM Model for only the 'Attainment' SPM latent variable 150
Table 27. Standardised regression coefficients (ß) for SPMs..... 151
Table 28. Key residual correlations among headline SPMs (SEM model)..... 153

Table 29. Number of schools in each cluster. 156
Table 30. Mean value of each variable for clusters 1 to 4 (overall mean is micro-averaged, i.e. it is calculated over schools rather than clusters and so is weighted to incorporate the different number of schools in each cluster). 157
Table 31. Summary of each variable for clusters 1 to 4. 158
Table 32. Model diagram for Initial School Information Model. 342
Table 33. Model fit indices for first Refined School Information Model. 342
Table 34. Model fit indices for second Refined School Information Model. 343
Table 35. Model fit indices for Final School Information Model. 343

Declaration

I, the author, confirm that the thesis is my own work. I am aware of the University's Guidance on the Use of Unfair Means (www.sheffield.ac.uk/ssid/unfair-means). This work has not previously been presented for an award at this, or any other, university.

1. Introduction

1.1 School Performance Measures

In 1992, the UK Government first published performance metrics for each state-funded secondary school in England, initiating what became commonly known as the school 'league tables' (Leckie & Goldstein, 2017a). These performance tables have their roots in the Education Reform Acts of 1988 and 1992, reforms that introduced the National Curriculum, standardised testing and market-style mechanisms into the education system (Flude & Hammer, 1990; Lawton, 1989; Wragg, 1988). From the outset, these metrics also supplied data for inspection by the Office for Standards in Education (Ofsted), embedding performance reporting into the wider school accountability architecture.

Over the ensuing three decades, the way school performance has been measured has changed substantially, driven by shifts in qualification structures, changing policy aims and methodological debates. Early league tables reported raw attainment: for example, the proportion of pupils achieving 'good' GCSE grades. In the early 2000s, 'value-added' measures—designed to capture students' progress relative to their prior attainment—were introduced (UK Statistics Authority, 2015). By the mid-2000s, value-added was replaced by Contextual Value Added (CVA), which adjusted for background characteristics such as free school meals, ethnicity, gender, special educational needs and cohort mobility (Leckie & Goldstein, 2017a). CVA was later dropped (around 2011) in favour of 'expected progress' (2011–2015), before being superseded by the current Progress 8 and Attainment 8 system in 2016 (Department for Education, 2014; Leckie & Goldstein, 2017a).

Attainment 8 captures the average attainment of a student cohort across eight specified subject slots, while Progress 8 aims to assess the progress that pupils make from the end of primary (KS2) to the end of secondary (KS4), benchmarked against national peers with similar KS2 prior attainment (Department for Education, 2025a). These measures reflect a significant conceptual shift: away from purely raw outcomes and towards value-added progress — yet without the heavy contextual adjustment that CVA once provided (Leckie & Goldstein, 2017a).

Crucially, the removal of CVA sparked debate about the implications of deliberately removing the contextual element of SPMs. Since the justification for this move was to maintain high expectations for all students, irrespective of their background (Department for Education, 2010; Leckie & Goldstein, 2017a), it reflects a worldview that all students effectively have the same opportunities to succeed academically, thereby dismissing the concept of socio-economic structural barriers as

hindrances to academic achievement. I will return to the arguments around contextualising SPMs in sections 7.1.2.3, 8.3.1.2 and 8.5.2. Progress 8, while more straightforward, fails to account for socio-structural inequalities, because it does not adjust for pupil background in the same way as CVA (Education Policy Institute, 2025). Furthermore, value-added measures are not without their statistical challenges: the reliability of these scores can be affected by factors such as small cohort sizes, outliers and volatility of progress over time (FFT Education Datalab, 2018; Prior et al., 2020).

Moreover, there is a growing argument for reframing accountability to reflect a broader conception of school effectiveness. For instance, some researchers propose multivariate value-added models that incorporate non-academic outcomes — like attendance, behaviour or exclusion — alongside academic attainment (Prior et al., 2020). Such models suggest a more holistic and equitable view of what schools contribute to students' lives.

Overall, the evolution of performance measures in England reflects tensions between simplicity and fairness, transparency and equity and statistical precision and policy practicality. Understanding this history is not just a matter of historical detail — it is essential to critically assess whether current measures serve their intended purposes and to envision alternative accountability frameworks that might better support schools and students in a changing educational landscape.

1.2 Secondary System in England

The secondary system in England typically serves pupils aged 11 to 16, with the majority of schools also offering post-16 provision for students up to age 18. Secondary education begins at Key Stage 3 (ages 11–14) and continues through Key Stage 4 (ages 14–16), culminating in national General Certificate of Secondary Education (GCSE) examinations and alternative vocational qualifications (Department for Education, 2024). Although statutory school-leaving age ends at 16, youth must remain in some form of education or accredited training until age 18. Post-16 pathways include academic A-levels, or vocational tracks such as BTECs and T Levels, hosted within school sixth forms or further education colleges (Department for Education, 2024).

English secondary schools operate within a highly decentralised and diverse institutional landscape. State-funded schools include local authority-maintained schools, academies and free schools, alongside a smaller independent (fee-paying) sector. Academies and free schools—which are publicly funded but operate independently of local government control—have come to dominate the landscape, educating over 82% of secondary pupils in England, often under the structural governance of Multi-Academy Trusts (Department for Education, 2025d). Furthermore, admissions structures vary regionally; while most areas utilise a comprehensive framework, some regions retain a selective system split between grammar schools (selective entry via examination) and secondary modern schools (non-selective entry) (Machin & Vignoles, 2018). Approximately 20% (163 grammar schools and approximately 450 secondary modern schools) of state maintained secondary schools are affected by academic selection, alongside a nominal group of roughly 35 partially selective schools that admit a minor percentage of pupils based on specific academic or artistic aptitude (Department for Education, 2025d; Machin & Vignoles, 2018).

Assessment and institutional accountability exert significant pressure on the English system. School performance is strictly monitored via national examination data—specifically through headline accountability metrics like 'Attainment 8' and 'Progress 8'—as well as school inspection outcomes published by the Office for Standards in Education, Children's Services and Skills (Ofsted) (Department for Education, 2025a; Ofsted, 2023). These structural mechanisms heavily dictate school priorities, curriculum choices and resource allocation, offering vital context for interpreting the findings in this study.

1.3 Policy Context for School Performance Measures

We find ourselves at a significant inflection point in education policy, shaped by political transition, renewed scrutiny of school accountability and broader social and economic pressures. The immediate post-pandemic period - initially characterised by concerns about 'unfinished learning,' recovery funding and cohort disruption - has evolved into a longer-term reflection on the resilience, fairness and purpose of England's education system. While the acute phase of COVID-19 has passed, its consequences continue to reverberate: persistent absence has reached historically high levels (Department for Education, 2025c), attainment gaps remain wider than pre-2020 (Education Policy Institute, 2023) and the emotional, behavioural and wellbeing needs of young people have intensified (NHS England, 2023). These continuing challenges have reframed debates about school performance, the adequacy of current measures and the expectations placed on schools.

Politically, the landscape has shifted considerably. The 2024 general election brought a change of government and a declared departure from the market-oriented, high-stakes accountability philosophy that has shaped English schooling for over three decades. The new administration has emphasised trust, professionalism and systemic improvement, signalling a desire to rebalance accountability with support. Central to this is the Curriculum and Assessment Review (CAR), which proposes wide-ranging reforms to curriculum breadth, assessment models and the principles underpinning school evaluation (Department for Education, 2025b). The CAR has proposed substantial changes to curriculum design and qualifications. Notably, the Review did not recommend removing Progress 8; instead, it advised retaining both Progress 8 and Attainment 8 as the core accountability measures (Department for Education, 2025b). The Review's major recommendation in this area was the removal of EBacc (%Ebacc) entry and attainment (Ebacc APS) as headline accountability measures. The CAR recommendation to scrap these measures was on the grounds that they distort curriculum breadth and contribute to a narrowing of Key Stage 4 provision.

However, the Department for Education (DfE) has not accepted the CAR recommendations in full. Instead of maintaining Progress 8 unchanged, the DfE has set out its intention to consult on structural reforms to Progress 8, focused on altering the subject "buckets." These proposals would reduce the EBacc-driven constraints by giving schools greater flexibility over which subjects can count, while still retaining a progress-based measure centred on English and mathematics (Department for Education, 2025e). The DfE has stated that any changes resulting from this consultation would apply for the first time in 2029, allowing schools to take the revised measure into account when determining GCSE subject choices for students in 2027 (Schools Week, 2025).

Simultaneously, Ofsted has embarked on its most substantial reform programme in a generation, informed by its national consultation - The Big Listen (Ofsted, 2024a). This review reopened debates about the legitimacy of the inspectorate, the emotional impact of inspection on staff and the appropriateness of single-word judgments. From 2025, Ofsted have replaced single overall inspection grades with multi-category Report Cards, grading schools separately across key areas such as curriculum, inclusion, behaviour, achievement and leadership. The reforms aim to reduce the high-stakes impact of one-word judgements and provide a more nuanced, contextual picture of school quality. Together, these shifts create an unprecedented moment of potential policy fluidity since the introduction of league tables in the 1990s.

Alongside these national developments, global events continue to shape domestic conditions. International economic uncertainty, armed conflict and pressures on energy and food security have driven inflation and reduced public spending power, directly affecting school budgets and staffing. England's teacher recruitment and retention difficulties—described by the National Foundation for Educational Research (NFER) as a "crisis" in its 2023 annual report (Worth et al., 2023) - interact with accountability pressures, shaping the real capacity of schools to deliver high-quality provision. Schools are therefore working within increasingly complex constraints and these contextual pressures have renewed attention on the validity of performance measures that do not fully account for them.

The pandemic period also transformed public engagement with data. Daily dashboards, modelling and statistical briefings brought terms such as 'R number,' 'rolling averages,' and 'exponential growth' into common discourse (European Centre for Disease Prevention and Control, n.d.; UK Health Security Agency, n.d.). At the same time, public debate highlighted shortcomings in the communication of statistical information. Professor David Spiegelhalter, for example, criticised aspects of government data presentation as "embarrassing, context-free number theatre" (BBC, 2020). Subsequent parliamentary reviews emphasised the need for transparent, high-quality statistical communication aligned with the Government Statistical Service Code of Practice (Office for Statistics Regulation, 2022).

This heightened visibility and contestation of data have important implications for how school performance information is understood. Accountability measures such as Progress 8, Attainment 8, EBacc entry rates and destination data fundamentally shape public perception of schools, yet they embody methodological assumptions that are often unclear to non-specialists. In 2022, the Office for Statistics Regulation conducted a full assessment of Key Stage 4 statistics and the Compare School Performance website, requiring improvements in clarity, interpretability and communication (Office

for Statistics Regulation, 2022). These concerns were further amplified by the impact of COVID-19 on assessment cohorts, leading the DfE to halt some performance measures during 2020 and 2021 and subsequently to emphasise that 2022 and 2023 results should be interpreted with caution (Office for Statistics Regulation, 2022).

At the core of these debates lies a broader recognition that school performance measures are profoundly shaped by factors beyond schools' control—including prior attainment, socio-economic context, special educational needs, ethnicity, language, neighbourhood characteristics and funding inequalities. The uneven distribution of these influences across England's schools raises critical questions about fairness and validity. The DfE's own 2022 statement acknowledged explicitly that COVID-19 affected schools differently and that published results should be considered only as "a starting point for discussion" (Department for Education, 2022a). Yet, as many researchers have argued, this principle is equally relevant to the wider structural and demographic contexts that shape student outcomes under normal conditions (Prior et al., 2023).

In this context of political change, heightened data awareness and renewed attention to educational inequalities, there is an important opportunity to re-examine the role, purpose and impact of secondary school performance measures. My research seeks to contribute to this emerging conversation by critically assessing the effectiveness, effects and future possibilities of these measures, drawing on both quantitative evidence and qualitative insights into how they shape practice, policy and public understanding.

1.4 Research Questions, Aims and Objectives

In my research, I explore the various aspects of the production, publication and impact of the school league tables as they apply to secondary schools in England. This will be achieved by conducting a series of three related research phases. The findings from these phases will then be integrated to provide a deeper understanding of the production, publication and impact of SPMs.

First, I aim to investigate the ideas underpinning the use of SPMs, which are, at least in part, reflected in their published aims and principles. Using a qualitative approach, I will gather and thematically analyse rich data from various government documents related to SPMs. My goal is to gain insight into the socio-political context surrounding the introduction and evolution of SPMs and to better understand their intended purposes.

Second, I will use a range of quantitative approaches to examine the breadth and meaning of the concept of 'performance' as represented in SPMs.

In the third phase, I will conduct a series of focus groups with participants made up of school stakeholders. These focus groups will examine, drawing on a variety of different stakeholder standpoints and perspectives, the purpose of SPMs, their effectiveness and their wider effects. They will then consider directions for alternative approaches to SPMs to begin to address issues identified in the other focus group stages.

The overarching research question for this thesis is:

How well do secondary school performance measures in England serve their stated purposes and what does empirical analysis reveal about how they might be improved?

This is further broken down into 3 research questions:

Research Question 1. What is the socio-political context of the production and publication of secondary school performance measures in England?

Research Question 2. To what extent is the production, publication and operationalisation of secondary school performance measures in England effective in meeting their stated aims, in this context?

Research Question 3. How can the findings of research and advances in technology be utilised to improve the communication and use of school data by stakeholders for school performance measurement purposes?

In each of the three research phases comprising this work, I have also developed sub-research questions that target the specific aims of each phase, to sharpen my focus on the insights sought from each.

The objectives of this study are:

- Evaluate the effectiveness of the production and publication of school accountability measures against their explicit and implicit aims.
- Examine how the published aims of English SPMs translate into practice and assess the extent to which current measures achieve these objectives.
- Critique the headline measures used and analyse how the information is published and operationalised.
- Investigate the role of school performance data in shaping, and being shaped by, multiple dimensions of the secondary school system, drawing on data gathered from a range of sources, including stakeholder focus groups, international comparisons and suggestions for alternative models.
- Appraise the consequences of SPM publication and use within schools and across the wider community.
- Develop critical theoretical-analytical frameworks for understanding the complex SPM data assemblage, using quantitative and qualitative methods alongside Critical Data Studies (CDS), FATE and Policy Logic research.

1.5 Researcher Positionality

My professional background in education spans over 28 years in secondary schools, including 14 years in senior leadership and 5 years as Headteacher of a large 11–18 secondary school. During this time, I gained direct experience of the complexities of school leadership and the practical implications of educational policies.

Through my leadership roles, I engaged extensively with school performance data and the school accountability system, developing a critical perspective on how data are used to represent school effectiveness. My work in establishing the National Association of Secondary Moderns (NASM) further shaped my positionality by focusing attention on non-selective schools in selective areas, highlighting challenges linked to school context, student intake and socio-economic factors. This experience has fostered a particular sensitivity to the ways in which contextual factors intersect with policy and performance metrics and the implications these have for equity and fairness in the education system.

In addition, my relationships with other school leaders, government officials and education researchers have provided a broad view of the educational landscape, but they also situate me within professional networks that influence my perspectives and interpretations. These experiences inform my understanding of school performance measures and the effects they have on schools, educators, students and communities, while acknowledging that my positionality shapes the questions I ask and the interpretations I make in this research.

Thesis Structure

This thesis is organised into nine chapters. The content of each chapter is as follows:

1. **Introduction**

This chapter sets out the background to the study, the research problem, aims and questions and the rationale for examining secondary SPMs in England. It also outlines the overall research design and structure of the thesis.

2. **Literature Review**

This chapter reviews relevant national and international literature on SPMs, accountability and school effectiveness, situating the study within existing empirical and theoretical debates and identifying key gaps the research addresses.

3. **Theoretical-Analytical Frameworks**

This chapter introduces the theoretical perspectives and analytical frameworks that underpin the study, explaining how they inform the conceptualisation of performance, accountability and the interpretation of SPMs.

4. **Methodology**

This chapter details the methodological approach adopted in the study, including the research design, data sources, analytical methods and ethical considerations, with particular attention to the mixed-methods, multi-phase structure of the research.

5. **Findings from Phase 1: Thematic Analysis of Documents Related to Secondary School Performance Measures in England**

This chapter presents findings from a thematic analysis of key policy and system-level documents related to secondary SPMs in England, examining their stated purposes, assumptions and theories of change.

6. **Findings from Phase 2: Analysis of the Structure of SPMSSPMs using Quantitative Analysis**

This chapter reports findings from the quantitative analysis of secondary SPMs, exploring their structure, interrelationships and the extent to which they reflect contextual versus school-level factors.

7. **Findings from Phase 3: Thematic analysis of focus group discussions on the subject of school performance measures**

This chapter presents findings from thematic analysis of focus group discussions with stakeholders, capturing perspectives on the use, interpretation and consequences of SPMs in practice.

8. Discussion

This chapter integrates findings from all three phases of the study, relating them back to the literature and theoretical frameworks, and critically examining their implications for understanding performance, accountability and effectiveness in secondary education.

9. Conclusion

This chapter summarises the key contributions of the study, reflects on its limitations and outlines implications for policy, practice and future research on SPMs.

2. Literature Review

2.1 Introduction and Scope

School performance measures (SPMs) have become a key feature of the education landscape in England and an increasingly important policy tool worldwide. Introduced to meet multiple and often conflicting objectives, SPMs aim to hold schools accountable, guide parental choice, inform inspection and interventions and promote system-wide school improvement (Grek, 2009; Ozga, 2009). Over time, they have developed from simple raw attainment indicators to far more sophisticated statistical models such as value-added approaches and Progress 8 (Leckie & Goldstein, 2017a). Despite their widespread presence, ongoing debates focus on their validity, fairness and the social, organisational and behavioural effects they produce for schools, teachers and students.

This literature review synthesises significant scholarship strands relating to the design, use and effects of SPMs. It addresses:

a) the historical and socio-political context of the development of SPMs,

b) the design and dissemination practices of SPMs,

c) critiques and alternative approaches,

d) methodological traditions used to investigate SPMs.

Across this literature, persistent tensions emerge concerning statistical validity, social equity, data governance and the political work that performance metrics perform. The review concludes by identifying key conceptual and empirical gaps of particular relevance to this study, especially the influence of school context on headline performance metrics and how they might be made more fair, the conceptualisation of 'performance' within SPMs, the communication of SPMs and the potential to reconfigure school accountability to a more supportive system.

To understand the debates surrounding the validity, fairness and consequences of SPMs, it is essential first to situate them within the broader historical and socio-political transformations that shaped their introduction and evolution. SPMs in England did not arise in a vacuum; they were constructed within a changing governance landscape marked by marketisation, managerialism and shifts in the role of the state and government priorities. The next section, therefore, examines the historical and socio-political foundations of SPMs, tracing the evolution of SPMs from the 1980s onwards and identifying the logics that continue to underpin current metrics.

2.2 Historical and socio-political context of the introduction and evolution of SPMs

A combination of ideological, economic and administrative shifts, including marketisation, managerialism and the rise of audit culture, created the policy conditions in which comparative performance data in the form of SPMs became central to the schools' system. The move towards a high-stakes accountability system was neither inevitable nor purely technical; instead, it reflected changing assumptions, shaped by political ideology, about the role of the state, the nature of public sector professionalism and the mechanisms expected to drive educational improvement.

These transformations were shaped by wider debates about quasi-markets in public services (Glennerster et al., 1991; Le Grand, 2007), the rise of performativity as a mode of governing professionals (Ball, 2003b) and the increasing reliance on data infrastructures to steer public sector organisations (Ozga, 2009). The emergence of what Power (1997) described as the "audit society" further normalised the expectation that schools would be continually monitored, compared and held accountable through quantified measures. This section traces the emergence and evolution of SPMs across these historical contexts, beginning with Neoliberalism and the 1988 Education Reform Act before exploring the rise of New Public Management (NPM) and the consolidation of market- and performance-based accountability regimes.

2.2.1 Neoliberalism

The conceptual foundations of neoliberalism emerged not as a simple continuation of nineteenth-century liberalism, but as a calculated intellectual response to the crises of the 1930s and the post-war rise of Keynesian economics and collectivism (Harvey, 2007; Pollitt, 2007a). Its formal origins are often traced to the Colloque Walter Lippmann in Paris in 1938, where thinkers debated how to revitalise liberalism to account for market failures while firmly rejecting totalitarian state control. This movement was further institutionalised with the formation of the Mont Pèlerin Society in 1947, bringing together key figures such as Friedrich Hayek, Milton Friedman and Ludwig von Mises, who argued that competitive markets were the only mechanism capable of processing complex information and preserving individual liberty (Eagleton-Pierce, 2016). Rooted in classical liberal economic thought, neoliberalism advocates for the primacy of markets, deregulation, privatization and a reduced role for the state in economic planning (Brenner et al., 2010). This set of ideas remained on the intellectual periphery until the crises of stagflation and the perceived failures of welfare states in the 1970s provided the political opening for its policy agenda to become dominant,

leading to its implementation by political leaders like Margaret Thatcher and Ronald Reagan in the 1980s (Hursh, 2000; Mirowski & Plehwe, 2009; Shermer, 2014).

There is currently no single, universally accepted definition of neoliberalism. It is a term with multiple meanings depending on the context in which it is used, a feature Matthew Eagleton-Pierce describes as a kind of "conceptual Swiss Army Knife" (2016, p. xiii). These meanings have also evolved (Shermer, 2014). While some commentators may emphasise neoliberal ideas and present it as an overarching ideology (Turner, 2007), others examine the policies and practices of neoliberalism, treating it as a set of 'rules' within which capitalism has operated over the last three decades (Monbiot, 2016). Moreover, some might even argue that neoliberalism is not a fixed 'thing' and that its ideas are more associated with processes and as such it might be referred to as 'neoliberalisation' (Maisuria, 2014). However, some core ideas remain that are generally associated with neoliberal ideology. These are the principles that underpin neoliberal thought and have influenced neoliberal policies over the last three decades. These underlying principles are what Eagleton-Pierce (2016, p. xiv) refers to as denoting "a set of policy tendencies" to what is "often a moving and ill-defined target":

- expansion of commercial markets
- privileging of corporations
- re-engineering of government as an entrepreneurial actor
- imposition of fiscal discipline

Central to the neoliberal project is the commitment to market forces as the preferred mechanism for both driving efficiency and improving standards (Hursh, 2000). It is beyond the scope of this project to conduct a detailed examination of the emergence of neoliberal thought and the related ideas of economic rationalism, which played a key role in normalising neoliberal principles. However, it is important to recognise the rise of New Public Management (NPM) practices and their role within the context of SPMs. Emerging during the 1980s, alongside or as a manifestation of neoliberal ideology (McDonald-Kerr & Boyce, 2024), NPM practices have helped establish market logic within England's school system. Neoliberal ideology and NPM practices, with their managerialist aims and focus on market-centric ideas, share many similarities (Knafo et al., 2019). According to Pollitt (2007b), the key aspects of NPM are:

- management (including private-sector management approaches and managemental/directorial decentralisation and freedom),
- increased focus on 'performance' standards and target-setting based around the measurement of output,

- competition and the role of market-type mechanisms in driving higher standards while maintaining financial efficiency.

Because of the shared core concepts, for this analysis, ideas related to NPM can be regarded as part of the expression of the neoliberal project and discussed within the framework of 'neoliberalism'. Importantly for this project, NPM practices, as a characteristic of neoliberal influence on education policy, can be seen as having created the conditions for the implementation of school accountability systems, including the publication of SPMs (McDonald-Kerr & Boyce, 2024). I will revisit this concept in more detail in section 2.2.3 below.

2.2.2 The Legacy of the 1988 Education Reform Act

The 1988 Education Reform Act (ERA) marked a defining moment in the restructuring of schooling in England, embedding a new architecture of marketisation, managerialism and accountability. The ERA introduced the National Curriculum, standardised testing at multiple key stages, open enrolment and delegated budgets, alongside the publication of performance information intended to support parental choice (Gorard, 1997; Whitty et al., 1998). These reforms built on earlier arguments about introducing market mechanisms into state schooling (Ball, 1993; Glennerster et al., 1991) and reflected broader New Right ideologies emphasising competition, consumer choice and the belief that market pressures would drive up educational standards (Gewirtz et al., 1997; Le Grand, 2007; Tooley, 1995).

Research from the 1990s and 2000s has shown that these quasi-market reforms reshaped relationships between schools, parents and the state. The foundational work of Gewirtz et al. (1997) demonstrate empirically that the capacity to exercise choice is socially stratified, with middle-class parents disproportionately benefiting from the publication of SPMs and league tables, which acted as signals reinforcing social segregation. These findings align with the broader hypothesis that quasi-markets tend to reproduce, rather than mitigate, social inequalities (Ball, 1993; Whitty, 1997).

A substantial empirical literature has examined how competitive pressures interact with school intake, funding and local demographics. Levačić (2000), drawing on the ESRC-funded ICOSS study of 300 schools, identified complex interactions between competition and socio-economic context. Later work by Levačić and Woods (2002; 2013) highlighted how schools serving disadvantaged communities were often placed at a competitive disadvantage. The publication of performance data contributed to recursive patterns: schools with socially advantaged intakes attracted more pupils, reinforcing reputational and resource differences. Levačić (2004) further explored how competitive conduct relates to perceived competition and the structural features of local markets, contributing

to a broader evidence base, including US studies synthesised by Belfield and Levin (2002), which found only modest positive correlations between competition and educational quality.

Bradley and Taylor (2002) brought these strands together, assessing the effects of quasi-markets on both efficiency and equity. They identified small efficiency gains but substantial increases in social segregation. However, their conclusions must be read in light of their use of the percentage of pupils achieving 5 A*–C GCSEs—a dominant SPM at the time but one strongly influenced by student intake (Goldstein & Thomas, 2008; Gorard, 2010)—and a staff productivity proxy based on the same metric, raising questions about the robustness and contemporary relevance of their findings.

More recent research finds similar patterns: modest performance improvements accompanied by increased inequality, including in systems outside the English state schools system such as Quebec (Kamanzi, 2022), Vietnam (Nguyen et al., 2021), higher education (Mizrahi, 2021) and the independent sector (López-Torres et al., 2021). Connell (2013) and Hursh (2007) interpret these developments as part of a globalised neoliberal restructuring of education, foregrounding competition, commodification and high-stakes accountability.

Understanding quasi-markets also requires attention to parental decision-making. Parents value not only academic performance but also safety, pastoral care, communication, reputation and extracurricular opportunities (Aldridge, 2021; Filer & Sinozich, 2021; OECD, 2016). These broader priorities contrast with the narrow academic focus of current SPMs. West, Mattei and Roberts (2011) provide a framework for analysing England's accountability regime, while Acquah (2013) draws on Pring (2012) to challenge the philosophical foundations of market-based accountability.

In her book "The Best for my Child" (Millar, 2018), Guardian Education columnist Fiona Millar looks at the principles of diversity, choice and competition within the English schools' system and the effect of these on the fairness and quality in education provision. This work raises some important questions about the reality of choice in a school place allocation system which, while allowing parental preference, is managed by admission authorities in line with the School Admissions Code (Department for Education, 2021) and is influenced by other factors, including, for example, the local housing market.

2.2.3 New Public Management (NPM) Practices and Audit Culture

The reforms of the late 1980s and 1990s were part of a broader shift toward New Public Management (NPM), which promoted output measurement, performance targets, cost-efficiency

and the managerial control of public services (Hood, 1995). NPM introduced a managerial rationality premised on measurable outputs and standardised indicators—directly shaping the development and use of SPMs.

Hood (1995) systematised the core doctrines of NPM, including disaggregation, competition and performance measurement. In schooling, this translated into an intensified focus on Key Performance Indicators (KPIs), target-setting and performance comparisons (Schedler & Proeller, 2000). These changes were accompanied by the rise of managerialism and internal monitoring systems intended to make schools accountable to external evaluators.

Critiques of KPI-driven public management emerged early. Smith (1995) warned of "goal displacement" and "measure fixation"; Bevan & Hood (2006) later show how high-stakes targets encourage gaming, manipulation and strategic behaviour. In education, these dynamics have been widely documented, including curriculum narrowing, focus on borderline students and strategic data handling (Gillborn & Youdell, 2000; Perryman, 2009).

The rise of audit culture further reinforced these trends. Power's (1997) concept of the "Audit Society" highlights how rituals of verification become ends in themselves, remaking organisational priorities. Ozga (2009) argues that education has become governed "through data," embedding datafication and surveillance into everyday practice. Ball (2003b) extends this analysis by showing how performativity reshapes teachers' identities, generating emotional labour, compliance pressures and a culture of constant self-monitoring. These developments form a crucial backdrop for understanding how SPMs evolved not only as technical tools but also as instruments of governance and disciplinary power.

2.2.4 Policy implications and governance of SPMs

The policy logics underpinning SPMs draw on broader theories of educational accountability. Elmore (2004) and Wei et al. (2009) distinguish between internal accountability (professional norms, capacity, coherence) and external accountability (targets, sanctions, inspections), arguing that sustainable improvement depends on aligning the two. In England, however, accountability has been dominated by external mechanisms, in the form of performance tables and inspections, placing significant pressure on schools' organisational practices and priorities (Lingard & Sellar, 2013).

Governance arrangements have oscillated between centralisation and decentralisation. While the ERA promoted decentralised decision-making at school level, accountability frameworks remained highly centralised, with the Department for Education (DfE) and Ofsted exerting substantial control over the definition, measurement and publication of school performance (Baxter & Clarke, 2013;

Leithwood & Earl, 2000). This hybrid model—decentralised responsibility combined with centralised oversight—has been described as characteristic of neoliberal governance regimes (Ball, 2003b, 2012).

Policy critiques argue that high-stakes SPMs distort organisational behaviour, undermine collaboration and may fail to support genuine school improvement (Wiliam, 2018). In parallel, researchers have identified emerging interest in alternative governance logics, including group-level accountability (e.g., within Multi-Academy Trusts), collective responsibility and network-based models that shift the locus of performance beyond individual schools (Greany & Higham, 2018).

2.2.5 Accountability regimes

The evolution of SPMs is closely tied to the rise of "high-stakes accountability" (Hamilton et al., 2002, p.28), characterised by the publication of performance tables, the embedding of performance data within inspection frameworks and the linking of outcomes to sanctions or rewards. English schools operate within a densely woven accountability regime in which Ofsted inspections, performance tables, Progress 8 targets, floor standards and intervention thresholds together form a complex evaluative infrastructure (Ehren & Shackleton, 2016).

Performance tables—introduced in 1992 and expanded through the 1990s and 2000s—became a central tool for governing the school system, shaping parental choice, influencing media narratives and exerting competitive pressure (Coldron et al., 2014). Ofsted inspection regimes have increasingly relied on performance data as evidence of quality, contributing to substantial consequences for schools and leadership teams, including forced academisation or leadership removal following unfavourable outcomes (Perryman, 2009).

Incentives and sanctions embedded in the system have generated well-documented behavioural responses: teaching to the test, narrowing the curriculum, off-rolling, strategic student selection and other forms of 'gaming' (Bevan & Hood, 2006; Gillborn & Youdell, 2000). Such practices highlight the complex interplay between data, accountability and organisational behaviour and raise fundamental questions about the purposes and unintended consequences of SPMs.

2.2.6 Summary of historical and socio-political context of SPMs

While the introduction and development of SPMs in England are widely recognised as reflecting a convergence of ideological, administrative and economic shifts, rather than solely technical innovation, there remains limited empirical research that systematically examines the socio-political context underpinning their production and publication. Reforms from the 1988 Education Reform Act onwards incorporated marketisation, managerialism and accountability into the school system,

introducing the national curriculum, standardised testing, open enrolment and performance data aimed at supporting parental choice. Research has documented how these reforms reshaped relationships between schools, parents and the state, often benefiting middle-class families and reinforcing social inequalities. Empirical studies (Amrein & Berliner, 2002; Leckie & Goldstein, 2018a) further indicate that while competitive pressures and quasi-market mechanisms occasionally lead to modest improvements in measured outcomes, they more consistently exacerbate inequalities, creating advantages for schools with socially advantaged intakes and incentivising strategic organisational behaviours, including, for example, 'teaching to the test' and curriculum narrowing.

These developments were further reinforced by the rise of New Public Management and audit cultures, which normalised performance monitoring, target-setting and managerial oversight as central instruments of governance. Schools became subject to overlapping accountability regimes, including performance tables, targets and Ofsted inspections, with external metrics exerting substantial influence on leadership, resource allocation and organisational priorities. The critiques discussed above highlight that high-stakes SPMs can distort educational practice, misalign internal and external accountability and inadequately reflect the broader qualities parents value in schools. Taken together, these historical and structural forces not only created the conditions for the development of SPMs but also embedded expectations of transparency, comparability and competition—setting the stage for their ongoing role in shaping behaviour, decision-making and governance across England's education system.

Despite these insights, there is a scarcity of detailed, empirically grounded accounts of the socio-political forces, decisions and negotiations that shaped the production and publication of SPMs. This gap limits understanding of how ideological, administrative and economic imperatives translated into specific measures and processes and how these have influenced the broader governance and operation of English secondary schools. Research Question 1 and the first research phase both aim to address this gap by providing a detailed exploration of the socio-political context in which SPMs were developed, highlighting the structural and policy dynamics that underpin their continued influence.

Building on this foundation, the following section explores how these governing logics manifest in the design and communication of current SPMs.

2.3 Design and Dissemination of SPMs

SPMs represent a diverse and evolving set of statistical indicators designed to evaluate schools, inform stakeholders and shape school priorities. In England, SPMs have included raw attainment

indicators, value-added and contextualised models, curriculum-based indices (e.g., EBacc, Attainment 8) and broader outcome measures such as destinations data or inclusion metrics. The proliferation of SPMs reflects long-standing debates about what should count as 'school performance' and how best to represent it (Gorard, 2010; Leckie & Goldstein, 2017a; Goldstein & Spiegelhalter, 1996). These debates draw on a substantial methodological tradition in school effectiveness research, most notably Goldstein's (1997) foundational contribution, which established key statistical principles for modelling school effects in hierarchically structured data.

2.3.1 Types of SPMs and their evolution in England

The English system has shifted over time from simple attainment indicators to increasingly complex progress-based models. Early critiques by Gray and Jesson (1990) and Mayston and Jesson (1991) cautioned that the construction of performance indicators inevitably involves negotiation and compromise. They warned that narrowly defined metrics risk distorting educational priorities. Despite these warnings, SPMs expanded in scope while retaining strong ties to quantifiable academic outcomes.

A wide range of methodological approaches have been employed to model school effectiveness. Raw attainment indicators assess outcomes without adjustment (Choi et al., 2005), while hierarchical/multilevel models (Aitkin & Longford, 1986; Zvoch & Stevens, 2008) and value-added models (Leckie & Prior, 2022; Tekwe et al., 2004) attempt to isolate school contributions by controlling for prior attainment. Goldstein (2001) highlighted the importance of recognising the multilevel structure of educational data and the need for models to account for variance at pupil, classroom and school levels.

Contextualised value-added (CVA) approaches further adjust for student background factors such as socioeconomic status (Troncoso, 2019). The extent to which such adjustments should be made remains a significant point of contention. Leckie and Goldstein (2018b, 2019c) demonstrate that omitting pupil background from Progress 8 can systematically disadvantage schools with challenging intakes, calling into question the validity of high-stakes comparisons using such a metric. These findings continue a long-standing methodological argument initiated by Goldstein and colleagues regarding how best to estimate and interpret school effects.

Variants of modelling include threshold measures (Leckie & Goldstein, 2017b), expected-progress or 'growth' models (Jennings & Corcoran, 2009; Leckie & Goldstein, 2017a), Bayesian estimators (Brunsdon, 2001), mixture models (Dominitz & Sherman, 2006) and latent-variable approaches (Fox, 2004). Spiegelhalter's (2019) work on statistical inference is increasingly influential here, especially

in emphasising uncertainty, probabilistic interpretation and the risks of overconfidence in point estimates presented as objective rankings.

Each model incorporates normative choices regarding which variables are important, how learning is understood and where responsibility is assigned. These design decisions profoundly influence how schools are ranked, compared and governed.

In current policy frameworks, the Department for Education (DfE, 2022) articulates clear aims for secondary SPMs: to reward broad and balanced curricula, prioritise core subjects, inform parental choice, promote improvement and provide signals of underperformance for regulatory action. Yet research repeatedly questions whether the technical design of these measures faithfully supports these intended functions. This is the subject of the next section.

2.3.2 Technical design considerations

A substantial literature highlights the methodological challenges inherent in modelling school effectiveness. 'Progress over time' versus snapshot attainment measures generate very different pictures of school performance (Harris, 2018). Choices about variables, model specification and pupil subgroup treatment directly influence results (Leckie & Goldstein, 2017a). Common problems include ceiling and floor effects, regression to the mean, measurement error and instability in annual rankings—issues extensively discussed in the methodological critique offered by Gorard (2010), who warns that statistical artefacts can easily produce misleading patterns in league tables.

Value-added (VA) models also face conceptual and empirical limitations. Goldhaber (2002) highlights both their promise (estimating school or teacher 'effects') and their inherent fragility, noting problems of omitted variable bias, model instability, sensitivity to assumptions and difficulties in attributing causality. These concerns align closely with Goldstein's (1997) argument that identifying causal school effects is extremely difficult given the complexity of educational processes and the non-experimental nature of administrative data.

Progress 8 exemplifies these tensions. Leckie and Goldstein (2019a) show that its omission of contextual background leads to systematic bias against schools serving disadvantaged populations. They argue that Progress 8 risks rewarding and punishing the wrong schools, with predictable consequences for school behaviour and student opportunities. Perry's (2016) review similarly identifies threats to validity and misuse in policy settings.

Numerous researchers have shown how SPMs interact with student demographics in patterned ways—e.g., disadvantage (Allen, 2015; Nye & Thomson, 2019), selection effects (Cullinane, 2020;

Richardson & Treadaway, 2016) and ethnicity and language (Nye, 2020)—raising normative questions about fairness and comparability.

As modelling becomes increasingly complex, Kelly and Downey (2010) caution that SPMs risk becoming "black boxes", where opacity obscures assumptions and undermines practitioner understanding. Their argument resonates with broader principles of Fairness, Accountability, Transparency and Ethics (FATE), which call for cautious, transparent and context-aware interpretation of high-stakes metrics. The long-standing warning from Goldstein & Spiegelhalter (1996)—that performance indicators carry uncertainty that must be explicitly communicated—remains fundamental.

2.3.3 Data infrastructures and data quality

SPMs depend on large administrative datasets such as the National Pupil Database (NPD), census returns and geodemographic indices (e.g., IDACI, IMD). While these infrastructures enable national-scale comparisons, they also introduce challenges of data quality, missingness, definitional drift and inconsistent linkages. Kitchin's (2014a) analysis of data infrastructures emphasises that these systems are not neutral: data collection and processing are historically contingent, political and shaped by institutional priorities. These insights underscore the importance of interrogating how SPMs come to be, not merely what they show.

Data governance and privacy concerns also surface when pupil-level data are repurposed for accountability, used beyond their original intent or accessed by third-party organisations without transparency.

2.3.4 Communication, visualisation and dissemination

The dissemination of SPMs—through league tables, dashboards and interactive platforms—profoundly shapes how the public interprets school quality. Visualisation plays a central role in rendering complex data simple, stable and authoritative (Williamson, 2017). However, scholarship on numbers and rankings warns that such presentations often mask uncertainty, amplify symbolic power and influence organisational behaviour (Espeland & Sauder, 2007).

Research on visualisation highlights how design choices, metaphors and cognitive framing embed implicit biases (Grainger et al., 2016; McInerny et al., 2014; Sculley & Pasanek, 2008). Spiegelhalter's (2019) emphasis on statistical literacy is germane here: the public can too easily interpret noisy, uncertain metrics as definitive, particularly when communicated visually.

Work on 'data emotions' (Hill et al., 2016; Kennedy, 2022; Kennedy & Hill, 2018) suggests that visual representations not only inform but also persuade, eliciting affective responses that reinforce or

challenge policy narratives. This is especially relevant at a time when both the DfE (2025) and Ofsted (2025) are consulting on changes to accountability and inspection, including the redesign of public-facing data displays. The design and dissemination of performance metrics are rarely neutral technical exercises; they embed normative assumptions, reflect political priorities and introduce statistical limitations that shape how school performance is understood. These design choices often amplify existing social inequalities or obscure contextual factors beyond schools' control, prompting sustained critical engagement from researchers, practitioners and policymakers. Accordingly, the next section synthesises major critiques of current approaches to SPMs, drawing on evidence of their unintended effects and exploring alternative models that seek to offer more equitable, context-sensitive or theoretically grounded assessments of school performance.

2.3.5 Summary of SPM design and dissemination

SPMs in England have evolved from simple attainment indicators to complex, progress- and (at times) context-adjusted models, reflecting both methodological advances and policy priorities. While raw attainment, value-added and contextualised metrics aim to isolate school contributions and inform stakeholders, research highlights persistent challenges related to validity, fairness and interpretability.

Modelling choices directly influence rankings and can systematically disadvantage schools serving socially disadvantaged populations. Meanwhile, increasingly sophisticated statistical techniques, while methodologically rigorous, risk creating 'black boxes' that obscure the underlying assumptions and limit practitioner understanding, raising wider concerns about transparency and ethical accountability.

These observations indicate that the design and dissemination of SPMs are not neutral technical exercises. Instead, they are deeply embedded in political, methodological and structural contexts, embedding choices and statistical limitations that shape how SPMs are both understood and operationalised. Decisions about modelling, adjustment for background, data infrastructures and visualisation can amplify or mitigate existing inequalities, yet empirical research has rarely explored how these socio-political, methodological and structural factors interact to produce and operationalise SPMs in practice.

This gap in the literature serves as a motivation for this research, prompting the 3 research questions which aim to move beyond critiques of individual measures to provide a holistic understanding of the production, operationalisation and improvement of SPMs in England.

The next section synthesises major critiques of current approaches to SPMs, exploring their unintended effects and considering alternative, more equitable and theoretically grounded models.

2.4 Critiques and Alternative Approaches

A substantial body of research raises concerns about the validity, fairness and broader consequences of SPMs. Many critiques stem directly from the design choices discussed in section 2.3. These, including decisions about what to measure, how to model it and how to present it to the public, carry deep social and political implications. Scholars highlight tensions between the technical aspirations of performance metrics and the socio-economic realities of schooling, as well as the problematic assumption that school quality can be reliably captured through a small number of quantifiable indicators. This section synthesises major strands of critique, including the role of context, bias in model design, alternative frameworks, international lessons, unintended consequences and sociological accounts of performativity. I begin by exploring quantitative critiques and their implications for the forms of bias baked into SPM design and for SPM reform proposals. I then turn my gaze to example models arising from international comparisons and broader sociological critiques of SPMs, which consider SPMs not solely as technical measurement instruments but also as powerful policy tools that actively shape the operation of the school system.

2.4.1 Implications of design and context for the effectiveness of SPMs

As noted earlier, separating design from critique is conceptually difficult because the limitations of SPMs often arise directly from modelling choices. A longstanding finding across educational research is that student prior attainment and socio-economic context explain a large proportion of the variation in educational outcomes as captured by SPMs. There exists a considerable body of research into the difference that schools have on student outcomes, with differing approaches taken to control for student characteristics. The Coleman Report (Coleman, 1966) is often cited as the source of this claim (Gorard, 2010).

While there have been significant challenges made to the methodologies employed within the Coleman Report (Jencks, 1975), its publication marks an important starting point for further work in this area, using more robust statistical techniques. Further work, typically using Multi-Level Modelling techniques with large administrative datasets aggregated at student level, while failing to provide a final, definitive percentage, suggest that the proportion of variance in student achievement can be attributed to pupil characteristics is relatively large, with only a small proportion attributed to school effects (Leckie, 2008; Lessof et al., 2018a; Wilkinson et al., 2018). Early syntheses of this emerging literature reached similar conclusions: a systematic review by Scheerens and Bosker

(1997), drawing on international school effectiveness studies, estimated that schools accounted for approximately 19% of variance in attainment prior to adjustment for student background, falling to around 8% once initial differences between pupils were taken into account. Further evidence comes from Strand (2014), who demonstrates that once ethnicity and deprivation are modelled appropriately, the apparent contribution of individual schools diminishes substantially. This reinforces arguments that intake characteristics dominate outcome variation and that SPMs can systematically disadvantage schools serving marginalised communities. Differences in the percentage variances explained by school effects suggested by these different studies are largely the result of a complex picture that is sensitive to each study's design and definitions. Claims that such research asserts that schools make no difference fail to appreciate the reality that a 'school effect' is neither static nor simple. Its measured magnitude is dependent on the timing and context of the study, the methodological approach, the variables included in the model and the very definition of 'achievement' used. The important finding from this work, however, is that student characteristics, particularly prior attainment and socioeconomic background, are the most powerful predictors of achievement as measured by SPMs. This has important and far-reaching consequences for the use of SPMs. When performance indicators insufficiently account for these contextual factors, comparisons between schools may be misleading. This 'contextual blind spot' risks misidentifying school effectiveness (Goldstein, 2001; Leckie & Goldstein, 2017a) and raises concerns about the fairness, validity and interpretability of high-stakes SPMs.

Research on differential school effectiveness examines whether schools are equally effective for all pupils or whether some schools are more successful with particular student groups, such as those defined by ethnicity, socioeconomic status, gender, or prior attainment. Early school effectiveness research primarily focused on average school effects, using value-added models to estimate the contribution schools made to student outcomes after accounting for pupil background characteristics (Goldstein, 1987; Sammons, Mortimore, & Thomas, 1996). However, researchers increasingly argued that schools may not benefit all students equally, leading to the development of research on "differential effectiveness" (Cuttance, 1989).

Studies conducted in England during the 1990s and 2000s found evidence that schools can vary in their effectiveness for different groups of pupils and across subjects or cohorts. For example, Sammons, Nuttall, and Cuttance (1993) demonstrated that schools differed in their effectiveness according to pupils' prior attainment in reading and mathematics, although less evidence was found for differential effects by gender or ethnicity. Similarly, Thomas et al. (1997) identified variation in secondary school effectiveness for different pupil groups across GCSE outcomes.

More recent research has focused particularly on educational inequality and attainment gaps. Strand (2010, 2016) examined whether some schools were more successful than others in reducing disparities linked to ethnicity, poverty, gender and prior attainment. Using large-scale national datasets, Strand found persistent inequalities in pupil progress, especially among disadvantaged and some minority ethnic groups, but relatively limited evidence that these gaps were primarily attributable to pupils attending less effective schools. Instead, many schools that were effective for advantaged pupils were also effective for disadvantaged pupils.

The literature also highlights important methodological debates surrounding school accountability and value-added measures (Scherer & Nilsen, 2019). Researchers have questioned whether school performance indicators should adjust for pupil background characteristics (Goldstein, 1997), since schools serving more disadvantaged intakes may otherwise appear less effective (Leckie & Goldstein, 2019). Others have argued that differential effectiveness analyses are essential because average school effects can conceal important within-school inequalities (Nuttall et al., 1989; Strand, 2010, 2016).

Quantitative estimates in the school effectiveness literature consistently reinforce the argument that student intake characteristics account for the majority of explained variation in educational outcomes, often leaving only a relatively small portion attributable to between-school differences. An early synthesis of school effectiveness research noted that national systems, schools and departments combined explain approximately 0–20% of the total variance in outcomes, with the remainder largely explained by student background, prior attainment and residual variance—a finding that underpins the oft-cited notion of an “80/20” split between contextual and school factors (Goldstein, 2001; OECD, 2016). Official analyses in the English context have reported comparable magnitudes: using national data for 2003, the Department for Education and Skills (DfES, 2004) estimated that approximately 20% of the variance in KS4 attainment occurred between schools, falling to 13% for value-added between KS2–KS3 and to around 8% between KS3–KS4 once prior attainment was taken into account. More recent value-added and multilevel modelling studies in the English context find that once intake differences in prior achievement are accounted for, between-school variation can fall from around 20–30% to as low as 4–7% of total variance, with the majority of explained variance associated with pupil-level characteristics rather than schools per se (Leckie, 2009; Lessof et al., 2018b). Analyses of value-added models further demonstrate that the remaining school-level residuals are often small and unstable over time, limiting their interpretability as indicators of enduring school effectiveness (Ray et al., 2009). Similarly, analyses of secondary school outcomes in England show that pupil background characteristics, including prior attainment and socio-economic deprivation, can explain around two-thirds or more of the variance in KS4 attainment

measures at school level (Strand, 2014; Wilkinson et al., 2018). These magnitudes are not unique to any single methodology or jurisdiction: international comparative research repeatedly demonstrates that within-school (student-level) variance substantially exceeds between-school variance, and that estimated school effects diminish further once robust contextual controls are introduced (OECD, 2016).

Thrupp's (1999) critique of school effectiveness research does not deny the existence of school effects, but rather challenges the way in which they are interpreted and mobilised in policy discourse. Drawing on multilevel studies which typically report post-adjustment school effects in the region of approximately 5–15%, Thrupp accepts that schools can and do make a measurable difference to pupil outcomes. However, he argues that these effects are socially and structurally constrained by school composition, or 'school mix', and that treating pupil background as a control variable risks obscuring its constitutive role in shaping school processes and outcomes.

Gorard (2010) questions the very foundations of school effectiveness measurement. By examining potential sources of error—from data generation and cleaning to model specification and interpretation—he argues that SPMs are fundamentally unreliable. His critique sits within the wider fields of School Effectiveness Research (SER) and School Improvement Research (SIR), both of which underpin contemporary debates about how effectiveness should be conceptualised. SER encompasses a wide range of topics—school leadership (Hallinger & Huber, 2012), teaching quality (Slater & Teddlie, 1992), socio-economic effects (Angus, 1993), equity (Thrupp, 2001), learning environments (Creemers & Reezigt, 1997) and assessment (Scherer & Nilsen, 2019) - with a core aim of understanding 'what works' and why (Scheerens & Creemers, 1989).

SIR, emerging partly in response, focuses more closely on policy and practice, emphasising leadership, professional development, community engagement and evidence-informed innovation (Deppeler & Ainscow, 2016; Rhodes & Brundrett, 2009). Research such as Bryk et al. (2010) further underscores that school improvement depends on organisational capacity, relational trust and professional culture—dimensions largely invisible within current SPMs. Together, these fields highlight the complexity of educational processes and the limitations of reducing school effectiveness to simplified quantitative metrics.

2.4.2 Bias in model design

Concerns about bias in the design of SPMs parallel debates about algorithmic fairness in other sectors (Barocas & Selbst, 2016). Decisions about which covariates to include, how to handle missingness or how to treat different subgroups embed value judgements (O'Neil, 2016). As Koretz

(2017) notes, performance metrics often produce distributional inequities, disproportionately affecting schools serving disadvantaged communities.

A substantial methodological literature emphasises the role of implicit assumptions in model construction (Jackson et al., 2000; Jakeman et al., 2006; Loehle, 2011; Ord, 1979; Silvert, 1981). Wieland et al. (2016) caution that complex models rely on numerous untested assumptions that are rarely revisited once operationalised. In measurement contexts, unstated assumptions can compromise comparability; as Medina et al. (2009) argue, without careful scrutiny "conclusions may be little more than arbitrary" (p. 333).

Performance measurement frameworks (Bitici, 2015; Fitzgerald et al., 1991; Neely et al., 2002) and structural equation modelling (Bollen, 1989; Diamantopoulos et al., 2008; Jöreskog Wallentin, 1996) provide tools for interrogating the relationships between observed indicators (e.g., Progress 8 scores) and latent constructs (e.g., school effectiveness). While multilevel modelling has generated significant insights into the limitations of existing SPMs (Leckie & Goldstein, 2008a; Ray et al., 2009; Troncoso, 2019), there remains comparatively little work examining measurement models through social, political or cultural lenses. The Fiscal Studies special issue on league tables (Allen & Burgess, 2011) made substantial contributions, yet the accountability architecture has evolved considerably since then and contemporary measures require renewed scrutiny.

2.4.3 Alternative models and reform proposals

Researchers have proposed a wide range of alternatives aimed at improving validity, fairness and usefulness. Perry (2016, 2019) and Leckie and Goldstein (2019c) advocate for the inclusion of contextualised value-added (CVA) approaches, a feature which was removed from SPMs in 2011 (Department for Education, 2010) (see also sections 1.1, 2.3.1 and 5.2.2), arguing that ignoring pupil background undermines the meaningfulness of comparisons. Allen (2019, September 29) and Richardson & Treadaway (2016) propose refinements to cohort definitions, such as weighting pupils by time enrolled rather than using the January census point, addressing distortions caused by student mobility.

More systemic reforms have also been explored. The DCSF and Ofsted's (Department for Children, Schools and Families, 2008, 2009) 'School Report Card' project introduced a broader, multi-indicator framework inspired by international models, with an unusually strong emphasis on presentation and stakeholder engagement. Schagen's (2006) work on standardised residuals informed the development of CVA measures in the national performance tables between 2006 and 2011. The

subsequent discontinuation of CVA—justified by the DfE on grounds of complexity, predictive limitations and differential expectations—remains controversial (Leckie & Goldstein, 2017a).

International reform experiences provide further inspiration for alternatives. Sahlberg's (2011) work on Finland demonstrates that high performance can be achieved through trust-based, low-stakes systems that prioritise teacher professionalism over data-driven accountability. Similarly, Bryk et al. (2010) offer a multi-dimensional framework—centred on organisational capacity and relational trust—that contrasts sharply with England's output-focused model and suggests richer ways to conceptualise school effectiveness.

Recent innovations include efforts to develop more holistic and user-centred dashboards. The Children's Commissioner for England (2021) has proposed a 'Good School Index' incorporating wellbeing and pupil experience. Software for Data Analysis (School Performance Tables, 2022) has created a 'National Single Indicator' that aggregates multiple measures into a customisable star-rating system. The Northern Powerhouse Partnership's 'Fairer Schools Index' (Northern Powerhouse Partnership, 2024, October 30) uses rich contextual adjustments to illustrate how school rankings change when background factors are considered. Lighthouse Data (Lighthouse, 2022) integrates achievement, finance, workforce and contextual data into a single analytic platform. Together, these initiatives reflect growing dissatisfaction with narrow attainment-based approaches and a desire for more nuanced, multidimensional models.

2.4.4 International comparisons

International evidence highlights both the diversity of accountability systems and recurring challenges. OECD's PISA operates as a global performance framework that has influenced national policy agendas (Goldstein, 2017a; Grek, 2009; Sellar & Lingard, 2013), though not without critique regarding validity and political effects.

Countries vary widely in the stakes attached to SPMs (Carnoy & Loeb, 2002). In the United States, School Report Cards have evolved across policy eras—from No Child Left Behind's (NCLB) proficiency emphasis (Hoff, 2007) to the incorporation of growth models (Carey & Manwaring, 2011) under the Every Student Succeeds Act (ESSA). Murray and Howe (2017) question their "validity as a policy instrument" (p.1), arguing that, despite their technical sophistication, these systems often fail to measure what they claim to measure. They also problematise the assumption that accountability empowers communities, suggesting that such systems may, in fact, demobilise public engagement.

Other jurisdictions offer contrasting approaches. In Italy, Paletta (2019) demonstrates how new evaluation instruments have reshaped school leadership roles and created tensions between

responsibility and influence. Germany's post-PISA reforms show how accountability can be used to support rather than sanction schools: underperformance triggers targeted assistance rather than punitive measures (Davoli & Entorf, 2018), contributing to reductions in inequality (Niemann et al., 2017). Studies from the Netherlands (Timmermans et al., 2013), Spain (Martinez-Arias et al., 2009), Chile (Taut et al., 2016) and Portugal (Ferrão & Couto, 2014) further enrich the evidence base, offering insights into the effects of value-added models in diverse policy environments.

Sahlberg's (2011) account of Finnish reform provides a particularly powerful counterpoint to high-stakes accountability. Finland's system demonstrates that excellence can coexist with minimal testing, high teacher autonomy and strong social equity.

A substantial body of research has examined how neoliberal education policies have travelled globally, often framed as part of a broader Global Education Reform Movement (GERM) characterised by standardisation, accountability, test-based evaluation and market-oriented governance (Sahlberg, 2011, 2016). Scholars argue that these reforms do not diffuse organically, but are actively promoted through transnational policy networks, international organisations and global benchmarking systems. Central to this process is the OECD, whose PISA assessments have become powerful tools of 'soft governance,' shaping national education agendas through comparison, ranking and policy advice (Grek, 2009; Sellar & Lingard, 2013). UNESCO, the World Bank and other multilateral organisations similarly contribute to the global circulation of School Autonomy with Accountability (SAWA) reforms, coupling decentralisation with managerial accountability frameworks (Mundy et al., 2016; Verger, Fontdevila & Parcerisa, 2019). Research from the REFORMED project (Verger et al., 2017; Verger et al., 2019) shows how policy instruments such as accountability metrics, school choice and performance-based management diffuse through complex advocacy coalitions and governance networks, often resulting in hybrid systems shaped by the countries' own political histories and institutional structures. Collectively, this literature highlights how neoliberal policy ideas travel, become institutionalised and reshape national schooling systems through global governance mechanisms.

2.4.5 Consequences and unintended effects

High-stakes accountability generates a spectrum of unintended consequences. Well-documented effects include curriculum narrowing (Au, 2007), teaching to the test (Booher-Jennings, 2005) and strategic behaviours such as off-rolling or exclusion of vulnerable pupils (Daniels & Cole, 2010; Gillborn & Youdell, 2000). Koretz (2017) provides a comprehensive account of gaming dynamics, demonstrating how incentives distort both teaching and data.

Recent empirical evidence from Jerrim and Sims (2022) shows that such gaming is widespread in England, including the strategic entry of pupils into specific qualifications and the prioritisation of borderline pupils. Hutchings (2015) highlights the emotional and organisational 'hidden costs' of accountability, including staff stress, identity conflict and shifts in school priorities, while also documenting the use of exclusion and off-rolling to protect headline measures.

The pressures associated with SPMs can also affect staff morale, wellbeing and retention. Hutchings (2015) finds that accountability regimes contribute significantly to stress and burnout among teachers and school leaders, affecting recruitment and stability—especially in disadvantaged areas.

2.4.6 Sociological critiques of performativity

Sociological analyses frame SPMs not merely as technical instruments but as disciplinary technologies embedded within broader audit cultures. Ball's (2000, 2003b, 2012) work on performativity shows how performance data reshape teachers' identities, professional values and organisational priorities. SPMs generate 'fabrications'—strategic representations of practice designed to satisfy accountability demands rather than reflect the lived reality of schooling. These insights are reinforced by Mansur & Thorsuwan (2008), who document contemporary forms of identity conflict, anxiety and emotional labour associated with audit demands.

Power's (1997) theory of audit culture explains how measurement systems expand over time, producing new forms of surveillance and symbolic authority. Au (2007) further argues that high-stakes testing not only reshapes teaching practices but also contributes to the standardisation and stratification of inequality. Together, these critiques highlight that SPMs do not simply measure performance; they actively shape it, often in unintended and undesirable ways.

These diverse critiques naturally lead to questions about how researchers study performance measures themselves. Understanding the limits and consequences of SPMs requires methodological approaches that can capture not only statistical properties, but also ethical considerations, social dynamics and the broader political economy of data. The following section therefore reviews the methodological traditions employed in this literature, highlighting approaches that explicitly address data ethics, fairness and the interpretive challenges of performance analysis.

2.4.7 Summary of Critiques and Alternative Approaches

Existing research offers a thorough critique of secondary SPMs, highlighting their methodological limitations, biases, contextual challenges, international alternatives and unintended consequences. Studies consistently show that pupil background and prior attainment explain most variance in outcomes captured by SPMs, raising long-standing concerns about fairness and validity, especially

for schools serving disadvantaged communities. Research has also emphasised the impact of modelling assumptions and design decisions in embedding structural inequities and shaping how SPM results are interpreted and used. Broader sociological analyses further demonstrate how SPMs reconfigure behaviour, priorities, teacher identity and organisational practice, revealing SPMs as powerful policy technologies, not neutral measurement tools.

However, despite this substantial critique, important gaps remain. Much of the literature treats problems of design, context, consequences and reform separately, rather than examining how they are connected through the socio-political conditions under which SPMs are produced, operationalised and governed. There is comparatively limited empirical work that:

- Interrogates the contemporary socio-political context in which SPMs are developed and published and how this context frames what SPMs are expected to achieve.
- Evaluates SPM effectiveness against stated aims, taking account of how accountability measures are understood and used in practice by schools and stakeholders.
- Explores how advances in analytical methods and insights into data communication might address the fairness, interpretability and transparency concerns highlighted by existing critiques.

These gaps provide an important rationale for this thesis. Taken together, this research seeks to move beyond identifying limitations in individual measures towards a holistic, context-aware understanding of what SPMs do, how they do it and how they might better support equity, transparency and informed decision-making across the education system.

2.5 Methodological Approaches in the Literature

Research into SPMs employs a wide array of methodologies, reflecting the complexity of educational systems and the multifaceted nature of accountability regimes. Scholars have used quantitative modelling, qualitative inquiry, mixed-methods designs and emerging data-ethical frameworks to interrogate the construction, interpretation and consequences of SPMs. This section reviews the major methodological traditions in the literature, highlighting their respective contributions, limitations and implications for understanding the effects and effectiveness of performance measures.

2.5.1 Quantitative methodological approaches

Quantitative approaches have long dominated the evaluation of SPMs, particularly in relation to questions of validity, reliability and causal inference. Much of the early and contemporary work relies on econometric and statistical modelling to quantify school-level contributions to student outcomes.

A substantial strand of research focuses on value-added (VA) and contextual value-added (CVA) models, employing regression-based and multilevel formulations to estimate school effects while adjusting for student characteristics (Harris, 2008). The methodological foundations of these models owe much to hierarchical linear modelling frameworks such as Raudenbush & Bryk (2002), whose work established the statistical tools necessary for analysing nested educational data structures. Multilevel modelling remains central to current accountability debates (e.g., Leckie & Goldstein, 2009), enabling analysts to account for variation at student, classroom and school levels.

The specification of educational production functions—central to many VA models—has been critically examined by economists such as Todd & Wolpin (2003), who argue that seemingly technical modelling choices embed substantive assumptions about learning, causality and the role of schools. Such concerns have shaped ongoing debates about omitted variable bias, functional form and the interpretability of 'school effects.'

Recent contributions have systematically compared alternative VA model specifications. Leckie & Prior (2022) provide a detailed comparison of competing VA frameworks used for accountability, illustrating how minor modelling choices can produce substantively different school rankings. Prior, Goldstein & Leckie (2020) extend this work by developing multivariate VA models incorporating both academic and non-academic outcomes, offering a more holistic approach that challenges the narrow focus of traditional SPMs.

Beyond multilevel modelling, researchers have adopted factor analysis, item response theory (IRT) and structural equation modelling (SEM) to examine the latent constructs underlying school performance. SEM, in particular, supports analysis of causal pathways, latent variables and measurement error (Byrne, 2016), allowing scholars to conceptualise school effectiveness as multidimensional rather than reducible to a single observed indicator.

More recently, machine learning models—such as random forests, gradient boosting and neural networks—have been used to predict student outcomes or classify school characteristics (Murphy, 2012). While these models often outperform traditional regressions in predictive accuracy, they sacrifice interpretability and raise concerns about fairness, transparency and replicability, limiting their suitability for high-stakes accountability.

Despite the sophistication of quantitative methods, challenges persist: contextual bias, omitted variables, hidden value judgements in covariate selection and difficulties interpreting complex models. These limitations have motivated the complementary use of qualitative methods in the mixed methods approach in this study.

2.5.2 Qualitative methodological approaches

Qualitative research has played a crucial role in illuminating the social, cultural and political dimensions of SPMs—dimensions not accessible through statistical modelling alone. These approaches reveal how measures shape professional identities, organisational routines and power relationships.

Ethnographic research examines how teachers, school leaders and policymakers engage with performance data in everyday contexts. Spillane (2012) highlights the interpretive work involved in making sense of metrics, while Ball (2003b) and others expose the emotional labour, strategic behaviours and identity reshaping associated with performative accountability.

Interviews, focus groups, surveys and case studies have been used to explore the perceived fairness, legitimacy and consequences of accountability metrics. Thrupp's (1999) ethnographic–qualitative work demonstrates how socio-economic context shapes both measured and perceived effectiveness, illustrating the limits of purely quantitative explanations. Case studies likewise reveal situated experiences of curriculum narrowing, gaming, organisational stress and the pressures exerted by high-stakes performance measures.

Documentary and historical analyses add further depth, tracing how policy agendas, ideological shifts and institutional reforms shape the development and use of performance metrics over time (Ozga, 2019; Power, 1997). Such approaches show how SPMs become embedded within broader

governance technologies linked to marketisation, managerialism and audit culture (Ball, 2003b; Strathern, 2000; Wright & Shore, 1999).

Qualitative methods thus ground performance metrics in their lived realities, offering insights that complement and challenge quantitative analyses.

2.5.3 Mixed-methods approaches

Mixed-methods research provides a powerful framework for studying SPMs because it integrates statistical modelling with rich contextual inquiry. This approach allows researchers to capture both the psychometric or econometric properties of SPMs and the everyday processes through which they are interpreted and enacted.

A well-established mixed-methods tradition pairs quantitative estimates of school performance (e.g., VA or P8-style measures) with qualitative investigations of leadership, pedagogy, school culture or community factors (Thrupp, 1999). This enables a more comprehensive understanding of variation between schools that appear statistically similar but differ organisationally or contextually.

Sequential designs use qualitative insights to refine quantitative model specifications—identifying missing covariates or mis-specified relationships—or deploy quantitative analyses to identify cases for qualitative follow-up (e.g., outlier or 'unexpectedly effective' schools) (Creswell & Plano Clark, 2018; Teddlie & Tashakkori, 2009). Embedded designs combine qualitative and quantitative data within integrated case studies (Yin, 2018).

Mixed-methods approaches have also been used to examine implementation fidelity, stakeholder interpretation or the political reception of performance metrics (Bryman, 2006; Greene, 2007). These studies illuminate how even technically robust measures may produce unintended or undesirable effects when applied in complex social systems (Meyer & Rowan, 1977; Spillane & Burch, 2006).

Although resource-intensive and sometimes epistemologically challenging, mixed-methods work provides some of the most nuanced accounts of accountability regimes.

2.5.4 Data ethics methodologies

As performance metrics have become increasingly complex and consequential, research has expanded to address the ethical, epistemic and socio-political dimensions of data use in school accountability. Scholars working at the intersection of education, data science and public-sector governance emphasise core principles of fairness, transparency, explainability and responsible data

stewardship, reflecting wider concerns about the risks posed by performance measurement in high-stakes environments (Griffiths, 2020; Pangrazio et al., 2024).

Building on foundational work in algorithmic accountability and fairness (Barocas & Selbst, 2016; Kroll et al., 2017; Mehrabi et al., 2021; O'Neil, 2016; Selbst et al., 2019), researchers have raised concerns about the fairness of SPMs. Key issues include whether statistical specifications disproportionately disadvantage particular social groups; whether choices about covariates and standardisation tacitly encode structural inequalities (Gillborn, 2018; Gorard, 2018); and how opaque modelling practices limit opportunities for scrutiny (Amrein-Beardsley, 2014; Goldstein & Thomas, 2008). These critiques connect directly to long-standing debates about whether performance indicators in public services generate distortions, bias or perverse incentives (Kelman & Friedman, 2009; Smith, 1995).

Increasingly, debates about SPMs draw explicitly on explainability and interpretability frameworks originating in machine learning (Guidotti et al., 2018; Molnar, 2019). These frameworks argue that accountability metrics must be comprehensible to educators, school leaders, policymakers and the wider public. Concerns have been raised about the opacity of multivariate or model-based measures such as Contextual Value Added (CVA) or Progress 8, which can obscure rather than illuminate causal processes and make it difficult for stakeholders to evaluate their validity or limitations (Goldstein, 2017b; Leckie & Goldstein, 2017b).

Data ethics methodologies complement quantitative and qualitative approaches by foregrounding issues of power, equity, responsibility and public legitimacy. This aligns with the broader field of Critical Data Studies, which examines how data systems reproduce social hierarchies and institutional power (Kitchin, 2014a; Dalton & Thatcher, 2014; boyd & Crawford, 2012). These perspectives illuminate how performance metrics can act not merely as neutral measurement tools but as instruments of governance, shaping organisational behaviour, constructing normative definitions of 'good performance' and embedding political values within technical systems.

Methodological approaches to SPMs are therefore deeply entangled with theoretical perspectives on data, governance and accountability. As the field has evolved, scholars have increasingly drawn on interdisciplinary frameworks—including FATE (Fairness, Accountability, Transparency and Ethics), Critical Data Studies and socio-technical systems theory—to interrogate how power, technology and policy intersect within performance regimes. This shift reflects a growing recognition that the ethical analysis of SPMs requires attention not only to statistical validity but also to the broader socio-political contexts in which metrics are produced, interpreted and used.

2.5.5 Summary of Methodological Approaches in the Literature

Research on school performance measures (SPMs) spans a wide range of methodological traditions, reflecting the complexity of accountability systems. Quantitative approaches dominate, especially value-added (VA) and contextual VA models, multilevel modelling and comparative evaluations of model specifications, with more recent work exploring predictive machine learning. These techniques are well suited to estimating school-level contributions to outcomes, but they face long-standing problems of contextual bias, omitted variables, value-laden modelling choices and limited interpretability. While some adjacent areas of educational effectiveness research employ structural equation modelling (SEM) or latent variable analysis, these approaches remain relatively rare within the core SPM literature itself, where accountability demands simplicity and transparency. This rarity highlights a methodological gap: the multidimensionality of school performance is often theorised but seldom modelled explicitly.

Qualitative research provides essential insight into how SPMs operate in practice. Ethnographies, interviews, case studies and documentary analyses illuminate the lived realities of high-stakes accountability, including its emotional effects, strategic behaviours and organisational distortions. These studies reveal the social and political work performed by metrics, often exposing consequences that are not visible through statistical modelling alone.

Mixed-methods studies combine quantitative estimates with qualitative inquiry to bridge these perspectives. They reveal why statistically similar schools may differ in practice, uncover mis-specified models and help explain the reception and enactment of accountability measures. Although powerful, mixed-methods approaches remain resource-intensive and are employed less frequently than single-method traditions.

A growing line of scholarship addresses the ethics, fairness, transparency and socio-political implications of performance metrics. Drawing on data ethics, algorithmic accountability and Critical Data Studies, this research critiques the opacity of model-based measures, the embedding of structural inequalities in covariate choices and the role of SPMs as tools of governance rather than neutral indicators.

The field faces persistent tensions: between statistical sophistication and interpretability; between technical validity and social legitimacy; and between standardised metrics and the diverse contexts of schooling. Quantitative methods dominate but often omit organisational or ethical dimensions; qualitative approaches capture these but lack generalisability; mixed-methods work offers integration but remains underdeveloped. Importantly, multidimensional or latent models of school

performance — such as SEM — are not yet mainstream in SPM research despite their potential to better represent the complexity of schooling.

These methodological trends inform my own approach, which combines advanced quantitative modelling with qualitative and ethical perspectives. By using SEM alongside multilevel methods, I address a rarely explored methodological gap: the latent structure of school 'performance' and the extent to which observed indicators (e.g., Attainment 8 or Progress 8) reflect contextual factors rather than school effectiveness. Integrating quantitative, qualitative and data-ethical lenses enables my research to examine not only the statistical properties of SPMs but also their consequences, legitimacy and governance implications within real school systems.

2.6 Synthesis of Findings and Identification of Gaps

Across the literature, a broad consensus exists that school performance measures (SPMs) in England are not merely technical instruments but products of deeper historical, political and socio-economic developments. From the 1988 Education Reform Act onwards, quasi-marketisation, managerialism and audit culture reshaped schooling, embedding expectations of comparison, transparency and competition. These structural conditions created both the demand for and the logic underpinning SPMs. Yet despite substantial historical and sociological critique, there remains limited empirical research that systematically examines the socio-political processes through which SPMs were produced, negotiated and justified. This gap is directly addressed in Research Question 1, which explores how policy, ideology and institutional pressures shaped the emergence, design and legitimisation of SPMs.

A second major theme concerns the design and dissemination of SPMs. While measures have grown more sophisticated—from simple attainment indicators to progress-adjusted and contextual value-added models—research consistently shows that methodological decisions embed implicit value judgements, disproportionately affect schools with disadvantaged intakes and can produce misleading or opaque representations of 'school performance.' Sophisticated modelling has introduced accuracy in some respects, yet often at the cost of transparency and stakeholder understanding. Despite this, few studies analyse how political priorities, modelling assumptions, data infrastructures and communication practices interact to shape the SPMs that are eventually published and used. This lack of integrative research underpins Research Question 2, which examines how design choices and dissemination practices influence interpretation, use and perceived legitimacy.

The third major strand concerns critiques and consequences. Current SPMs capture variance that is overwhelmingly driven by pupil background and prior attainment, raising long-standing concerns about fairness and validity. Sociological studies show how SPMs reshape behaviour, curriculum, identities and organisational practice, positioning metrics as powerful policy technologies that govern schools through comparison, surveillance and competition. Yet much of this work treats design, context, consequences and reform in isolation. There is comparatively little research that connects how SPMs are produced, what they measure, how they are interpreted and what effects they generate within schools and communities. Addressing this fragmentation, Research Question 3 investigates the conceptual and practical adequacy of current SPMs and explores how more transparent, equitable and theoretically grounded models might be developed.

Across these strands, several methodological gaps become apparent:

1. Lack of conceptual clarity about what 'performance' measures actually measure.

Although scholars critique validity and fairness, very few studies interrogate the latent structure of 'school performance' itself. There is almost no use of exploratory or confirmatory factor analysis to examine whether attainment, progress, EBacc measures and destination metrics represent coherent underlying constructs or multiple dimensions. Treating SPMs as single indicators masks multidimensionality and obscures theoretical understanding of 'effectiveness.'

My research addresses this gap directly by applying Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) to conceptualise the structure of 'performance'—a methodological contribution currently absent from the field.

2. Minimal use of Structural Equation Modelling in SPM research

While SEM is used in broader educational research, it is rarely used to interrogate SPMs directly, particularly with nationally available data. SEM offers the capacity to model measurement error, latent constructs and causal pathways, but these strengths remain under-exploited in high-stakes accountability research.

This thesis contributes by applying SEM to examine how contextual factors shape headline SPMs and to test theoretical relationships that existing models take for granted.

3. Over-reliance on pupil-level modelling, with limited analysis at school level

Much quantitative literature uses pupil-level NPD data for VA modelling. While powerful, this approach can obscure school-level relationships and constrain conceptual work on performance at the organisational level.

By working primarily with school-level aggregated data, my research foregrounds patterns and structures that become visible only at the organisational scale, complementing rather than replicating VA-based analyses.

4. Fragmented treatment of context, design and effects

Existing research tends to analyse:

- the socio-political origins of SPMs,
- the technical design of metrics and
- the everyday consequences in schools

as separate domains. Very little work integrates these strands to show how SPMs function as socio-technical systems shaped by power, history, measurement choices and institutional incentives.

My research adopts a holistic approach—historical, social, cultural and political—explicitly integrating these dimensions across three research phases.

5. Limited attention to communication, interpretability and public understanding

While some researchers highlight opacity and complexity, there is little empirical work examining how SPMs are communicated to—and interpreted by—stakeholders, particularly in relation to design choices, fairness and model assumptions.

This thesis contributes by analysing how SPMs are operationalised, interpreted and acted upon in practice, drawing on qualitative data and data-ethical frameworks.

2.6.1 Positioning This Research

Bringing these strands together, the literature reveals consensus that England's SPMs are shaped by powerful socio-political drivers, embed structural inequalities and suffer from methodological and communicative weaknesses. Yet it also reveals key tensions:

- between simplicity and accuracy,
- between fairness and comparability,
- between policy demands and methodological limitations,
- between what SPMs claim to measure and what they actually capture.

Crucially, what remains *unknown* is how these elements connect. There is no comprehensive account of SPMs that simultaneously addresses their historical origins, methodological assumptions, socio-political functions and conceptual validity as measures of 'performance.'

This thesis is positioned directly within these gaps, offering original contributions by:

- analysing the socio-political production of SPMs (Research Question 1),
- examining how design and dissemination shape interpretation and use (Research Question 2),
- advancing conceptual and methodological tools—including factor analysis and SEM—to develop more transparent, multidimensional and context-aware approaches to performance measurement (Research Question 3).

Through this integrated, multi-method and holistic approach, this research aims to move beyond critique towards a theoretically grounded and practically useful rethinking of how school performance is measured, understood and governed in England.

3. Theoretical-Analytical Frameworks

Three decades of the publication of SPMs in England has created the illusion that they have always been an integral part of the education system. The majority of staff working in English schools today will only have had experience of an education system which, in many ways, is significantly shaped by the existence of these high-stakes accountability measures. Their annual publication is received, even following disruption resulting from the COVID-19 pandemic, with a sense of inevitability. The way SPMs are presented, via government websites and announcements and through the national media, is extremely influential in imbuing them with meaning, suggesting that they are fair, reliable, independent and authoritative measures of school quality. The way in which the measures have been generated, adjusted, published and used has been questioned in only a relatively limited way. The bulk of the previous critiques and research into school accountability is dominated by work which explores the performance measures through a technical, quantitative examination and review of the statistical techniques used. 'Solutions' to the issues raised using this lens are therefore seen as being one of technical adjustments to these instruments. However, the statistical methods used in the production of SPMs are a small part of a much larger SPM 'data assemblage' (Kitchin, 2014b). To more fully understand SPMs, I will need to turn my gaze not only to their technical dimensions but also to a much wider conceptualisation from the viewpoints of their ethical, economic, political, temporal, spatial and philosophical dimensions.

In the next three sections, I explain the three theoretical-analytical frameworks I use in my evaluation of SPMs. These frameworks draw on ideas from Critical Data Studies, from the Fairness, Accountability, Transparency and Ethics (FATE) framework and Policy Logic scholarship. Each framework presents a different set of perspectives, viewpoints, foci, priorities, starting points, underlying scholarship, purposes, levels of analysis, assumptions, conceptual tools and inherent limitations. However, together they enable a deeper and more critical interrogation of SPMs, illuminating dimensions that would remain obscured if considered in isolation. Indeed, the complementary combination of these three frameworks is important for constructing a multi-layered critique of how school performance measures ultimately shape educational policy and practice. The three frameworks are united by their common function as critical lenses. In my explanations below, I will emphasise the points at which the frameworks align. However, it is worth noting that such alignments are not fixed. Instead, their synergy is dynamic, shifting in response to different contextual factors, including the situation under scrutiny, the research purpose and the interpretative choices the researcher applies.

3.1 Critical Data Studies

The emerging field of Critical Data Studies, as framed by, amongst others, Dalton and Thatcher (2014) and Iliadis and Russo (2016) will provide a theoretical-analytical framework for the exploration and evaluation of SPMs. CDS can be considered to be based around a number of 'axioms'. These provide an epistemological critique of the use of data by stating:

a) Data are never raw, an idea advanced by Kitchin (2014a), boyd and Crawford (2012), Lupton and Micahel (2016) and van Dijk (2014) and encapsulated in the term from used by Lisa Gitelman (2013) as the title of her book, taken from a quote by Geoffrey Bowker (2005), "'Raw Data' is an Oxymoron". This means that data do not simply 'exist'. Instead, they are created, collected, curated, cleaned, classified and processed through a variety of complex interacting social, political and technical decisions and mechanisms.

b) Data are never neutral, rather they are shaped by their model design, analytical methodology and interpretation which themselves are influenced by biases, assumptions and contexts (Kitchin, 2014a; D'Ignazio and Klein, 2020). Since they go through these various layers of transformation, they include embedded values and power relations. Every dataset therefore reflects the intentions, biases, assumptions and limitations of the systems and people who produced it.

c) Data have context, providing partial representations of the world based upon their creation through their design and the selection of analytical techniques used (Kitchin, 2014b). They gain meaning only within the social, political, historical and institutional environments in which they are collected and interpreted and their implications change when they are used in different contexts.

While still in its relative infancy, CDS as a distinct field brings together interdisciplinary work which reflects upon the wider philosophical concerns of the production and use of data. This includes its ideological, ontological, epistemological and methodological dimensions (Kitchin, 2014b). CDS provides an approach for not only reconceptualising the uses of school performance data but of unpacking the entire data assemblage associated with them; the technological, political, social and economic dimensions which frame their creation, publication and use. In this way I aim to consider the diverse mechanisms in which such school data are imbued with different meanings.

While some have questioned whether 'critical' should remain central, instead arguing for a broader field of 'Data Studies' (Kennedy, 2018), its inclusion reflects the continued influence of critical social theory within CDS, where the aim is to question, challenge and ultimately change the use of data within society. The concepts of 'politics' and 'power' have become central to CDS (Taylor, 2017). The practices within much of the body of work arising from CDS demonstrate a desire to uncover data

power, to challenge its operation and to suggest alternative approaches (Kitchin, 2022). This approach enables me to make visible the power which is at play in SPMs and the wider school accountability system and to interrogate the interpretations and behaviours that such power generates. Building on this critique, I will propose alternative approaches that foreground the inequalities produced by current SPMs – particularly their limitations in terms of fairness and transparency – while also identifying pathways toward more equitable, socially just and constructive ways forward.

The Critical Data Studies (CDS) framework I present below in Table 1 offers a theoretical-analytical lens for interrogating how data-driven systems are constructed, used and experienced. It builds on the three core axioms discussed above. Each axiom is elaborated through key themes which I have derived from established strands within Critical Data Studies as well as the related fields of Science and Technology Studies and critical quantification research. These themes represent widely recognised concerns about how data are produced, operationalised and experienced and collectively allow for the interrogation of the sociotechnical, political and ethical dimensions of data systems. These themes - social construction, politics, performativity, visibility, literacy, imaginaries, infrastructures, power, temporality, justice and reflexivity - highlight the sociotechnical and ethical complexity of data systems. Before providing greater details of these themes, I will set out how they relate to the three CDS axioms outlined above.

First CDS axiom - data are never raw.

This axiom emphasises that data are produced through classificatory choices, technical affordances and institutional priorities. Bowker and Star (1999) demonstrate how taxonomies encode assumptions and reproduce social order. This is reflected in the framework's (Table 1) emphasis on the idea that *Data are Socially Constructed*, where decisions about what to measure determine what becomes visible or invisible in policy fields. Strathern's (2000) critique of audit cultures similarly foregrounds how metrics narrow the scope of what is valued, a theme captured under *Inclusion/Exclusion/Invisibility* in the framework.

Second CDS axiom - data are never neutral.

This axiom aligns with scholarship that conceptualises data as political instruments. Kitchin and Dodge (2011) and Beer (2016) argue that data systems enact governance by shaping behaviour, expectations and institutional priorities. The framework's (Table 1) themes of *Data are Political*, *Data Ethics & Power* and *Data Justice & Equity* connect directly to these insights. Publishing metrics is not merely descriptive; it is performative, generating feedback loops that reshape the practices they evaluate. This echoes the performativity literature and the broader politics of data-driven ordering (Kitchin, 2014) as captured under the theme of *Data are Performative*.

Third CDS axiom - data have context.

The third axiom foregrounds the embeddedness of data within sociotechnical assemblages and institutional infrastructures. Kitchin (2014b) emphasises that data exist within data assemblages comprising technologies, organisations and regulations. The framework's themes of *Data Infrastructures & Assemblages* and *Temporality of Data* reflect this, noting that data's meaning depends on when it was produced, by whom and for what purpose. Issues of *Critical Data Literacy* and *Reflexivity* underscore that interpretation is shaped by the positionality and expertise of data users, echoing CDS calls for interpretive awareness and methodological transparency.

Collectively, these themes show how CDS exposes the contingent and contested nature of current data systems. *Alternative Imaginaries* are therefore necessary for envisioning fairer and more accountable forms of datafication. This theme highlights that metrics and accountability regimes are neither inevitable nor fixed and invites critical reflection on the kinds of futures and values that different data designs could enact. Similarly, CDS exposes how data designers and data users participate in producing meaning. *Reflexivity in data work* is therefore essential for interrogating how their situated perspectives shape the resulting metrics and interpretations. This theme challenges the idea of the neutral expert by foregrounding the social, professional and institutional contexts in

which data work takes place. As such, this CDS framework provides a rigorous foundation for analysing the SPM policy domain, highlighting how such data-driven governance is never merely technical but always ethical, political and relational and for proposing alternative approaches.

Table 1. Summary of Themes in Critical Data Studies Framework

Critical Data Studies theme	Explanation	Key Questions	Key Questions for the use of SPMs	Example SPM assumption challenged
Data are socially constructed	Measures reflect choices, e.g. about what/whom to count and how to classify them (Desrosières, 1998; Dhawka & Dasgupta, 2023)	Who decides what is measured and why? Whose perspectives shape the categories?	Who decides what SPMs are generated and prioritised? What alternatives were considered and rejected?	SPMs are objective reflections of schools quality i.e. effectiveness
Inclusion / exclusion / invisibility	Choices about what is measured determine what is valued. What is omitted becomes devalued [or just unseen] (Eubanks, 2018; Vono de Vilhena & Loi, 2020)	Who or what is missing from the data? What is underrepresented or ignored?	Who/what is missing? How are for example, SEND, vocational routes, wellbeing represented?	SPMs are fair and they capture what matters most
Data are political	Data act as tools of governance that steer behaviour, enacting and legitimising policy (Gitelman, 2013; Ruppert et al.,2018)	Who benefits from this data? How is it used to justify actions or policies?	Who benefits/loses from the use of these metrics? How do they contribute to interventions?	Accountability is a natural, necessary part of the system – driving required improvement
Data ethics & power	Focus on harm, surveillance and control (Kitchin & Dodge, 2011)	Who controls and benefits from the data?	What harms arise (stress, stigma, inequity)? Who controls access to data and the narratives that arise from it?	SPMs are harmless

Data justice & equity	Tackles structural inequalities reproduced by data systems (Eubanks, 2018; Noble, 2018)	How do data systems perpetuate social inequalities?	How do measures penalise disadvantage and/or entrench segregation?	Treating all schools 'equally' is inherently fair
Data are performative	Publishing metrics changes the behaviours and outcomes they measure by influencing practices and feedback loops (Amoore, 2011; Van der Velden & Harambam, 2016)	How does publishing data alter behaviour? What feedback loops exist?	What behaviours do SPMs induce? What practice gets optimised - or neglected?	SPMs report school performance *[they don't shape practice?]* – they inform and empower school choice
Data infrastructures & assemblages	Metrics sit within complex sociotechnical systems (Kitchin, 2014)	What institutions, technologies and policies underpin the data?	Which institutions, technologies, ideologies, policies and practices shape data and visibility?	SPMs operate independently of wider structures (e.g. DfE, Ofsted, media or data platforms)
Temporality of data	Data are time-bound snapshots that can have long afterlives (Drucker, 2011; Gitelman, 2013)	How long should data influence decisions? How quickly does it become outdated?	How long should a cohort result shape school reputation?	Scores are timeless, comparable indicators
(Critical) data literacy	Without context, data can mislead and/or entrench inequalities. Use requires grasping methods, limits and uncertainties (D'Ignazio & Bhargava, 2016; Kitchin, 2014)	Can stakeholders interpret the data meaningfully? How is uncertainty conveyed?	Do stakeholders understand context, cohort effects or confidence intervals?	Metrics are interpreted accurately by stakeholders (parents, schools and policymakers etc)

Alternative imaginaries	CDS invites rethinking towards fairer, more accountable, transparent and ethical data systems (Couldry & Mejias, 2019; Lehtiniemi & Ruckenstein, 2019)	What could data look like if designed for equity? What other metrics could be included?	What would equitable school accountability look like? Whose values would guide it?	The current model is the only viable option
Reflexivity in data work	Both designers and users form part of the system; their positionality matters (Crawford, 2013; D'Ignazio & Klein 2020)	How does stakeholder position shape the way they collect and/or interpret data?	How do roles and associated assumptions shape SPMs and the interpretations and actions arising from them?	Systems are designed by neutral 'experts'

3.2 Fairness, Accountability, Transparency and Ethics (FATE)

The advance of Big Data into many aspects of our lives and the data analytics which accompany them, raises the question of the ethical considerations of this rapidly changing field. Some attempt to address the questions raised by this are framed within the principles of Fairness, Accountability, Transparency, Ethics (FATE). The term FATE was originally associated with the Microsoft's FATE Research Group, whose aim is to "facilitate computational techniques that are both innovative and responsible, while prioritizing issues of fairness, accountability, transparency and ethics" (Microsoft Research Lab, 2025) by drawing on the deeper context surrounding these issues from information science, sociology, anthropology, human–computer interaction research, history, science and technology studies, political science and law. The term FATE has now taken on a significance beyond the work of Microsoft Research Lab. It refers to a set of key principles in the responsible use of data, particularly in public services, which aim to ensure that data-driven technologies are used in a way that is ethical and minimises harm. FATE also has a growing community which seeks to deal with fairness and ethics in the use of data-driven technologies in the form of, for example, the ACM Conference on Fairness, Accountability and Transparency (FAccT).

Others working in this field have developed 'toolboxes' which are designed to inspect the fairness, accountability and transparency aspects of such systems and to provide a mechanism for these to be objectively reported, for example FAT Forensics (Sokol et al., 2020) or CERTIFAI (Sharma et al., 2020). Such approaches are extremely valuable in the context of the ever-advancing use of Big Data in an educational setting.

FATE focuses on ensuring that data-driven and algorithmic systems operate fairly, transparently and accountably, often through normative principles and technical interventions to reduce biases and harm. Compared with this more applied orientation, CDS offers a broader socio-political critique, examining how datafication and algorithmic governance reproduce power relations and shape social realities. While both challenge the neutrality of data, FATE seeks to improve systems from within, whereas CDS interrogates the deeper structures and ideologies that give rise to those systems in the first place.

While much of the focus of those using FATE principles as a basis for their research is on the use of artificial intelligence (AI) or machine learning (ML), the same principles can be, and often are, considered to apply to any data-driven decision-making system, including traditional statistical models. For example, the principles of FATE (Table 2) can be applied to better understand a) the manner in which crime data can reinforce systemic bias in predictive policing (Alba, 2024; Chiao,

2019), b) how health risk assessment scores, trained on biased datasets, might underrepresent certain populations (Singhal et al., 2024) or c) how credit risk models can be discriminatory if based on variables which correlate with socioeconomic status (Valdrighi et al., 2024). Much can be learned by applying the same powerful FATE principles to the use of data in relation to schools.

Table 2. FATE Framework for School Performance Measures (SPMs)

FATE Theme	Explanation	Key Questions (general)	Key Questions related to SPMs	SPM Assumption Challenged
Fairness	Fairness concerns whether data systems and their uses produce or reproduce unjust advantages, disadvantages, or biases among groups or individuals. In education, this includes equity of opportunity, representational fairness and distributive outcomes.	Who benefits and who loses from the system's design, measurement choices and consequences? Are data-driven comparisons equitable across social, contextual and structural differences?	Do SPMs fairly represent the performance of schools serving different intakes and contexts? How do measures like Progress 8 or Attainment 8 treat variation in prior attainment, disadvantage, or inclusion?	That identical metrics applied universally produce fair comparisons; that differences in context are 'noise' rather than meaningful determinants.
Accountability	Accountability relates to who is answerable for decisions made through data, who holds power to define 'performance,' and how responsibility is distributed across actors (schools, policymakers, inspectors, communities).	Who defines the criteria of accountability and on what basis? Are those held accountable empowered to influence the metrics that judge them? What mechanisms exist for contestation or redress?	Who is accountable when SPMs misrepresent school effectiveness or create perverse incentives? Are schools, local authorities, or central government responsible for mitigating data-driven harms?	That accountability can be effectively and justly operationalised through quantified outcomes; that transparency of results alone ensures accountability.
Transparency	Transparency involves the visibility and intelligibility of data practices — including how data are collected, processed, interpreted and presented. It is not just disclosure, but also comprehensibility and interpretability.	Are the data processes, limitations and modelling assumptions open and understandable to all stakeholders? What is obscured or simplified in the pursuit of clarity?	Are the statistical adjustments (e.g. value-added models) and data cleaning processes behind SPMs transparent to educators, parents and the public? Do published tables adequately explain uncertainty and bias?	That data and models are neutral representations of 'facts'; that publication equals transparency without the need for interpretive literacy.

Ethics	Ethics concerns moral principles guiding the collection, interpretation and use of data — including considerations of harm, consent, proportionality and the public good.	What ethical frameworks underpin data practices? Are potential harms (e.g. stigmatisation, narrowing of curriculum, strategic exclusion) identified and mitigated? Who has the moral right to decide what counts as good performance?	How are ethical risks considered when SPMs influence funding, inspections, or public reputation? What ethical obligations exist to contextualise or qualify data before publication?	That metrics and rankings are inherently legitimate because they pursue accountability or improvement; that ethical considerations are secondary to technical accuracy.

3.3 Policy Logic Framework

The generalised Policy Logic Framework (PLF) which I have formulated and presented below (Table 3) offers a structured approach for analysing public policy. It integrates insights from policy design, critical policy studies and governance theory. In contrast to the CDS and FATE frameworks, which foreground the sociopolitical construction of data and the ethical consequences of their use, the PLF adds value by interrogating the internal rationale of performance measures themselves, revealing, for example, policy assumptions, causal logics and intended mechanisms that these other lenses do not explicitly unpack. The PLF draws on established traditions in policy sociology and evaluation to examine four dimensions. These are the Conceptual and Ethical Foundations of policy (Ball, 1994; Gewirtz & Cribb, 2002), the policy's Aims and Rationales (Bacchi, 2009; Weiss, 1995), the Values and Assumptions embedded in policy design (Apple, 2001; Stone, 2012) and, crucially, it foregrounds the Conditions and Processes of Operation (Hill & Hupe, 2014; Pressman & Wildavsky, 1984) which collectively aim to surface the intended rationales and internal coherence of SPMs. The fourth dimension, Conditions and Processes of Operation, is further divided into a sub-framework (Table 4) that provides a realistic approach to the evaluation of neoliberal-influenced policy by pragmatically considering a) the operational conditions (Honig, 2006; Pawson & Tilley, 1997) and b) the ethical-structural enabling conditions (Fraser, 2008; Sen, 2009) for SPM policy. These dimensions are discussed in more detail next.

The first dimension of the PLF, Conceptual and Ethical Foundations, emphasises that all policies are grounded in specific ideological, historical and ethical assumptions that influence how problems are defined and what is considered legitimate intervention. Drawing on framing theory (Bacchi, 2009; Fischer, 2003), this emphasises that problem definition is not neutral but mirrors underlying political and epistemic commitments.

The second dimension, Aims and Rationales, identifies both explicit and implicit goals that policies pursue. Policy instrument scholarship demonstrates that rationales often combine functional objectives, such as efficiency, accountability, legitimacy or behavioural change, with political imperatives (Hood, 1991; Howlett, 2011). This dimension clarifies for whom a policy is designed, what it intends to achieve and how it justifies its existence.

The framework's third dimension, Values and Assumptions, examines the normative and epistemic commitments embedded in policy design. Literature on valuation and governmentality (Espeland & Stevens, 1998; Miller & Rose, 2008) shows that policies privilege particular ways of knowing, often

involving quantification, standardisation and comparability, while marginalising others. This step surfaces the implicit hierarchies of value that shape policy priorities and consequences.

The fourth dimension, Conditions and Processes of Operation, focuses on how policy instruments work in practice. Implementation theory (Pressman & Wildavsky, 1984) and policy learning (Dunlop & Radaelli, 2018) demonstrate that policy outcomes emerge from interactions among actors, incentives, institutional capacities and wider systems.

This fourth dimension includes an additional sub-framework of pragmatic conditions for the operation of neoliberal policy. By adding this to the framework, I am choosing to reframe the assumptions that underpin neoliberal policy design, 'perfect information', 'rational actors' and 'competition-driven improvement', into realistic governance conditions.

Idealised neoliberal policy models assume perfect information (Arrow & Debreu, 1954; Friedman, 1953; Marshall, 1890/2013; Samuelson & Nordhaus, 2009): that all actors have full, accurate and comparable data with which to make optimal decisions. The use of a more pragmatic condition of information integrity rejects this abstraction, recognising that public information systems are socially and institutionally produced. Effective governance, therefore, depends on data that are accurate, comparable, intelligible and timely, enabling meaningful decision-making and accountability. Without such integrity, information asymmetries distort behaviour and weaken legitimacy.

The idealised neoliberal actor is fully rational, capable of interpreting complex information and acting logically upon it (Friedman, 1953; Robbins, 1932/2014; Samuelson & Nordhaus, 2009). In practice, people are shaped by bounded rationality: cognitive limits, heuristics and contextual cues. The pragmatic condition of actor rationality and use therefore requires policy information to be interpretable, accessible and framed in ways consistent with how real people reason and act. This aligns with behavioural insights that highlight the interpretive and affective dimensions of policy use.

Neoliberal theory also assumes that competition leads naturally to equilibrium and improvement (Arrow & Debreu, 1954; Friedman, 1953; Marshall, 1890/2013; Stiglitz, 2002), as organisations respond rationally to incentives. The more pragmatic condition of organisational responsiveness emphasises instead that improvement depends on organisational capacity, resources, leadership and the operation of the wider system. Schools and other public institutions vary widely in their capacity to adapt.

Finally, the framework aligns these conditions with broader critiques of neoliberal governance and alternative principles—fairness and equity, stability and trust, accountability for the unmeasured and system stewardship—rooted in literatures on social justice (Fraser, 2008; Rizvi & Lingard, 2010),

public value (Moore, 1995), trust-based governance (O'Neill, 2002) and systems thinking (Meadows, 2008). These principles highlight the need for governance arrangements that are ethically grounded, contextually sensitive and oriented toward long-term public value rather than narrow performance metrics.

Table 3. Policy Logic Framework for School Performance Measures (SPMs)

Section	Purpose in Framework	Relation to Policy Logic / Policy Design Literature	Key Considerations for SPMs	Application to SPMs	Original Ideas (Expanded Formulation)	Key Evaluative Questions
1. Conceptual and Ethical Foundations *(incorporating Problem Definition)*	Establishes the ideological, historical and ethical premises of the policy; defines what counts as the 'problem' and what constitutes 'success.'	Policy design and framing theory (Bacchi, 2009; Fischer, 2003); critical policy sociology; education governance studies.	Ideological roots (neoliberalism, NPM, neo-conservatism, Third Way); ethical assumptions about autonomy, control and measurement; problem framing.	Defines education as an economic investment; constructs 'school performance' as a measurable output problem; legitimises state intervention via data.	SPMs in England are grounded in neoliberal ideology and New Public Management, framing education in terms of efficiency, competition and accountability. These merge with neo-conservative and Third Way values to create a system where education is treated as an economic investment and schools operate in quasi-markets. Ethically, this narrows education to what can be measured, turning autonomy into control and fostering data-driven performativity. SPMs thus act as instruments of governance that reinforce inequalities while appearing objective. Reimagining SPMs would require a broader evaluative framework that values qualitative evidence, contextual understanding and collective accountability, restoring education's purpose as a public good.	What assumptions about the purpose of education and the nature of performance underpin SPMs? How does problem definition shape their logic and legitimacy?

Section	Purpose in Framework	Relation to Policy Logic / Policy Design Literature	Key Considerations for SPMs	Application to SPMs	Original Ideas (Expanded Formulation)	Key Evaluative Questions
2. Aims and Rationales	Articulates the intended goals and justifications of the policy; what it claims to achieve and for whom.	Policy instrument rationales (Hood, 1991; Howlett, 2011); accountability theory; performativity (Ball, 2003b).	Functional aims (choice, accountability, improvement, transparency, legitimacy); tensions between public and private rationales.	SPMs serve multiple audiences—parents, policymakers, regulators—but their competing aims create tensions.	SPMs aim to make school performance visible, comparable and actionable for multiple stakeholders. Their purposes include: (1) informing parental choice and stimulating competition; (2) supporting inspection and accountability (e.g., Ofsted); (3) enabling school improvement through self-evaluation and benchmarking; (4) guiding intervention and funding decisions; (5) promoting transparency and system learning through published data and international comparability. Evaluation of SPMs must assess whether these aims are realised or instead produce inequity, distortion and performativity.	Do SPMs achieve their stated aims, or generate unintended consequences (e.g., inequity, gaming, reputational harm)?

Section	Purpose in Framework	Relation to Policy Logic / Policy Design Literature	Key Considerations for SPMs	Application to SPMs	Original Ideas (Expanded Formulation)	Key Evaluative Questions
3. Values and Assumptions	Identifies normative and epistemic values embedded in the policy; what is prioritised and what is marginalised.	Valuation studies (Espeland & Stevens, 1998); governmentality (Miller & Rose, 2008); performativity theory.	Measurability, comparability, standardisation and productivity; implicit valuation of quantification over judgement.	SPMs equate quality with exam results, prioritising academic metrics and obscuring social and moral purposes of education.	SPMs assume that educational quality is measurable, comparable and improvable through transparency and competition. They embody neoliberal values—accountability, standardisation and quantifiable outcomes—particularly in core academic subjects. These assumptions define what counts as 'success' while neglecting contextual and broader educational aims, reinforcing inequities and narrowing the conception of education.	What values are encoded in SPMs and how do these shape what schools do, measure and value?

Section	Purpose in Framework	Relation to Policy Logic / Policy Design Literature	Key Considerations for SPMs	Application to SPMs	Original Ideas (Expanded Formulation)	Key Evaluative Questions
4. Conditions and Processes of Operation *(incorporating Feedback and Effects)*	Examines how policy instruments function in practice: the practical, social and ethical conditions that shape their operation and effects.	Implementation and learning (Pressman & Wildavsky, 1984; Dunlop & Radaelli, 2018); complex systems in governance.	Interaction between actors, incentives and structures; feedback loops; adaptation; systemic effects.	SPMs function as instruments of behavioural steering via data, reshaping trust, legitimacy and governance.	SPMs depend on market-based mechanisms—rational choice, perfect information and fair competition—which seldom exist in practice. A more realistic framework highlights the need for: (1) high-quality, transparent information; (2) measures understandable to and usable by, parents and stakeholders; (3) school capacity and fairness in accountability; (4) equitable design and communication; (5) consistency, trust and legitimacy. These underpin whether SPMs produce meaningful accountability or distortive effects.	Do SPMs operate as intended? What feedback loops, distortions, or adaptive responses emerge in practice?

Table 4. Sub-framework: Pragmatic Conditions for Neoliberal Policy Operation

Category	Condition	Idealised Neoliberal Assumption	Generalised Pragmatic Governance Condition	SPM-Specific Example	Purpose in Framework	Example Evaluative Questions related to SPMs
Operational Conditions	**1. Information Integrity**	Perfect information	Information must be accurate, comparable and intelligible to support informed decision-making and accountability.	Accuracy and usability of SPM data; publication, interpretation and media use.	Assess the informational foundations of policy legitimacy.	How reliable, transparent and comprehensible are SPM data to their users?
	2. Actor Rationality and Use	Rational actors	Policy assumes users act logically; in reality, bounded rationality means data must be interpretable and contextually relevant.	Parental interpretation and media framing of school performance.	Examine cognitive and interpretive dimensions of policy use.	How do different actors interpret and act upon SPM information?
	3. Organisational Responsiveness	Competition drives equilibrium and improvement	Improvement depends on schools' adaptive capacity, resources and system fairness.	Schools' responses to accountability pressures and rankings.	Evaluate behavioural and systemic effects.	Do SPMs promote genuine improvement or gaming and avoidance?

Category	Condition	Idealised Neoliberal Assumption	Generalised Pragmatic Governance Condition	SPM-Specific Example	Purpose in Framework	Example Evaluative Questions related to SPMs
Ethical–Structural Enabling Conditions	**4. Fairness and Equity**	Level playing field	Governance requires attention to context, disadvantage and differential capacity to compete.	Context-adjusted accountability models (e.g., contextual value added).	Assess distributive justice and system inclusiveness.	Do SPMs mitigate or exacerbate structural inequities?
	5. Stability and Trust	Self-stabilising market	Effective governance depends on predictability, institutional credibility and stakeholder trust.	Consistency of SPM frameworks and confidence in their legitimacy.	Evaluate institutional reliability and continuity.	Are SPMs trusted, consistent and viewed as legitimate by stakeholders?
	6. Accountability for the Unmeasured	Measurement completeness	Governance must protect unmeasured but valuable educational outcomes.	Supplementary accountability mechanisms (e.g., Ofsted, qualitative review).	Safeguard educational purposes beyond quantification.	What mechanisms ensure that non-measurable aspects of education remain valued?
	7. Accountability for System Stewardship	Self-regulating market	Markets need oversight to sustain capacity, diversity and public value.	Oversight of system balance (e.g., school place planning, diversity of provision).	Protect system-level coherence and sustainability.	How does the system maintain capacity and equity across schools and regions?

3.4 Conclusion to Theoretical-Analytical Frameworks

School Performance Measures are complex policy instruments. As such, their effects cannot be adequately understood from a single analytical perspective. The PLF, CDS and the FATE framework therefore function in this research as complementary lenses, each illuminating different dimensions of how SPMs are constructed, interpreted and enacted.

FATE provides a normative lens through which to evaluate whether SPM systems operate fairly, transparently and accountably and whether they adhere to ethical principles in both design and consequence. CDS complements this by shifting attention upstream, to the data infrastructures, classifications and epistemological commitments that make SPMs possible in the first place. It challenges the neutrality of metrics, exposing how data practices reflect and reproduce broader patterns of power, inequality and governance. The PLF extends this by acting as a crucial bridge, grounding the evaluation in policy analysis by critiquing SPMs as policy instruments driven by neoliberal ideology and market logic. It does this by interrogating the policy rationale and design logic embedded in SPMs, including the assumptions about what constitutes educational quality, the behavioural incentives that performance measures generate and the effects, both intended and unintended, that follow from their implementation.

Ultimately, a robust critique of SPMs requires integrating the Policy Logic (ideology and purpose) with the Data Logic (CDS) and the Fairness/Systemic Logic (FATE). By using this three-part lens, this research can address key evaluative questions—such as whether SPMs mitigate or exacerbate structural inequities—and conclude that simply improving data integrity or introducing technical fixes is insufficient. Instead, a comprehensive critique demands that the ideological framework underpinning the policy be challenged.

Together, these frameworks not only expose the limitations of current SPMs but also guide the development of credible alternatives. CDS challenges the underlying data assumptions, the PLF tests the coherence and consequences of policy design and FATE can evaluate whether proposed models meet higher standards of fairness, accountability, transparency and ethics.

4. Methodology

4.1 Introduction to Methodology

This chapter outlines the methodological foundations and research methods employed in this thesis, which examines secondary school performance measures in England: their effectiveness, their effects and considers the potential for alternative models. The chapter considers the philosophical assumptions which underpin the study, the overall mixed-methods research design, the rationale for methodological choices and the procedures for data collection and analysis.

4.2 Research Philosophy

Research paradigms shape how researchers generate knowledge and judge its quality. They set the assumptions, boundaries and expectations of inquiry—much like the rules of a game. Making these assumptions explicit is essential for evaluating the strength of findings and identifying where alternative forms of evidence are needed.

The nature of this topic—spanning statistical models, policy design, organisational behaviour and lived experience—requires engagement with both numerical data and the meanings, values and consequences attached to it. My research approach therefore developed naturally into a mixed-methods design. This was not a predetermined methodological stance, but the most appropriate response to my research questions, as well as the practical expertise and philosophical commitments I bring to the work.

Mixed-methods research is often viewed as an uneasy compromise between two ostensibly incompatible paradigms (Bryman, 1988; Guba & Lincoln, 1994; Tashakkori & Teddlie, 1998). Quantitative inquiry is typically linked to a realist ontology, a positivist epistemology and a deductive logic; qualitative inquiry, to a constructivist ontology, an interpretivist epistemology and an inductive logic (Clark, 2019). This apparent division has fuelled debates—sometimes framed as 'paradigm wars'—that question whether these philosophical traditions can meaningfully coexist (Alasuutari ewt al., 2008; Denzin & Lincoln, 2011).

These tensions matter directly for the study of SPMs. High-stakes accountability metrics are often presented as objective, technical outputs of rigorous quantitative analysis. Yet they also carry social meanings, influence behaviours, distribute power and shape how schools are judged. Understanding SPMs therefore requires both: measurement of performance patterns and critical interpretation of what those measures represent, who they serve and with what consequences.

The philosophical terms used in these debates—positivism, interpretivism, pragmatism, critical realism—are themselves variable in meaning. They may be described as paradigms, lenses, traditions, or stances, each carrying different emphases and histories. Rather than treating them as rigid categories, it is necessary to engage with the assumptions they represent and how they relate to my research purpose.

In this section, I clarify the philosophical foundations of my study. I first examine positivist and interpretivist perspectives and their conventional alignment with quantitative and qualitative research. I then consider approaches that seek to reconcile these, focusing on pragmatism and critical realism as potential bridges. Finally, I reflect on the implications for my research design and articulate the philosophical position that underpins my analysis of school performance measures.

4.2.1.1 Positivism

Positivism is a broad philosophical tradition emphasising the role of scientific method, objective observation and logical or mathematical reasoning in the generation of knowledge. It is typically associated with a realist ontology—the belief that there is a single reality independent of perception—and a deductive epistemology that seeks to uncover generalisable laws about how the world works. Within the social sciences, this has often translated into efforts to measure human behaviour using tools and logics drawn from the natural sciences (Turner, 2001).

There is no single, universally agreed definition of positivism. It has been described variously as a philosophy of science, a research paradigm, an intellectual movement and even a quasi-political stance. Historical and disciplinary variations include Comte's 'social physics,' logical positivism and Popper's falsificationism. These versions share a commitment to empirical observation and systematic testing, but differ in assumptions and priorities. The plurality of meanings can make 'positivism' both difficult to define and easy to critique.

A common characterisation in introductory sociology captures positivism's aspirational logic: scientific inquiry should uncover the laws that govern human behaviour, just as scientists reveal the laws of the physical world (ReviseSociology, 2015). In its strongest form, this implies that social life operates according to discoverable, predictable structures and that scientific enquiry can reveal stable, objective truths.

This logic has shaped the development and use of school performance measures in England. Metrics such as Attainment 8 and Progress 8 are often presented as neutral, technical outputs that provide objective insight into school quality. Positivist assumptions are embedded in the statistical

modelling, the focus on quantifiable outcomes and the search for causal explanations. These measures therefore appear to offer a scientific basis for accountability.

However, critics argue that a purely positivist stance struggles to capture complex social realities. In education, outcomes are shaped by context, culture, policy incentives and lived experience—all of which resist reduction to universal laws. The emphasis on objectivity and hypothetico-deductive reasoning risks marginalising interpretation, meaning and power and can obscure the value-laden choices involved in constructing metrics. In the case of SPMs, this may lead to oversimplified conclusions about 'effectiveness' that ignore structural inequalities and unintended consequences.

Even the analogy with the natural sciences has weakened. Modern physics acknowledges uncertainty, non-linearity and observer effects; the notion that education systems can be understood through fixed predictive laws becomes even more problematic.

Positivism has played an important role in establishing social science as an empirical enterprise. Yet its limitations highlight the value of complementary approaches when studying phenomena like school performance measures—where numbers alone cannot capture how accountability systems are constructed, interpreted and enacted in practice.

4.2.2 Interpretivism

Interpretivism refers to social science approaches that reject the assumption that social phenomena can be studied using the same tools and assumptions as the natural sciences (Bryman, 2012). The argument is not simply methodological but ontological: the subject matter of social science—human action, meaning and interaction—is fundamentally different from the physical objects studied in physics or chemistry (Moon, 2014). Social life is already interpreted by the people who participate in it; understanding it requires grasping those interpretations, not imposing external ones (Lewis-Beck et al., 2004).

Interpretivist paradigms typically adopt a constructivist ontology in which reality is actively created through perception, discourse and interaction (Lincoln & Guba, 1985; Schwandt, 1997). This aligns with an epistemology that seeks to understand how individuals and groups make sense of their world, often through qualitative and inductive inquiry (Clark et al., 2019). Knowledge is situated, partial and shaped by historical and cultural context. Because different people construct different meanings, interpretivists assume multiple realities rather than a single, universal one.

The distinction becomes clearer when considering the concept of 'reality.' In positivism, reality refers to a mind-independent material world; in interpretivism, it concerns lived meaning. These two senses are often conflated. For example, a school's purpose, culture and power relations exist

through shared understandings rather than as tangible matter. Schools persist not because of bricks and mortar, but because communities collectively enact their existence (Capper, 2018). Interpretivism focuses on those enacted meanings.

Ontology and epistemology are tightly coupled: if social reality is constructed through consciousness and interaction, then knowledge is necessarily interpretive. What we can know is inseparable from how we experience and make sense of the world.

These assumptions are highly relevant to SPMs. Such measures do not neutrally reflect school 'effectiveness'; they embody social judgements about what counts as success and who is responsible for achieving it. Their consequences—how teachers respond, how students feel labelled, how communities perceive schools—cannot be understood through statistical analysis alone. An interpretivist lens reveals dimensions of accountability systems that numbers obscure: values, perceptions, identities and power.

4.2.3 Positivism vs Interpretivism? The 'Paradigm Wars'

The idea of 'paradigm wars' refers to debates about the merits and assumptions of quantitative versus qualitative approaches. These tensions largely arise from the contrasting ontological and epistemological commitments associated with the two traditions (Bryman, 2008). The resulting binary oppositions—quantitative vs qualitative; realist vs constructivist; objectivist vs subjectivist—suggest that the paradigms are fundamentally incompatible (Lincoln & Guba, 1985).

However, this characterisation oversimplifies the complexity of research practice. Differences within each tradition can be as substantial as those between them (Bryman, 2006). Researchers across both traditions adopt diverse logics, aims and assumptions; and methods do not map neatly onto paradigms. Positivist principles can underpin qualitative designs, while interpretivist principles can inform quantitative work.

Furthermore, the paradigm binary is itself socially constructed. The idea of incompatibility has been reproduced through debate and tradition rather than reflecting inherent contradiction (Bloor, 2001).

A more productive way to view the relationship is as a set of continua (Moon, 2014). Researchers can position their ontological and epistemological commitments along these continua, leaning toward one end in some respects while drawing from the other where appropriate.

This perspective aligns closely with the study of SPMs. Accountability metrics such as Progress 8 rest on positivist assumptions about measurement and causality. Yet they also shape behaviour, carry symbolic power and influence how success is understood—phenomena requiring interpretive

insight. Recognising paradigms as continua supports research that uses statistical tools to analyse performance patterns while also exploring what those measures mean and do in practice.

From this standpoint, the 'paradigm wars' can be moved beyond by embracing methodological pluralism. The following sections consider philosophical approaches—particularly post-positivism, pragmatism and critical realism—that explicitly seek to bridge these traditions.

4.2.4 Moving Beyond the Paradigm Wars

Various philosophical approaches have sought to reconcile distinctions between positivist and interpretivist paradigms. The following sections summarise three prominent reconciliatory positions, before reflecting on their relevance for mixed-methods research into SPMs.

4.2.4.1 Post-positivism

Post-positivism is best understood as an evolution of positivist philosophy that emerged as a critical response to its earlier forms. It maintains that an external reality exists, but rejects the premise that researchers can observe it with complete neutrality. Observations are shaped by perspectives, social contexts and theoretical assumptions. Knowledge is therefore fallible and objectivity is achieved not through a mythical neutrality but through reflexivity and systematic attempts to address bias.

In the context of SPMs, a post-positivist stance allows for the continued use of robust statistical models while acknowledging that such models reflect the value-laden choices of their designers: which outcomes count, how progress is defined and which contextual factors are included or excluded.

4.2.4.2 Pragmatism

Pragmatism holds that research should be guided by what best addresses the problem at hand, rather than by philosophical allegiance to a specific paradigm (Tashakkori & Teddlie, 1998). It is commonly associated with mixed-methods designs, where focus is on the consequences of research rather than on adherence to particular methodologies (Creswell & Plano Clark, 2018). Pragmatists typically accept that knowledge is socially constructed (Morgan, 2014) and argue that debates over ultimate ontology and epistemology cannot be definitively resolved because meaning is inseparable from human experience and context (Dillon et al., 2000).

For SPM research, pragmatism supports using quantitative methods to analyse performance patterns while also using qualitative methods to understand how those patterns are interpreted and acted upon. The choice of method is determined by its utility, not by philosophical purity.

4.2.4.3 Critical Realism

Critical realism, developed by Bhaskar (1975, 1979), challenges both positivism's focus on observable facts and postmodernism's scepticism about reality. It posits an underlying causal reality - what Bhaskar terms "transcendental realism" - that exists independently of perception. However, our access to this reality is indirect, mediated through the actual events it generates and the empirical experiences through which we interpret those events.

By recognising that social phenomena are shaped by both underlying mechanisms and human interpretation, critical realism offers a structured bridge between positivism and interpretivism. It supports research that combines causal investigation (e.g., how prior attainment and resources affect outcomes) with interpretive analysis of meaning and agency.

This layered ontology is particularly useful for examining SPMs: measurable outcomes can be understood as the surface expression of deeper structures—such as socioeconomic inequality, policy incentives and institutional constraints—that shape school-level performance.

4.2.5 Discussion of Research Philosophy

The so-called 'paradigm wars' have often framed quantitative and qualitative approaches as mutually exclusive. However, Kuhn's (1962) concept of incommensurability suggests that paradigms are not adversaries but different ways of seeing. They involve distinct assumptions and logics that shape how knowledge is produced—but do not invalidate one another.

Viewed this way, quantitative and qualitative approaches offer complementary vantage points on complex social phenomena such as SPMs. Metrics are simultaneously numerical indicators and socially constructed artefacts: they are made from data, but they also make meaning—shaping accountability, behaviour and perception. Understanding how SPMs function, therefore, requires tools to interrogate both measurable outcomes and the values, interpretations and power relations behind them.

Developments in the sciences illustrate that seemingly contradictory models can coexist and, together, deepen understanding. Quantum mechanics and relativity explain different dimensions of the physical world; their coexistence strengthens insight rather than undermining it. Likewise, mixed-methods designs can generate knowledge that neither paradigm alone could provide.

Post-positivism acknowledges that knowledge is always provisional; pragmatism keeps inquiry focused on the research problem; critical realism offers conceptual tools for linking observed

outcomes to underlying causal structures. Each contributes to an integrated approach capable of capturing both what SPMs measure and what they do.

For my study, pragmatism and critical realism are particularly valuable. Both recognise that numerical performance scores reveal only part of the reality of schooling. My aim is therefore not to choose between paradigms but to strategically draw on each to explain the construction, operation and consequences of England's accountability system.

4.2.6 Conclusion to Research Philosophy

This section has reconsidered the perceived incompatibility between quantitative and qualitative approaches, showing that mixed-methods research is not only philosophically defensible but essential for analysing SPMs.

SPMs sit at the intersection of measurable outcomes and socially embedded processes. They claim to quantify school 'effectiveness,' yet their design, interpretation and impact are shaped by political priorities, institutional constraints and human behaviour. A quantitative lens can reveal patterns in national datasets—who is labelled successful, where disparities lie and how performance varies across contexts. A qualitative lens can explore how those metrics are constructed, contested and experienced by schools, policymakers and communities.

The methodological choices in this thesis are driven by the nature of the research questions:

- How do SPMs define performance?
- What factors drive variation in measured outcomes?
- How do accountability pressures shape practice and policy?

These cannot be fully addressed by a single paradigm. A mixed-methods design allows me to analyse both what the measures show and what they do.

I draw on pragmatism's principle that methods should serve the inquiry and on critical realism's focus on underlying mechanisms that shape the patterns seen in performance data. This combined stance acknowledges that SPMs reflect aspects of reality and reshape that reality through their enactment.

Ultimately, quantitative and qualitative approaches are not adversaries in investigating school performance but partners. Each contributes uniquely to a more complete understanding of how England's SPMs operate, what they illuminate and what they obscure—and it is at their intersection that the most meaningful insights emerge.

4.3 Research Design

This research uses a multi-phase mixed-methods approach that combines qualitative document analysis, advanced quantitative modelling and discursive stakeholder engagement. This design follows best practices in mixed-methods education and policy research, in which each phase builds on and deepens insights from the previous one (Creswell & Plano Clark, 2018; Tashakkori & Teddlie, 2010). The structure supports both conceptual development and empirical testing, followed by stakeholder deliberation, providing a thorough and triangulated understanding of SPMs. This design is especially appropriate when investigating complex policy instruments whose meanings, effects and assumptions are contested and multi-layered (Howlett, 2011; Schneider & Ingram, 1997).

Although the specific research design details for each phase are discussed later, here I explain the rationale behind the chosen methods for each stage, linking them to the overall research questions of this thesis. Figure 1 provides a high-level overview of the research design.

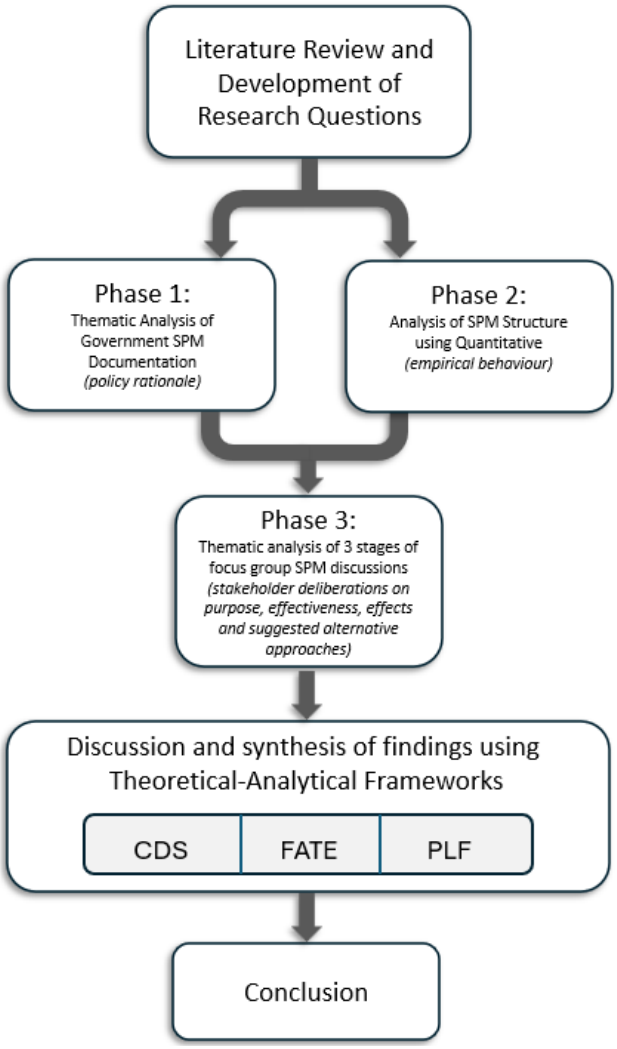


Figure 1. Overview of research design

4.3.1 Phase 1: Policy Documentation Analysis

The first phase of this research involved analysing the introduction and development of SPMs through thematic coding of government documentation related to SPMs, including ministerial speeches, parliamentary debates, government white papers, Ofsted frameworks and DfE technical guidance. The purpose of this work was twofold.

First, it aimed to gain a deeper understanding of the socio-political context of SPMs. This is crucial in revealing the principles that underpin the creation, publication and use of SPMs. Second, it consequently informs the application of the SPM Policy Logic (section 8.2), offering insights into the aims and assumptions embedded in SPMs, both explicit and implicit, which in turn provide a foundation for evaluating their effectiveness, addressing Research Question 2.

The decision to undertake a thematic analysis of government documents is grounded in the view that policy texts are central artefacts through which governments construct problems, justify interventions and signal intended mechanisms of change (Bacchi, 2009; Yanow, 2000).

Thematic coding is an established method for identifying underlying assumptions, policy logics, rationales and discursive shifts over time (Bowen, 2009; Braun & Clarke, 2006). It enables the systematic tracing of how SPMs have been framed, prioritised, problematised and legitimised by government actors. This approach supports the development of the SPM Policy Logic, reflecting methodological standards for reconstructing policy instruments and their underlying theories of change (Howlett, 2011; Schneider & Ingram, 1997).

4.3.2 Phase 2: Quantitative Statistical Modelling

For the second research phase, I used national datasets of SPMs and wider school information variables aggregated at school level and derived from publicly available UK government's KS4 performance tables. I used three advanced quantitative approaches - factor analysis, structural equation modelling and cluster analysis - to examine SPMs. This research phase had three aims. First it aimed to examine the hidden, latent constructs underlying SPMs to gain insight into the conceptualisation of 'performance' within SPMs. Second, it sought to consider the breadth of 'performance' within the current headline SPMs. Finally, it examined the extent to which SPMs encapsulate elements which are beyond the influence of schools. The approaches used, point towards potential uses to identify patterns related to, for example, common challenges faced by schools in similar contexts and so inform policy related to this.

The three methods used here - factor analysis, structural equation modelling and cluster analysis - were chosen in combination because they offer complementary insights. Their use aligns with

longstanding standards in educational measurement, in which policy instruments such as performance metrics are assessed for construct validity, dimensionality and systemic bias (Kline, 2016; Messick, 1995). Used together, they provide a novel, robust and triangulated quantitative foundation for the evaluation of SPMs.

Factor Analysis is appropriate because it identifies latent constructs and helps reveal how measures cluster conceptually. This is important for assessing how well SPMs genuinely represent 'school performance' (Fabrigar et al., 1999). Building on this, Structural Equation Modelling (SEM) allows for theory-driven causal modelling, enabling the evaluation of how much contextual factors impact performance measures, in line with best practice in, for example, educational effectiveness research (Byrne, 2016). Cluster Analysis adds the opportunity to detect otherwise obscured patterns of similarity across schools and is recommended for revealing system-level differences relevant to policy targeting (Everitt et al., 2011).

Using these methods in combination provides for quantitative triangulation, enhancing construct validity, explanatory power and interpretive robustness (Bryman, 2016; Flick, 2018). Meanwhile, the sequencing aligns with best practices in mixed-methods research. Phase 1 develops conceptual insight regarding a) the purpose of SPMs and b) their underlying policy logic, while Phase 2 tests these insights empirically (Creswell & Plano Clark, 2018).

4.3.3 Phase 3: Qualitative Focus Groups with Stakeholders

In the final research phase, focus groups with stakeholders, typically in groups of three or four, took place online in 90-minute sessions over three stages. The size of the groups and the length of the sessions provided a meeting structure which allowed each participant's ideas to be expressed and responded to. For each stage the composition of each group was changed, providing different combinations of participants to facilitate richer discussions. The sessions were organised around discussion guides with semi-structured questions. Each stage had a different focus. The first stage of focus groups explored participants' perceptions of the purposes, effectiveness and effects of SPMs. The second stage examined suggestions for alternative approaches to SPMs to address the issues identified in stage 1. Finally, the third stage revisited these suggested alternative approaches, considering how they might be prioritised and introduced and identifying any potential issues which might arise from their operation in practice.

Focus groups were chosen because they are well suited to exploring how actors make sense of policy instruments collectively, especially when dealing with socially constructed concepts such as 'performance,' fairness, accountability, or quality (Barbour, 2018; Kitzinger, 1995). Group interaction

generates ‘interactional data’ that cannot be obtained through individual interviews, allowing participants to question, challenge and refine one another's assumptions (Morgan, 1996).

The three-stage, iterative design reflects principles of deliberative qualitative methodology, where group processes are used to explore complex problems, generate alternatives and refine recommendations (Patton, 2015). Rotating group membership between stages enhances the diversity of perspectives and reduces groupthink (Bloor et al., 2001). Small groups (3–4 participants) and extended 90-minute sessions align with methodological guidance for facilitating deep engagement, allowing every participant to contribute meaningfully (Krueger & Casey, 2015).

This final phase builds directly on the earlier findings: participants, in addition to sharing their direct experience of SPMs, were invited to reflect on some of the insights which emerge from Phase 1 and Phase 2 and then to co-produce potential policy alternatives. This aligns with best practice in participatory and co-production research, which emphasises the value of stakeholder deliberation in evaluating and designing public policy (Bradbury & Reason, 2008).

4.3.4 Sequential Design

The sequential design allows insights from each phase to feed into the next, culminating in an integrated understanding of both current policy and potential alternatives. Each phase was purposefully structured to build on the one before it. The policy documentation analysis established the conceptual and historical foundations of SPMs. The quantitative modelling then examined how these ideas manifest empirically within the current system. Finally, the focus groups explored how stakeholders interpret these findings and how they might inform more workable futures. This progression, from a) policy rationale, to b) empirical behaviour, to c) stakeholder deliberation, enabled a comprehensive and coherent analysis that would not have been achievable through a single method alone.

4.4 Context of the Study

The English secondary school accountability system has, since the early 1990s, been shaped by an expanding regime of public performance comparisons centred on GCSE outcomes at Key Stage 4, evolving from simple threshold measures to increasingly complex value-added and attainment SPMs. Over time, headline SPMs such as Attainment 8, Progress 8 and the English Baccalaureate (EBacc) were introduced to promote particular curriculum priorities and to measure both raw attainment and pupil progress, profoundly influencing school behaviour, including curriculum narrowing, strategic resource allocation and pressures within teacher labour markets. The system is supported by extensive national and regional datasets, including performance tables, the National Pupil Database and the School Workforce Census, which together enable detailed policy analysis and empirical modelling of how accountability metrics shape school practices and outcomes.

4.5 Data Sampling

4.5.1 Phase 1: Policy Documentation Analysis

School Performance Measures in England are badged as 'National Statistics' through regulation by the Office of Statistical Regulation. They were introduced through legislation, in particular the Education Reform Act of 1988 (EFA) and the Information on School Examination Results (England) Regulations of 1992 and are administered by the Department for Education. The data collected for use in this analysis was therefore collected from UK Government bodies as these were perceived to be primary sources for the subject of school performance measures.

The 39 documents, listed below, collected for this analysis can be categorised into one of four groups, based on their context and intended audience.

1. Government department – the Department for Education (DfE)
2. Non-departmental government bodies – Office for Standards in Education, Children's Services and Skills (Ofsted), Office of Qualifications and Examinations Regulation (Ofqual) and Office for Statistics Regulation (OSR)
3. Legislation – the texts of the specific acts of parliament/regulations and transcriptions of the parliamentary debates when these acts/regulations were presented to parliament.
4. Government Education White Papers, Green Papers and Reports

The 39 documents can also be categorised by their purpose, with documents falling into one of four categories: a) methodology, b) advice/information, c) legislation and d) education white papers, green papers and government reports.

Table 5. Documents related to SPM Methodology (n = 7)

These documents include technical information about the calculation of SPMs as well as information about their use within the wider school accountability framework.

DfE Key stage 4 performance, Academic year 2021-22 – Explore education statistics – GOV.UK
DfE Key stage 4 performance, Methodology – Explore education statistics – GOV.UK
DfE Secondary accountability measures
Ofsted Education inspection framework for September 2023 - GOV.UK
Ofsted School inspection handbook for September 2023 - GOV.UK
Ofsted School monitoring handbook for September 2023 - GOV.UK
OSR Assessment Report KS4 Statistics

Table 6. Documents related to SPM Advice/Information (n = 4)

These documents provide advice and information to stakeholders, particularly schools and parents, about the use and interpretation of SPMs.

DfE How to understand school performance – everything you need to know - The Education Hub
DfE Understanding school and college performance measures - GOV.UK
Ofqual Analytics
Ofqual Infographics for GCSE results, 2022 - GOV.UK

Table 7. Documents related to SPM Legislation (n = 8)

These documents provide information about the legislative changes which brought about the introduction and development of SPMs.

Hansard Education Reform Bill - Hansard - UK Parliament - 2nd Reading 1 December 1987
UK Gov The 1988 Education Reform Act
Hansard Education Reform Bill (Allocation Of Time) 18 July 1988
Hansard Education Reform Bill 03 May 1988
Hansard Education Reform Bill 22 June 1988
Hansard Education (Information On School Examination Results) (England) Regulations 1991 1991-07-12
UK Gov Education (Schools) Act 1992
Hansard Schools (Self-government) (Hansard, 25 June 1996)

Table 8. Government Education White Papers, Green Papers and Reports relating to SPMs (n = 20)

These documents present the changing policy rationale, objectives and reform proposals underpinning England's system of SPMs across different periods of educational reform. White papers and government reports which did not directly relate to the use of SPMs were not included.

Better Schools (1985)	White Paper
Choice and Diversity - a new framework for schools (1992)	White Paper
Learning to Compete - Education and Training for 14-19 Year Olds (1996)	White Paper
Self-Government for Schools (1996)	White Paper
Excellence in Schools (1997)	White Paper
Schools Building on Success (2001)	Green Paper
Schools Achieving Success (2001)	White Paper
14 to 19 Opportunity and Excellence (2003)	Green Paper
Every Child Matters	Green Paper
Secondary Education (2005)	Education Select Committee Report
14-19 Education and Skills (2005)	White Paper
Higher Standards, Better Schools for All (2005)	White Paper
Childrens Plan (2007)	Policy Paper
21st Century Schools (2008)	Green Paper
Excellence and fairness Achieving world class public services (2008)	Green Paper
New opportunities Fair chances for the future (2009)	White Paper
Your child your schools our future	White Paper
School Accountability (2010)	Education Select Committee Report
The Importance of Teaching (2010)	White Paper
Educational Excellence Everywhere (2016)	White Paper

In this study, the selection of government documents was guided by principles of purposive sampling. This approach prioritises the relevance and richness of cases over what might be considered statistical representativeness (Palinkas et al., 2015; Creswell & Plano Clark, 2018). Documents were chosen to reflect a comprehensive perspective on school performance measures, including policy formation (Department for Education reports and White Papers), regulatory oversight (Ofsted, Ofqual, OSR) and legislative frameworks (acts of parliament and parliamentary debates). In practice, this was a combination of two key purposive techniques. The first technique was critical case sampling. This approach was used to select the form of documents that represent the most important or influential sources (e.g., a foundational Education White Paper or the primary Act of Parliament) (Patton, 2015). If the ideological and practical rationale for SPMs is present in these documents, it is highly likely to be representative of the system's core logic. The second is

criterion sampling, in which documents were included if they met the predetermined criterion of being official policy or legislative instruments that define, regulate or comment on SPMs. This ensures a comprehensive coverage of all relevant statutory and regulatory texts.

This approach aligns with the notion of information-rich sampling, which seeks sources that provide deep insight into the phenomenon of interest (Gentles et al., 2015). Moreover, the inclusion of diverse document types, in this case methodological guidance, advice/information, legislation and policy documents and reports, ensures that multiple dimensions of the policy process are captured, supporting triangulation (Yin, 2018), a key principle for increasing trustworthiness and interpretive validity in qualitative research (Bowen, 2009; Flick, 2018). It also affords a nuanced understanding of how school performance measures are conceptualised, implemented and evaluated. This approach ensures that the analysis is grounded in the most relevant, authentic and authoritative sources (Patton, 2015), moving beyond potentially superficial policy statements to interrogate the full policy lifecycle from ideological conception to legislative enactment and regulatory oversight. As such it aligns with best practice in policy scholarship that emphasises studying the full policy structure, not only formal legislative output (Shore & Wright, 2011).

4.5.2 Phase 2: Quantitative Statistical Modelling

All quantitative data for this research phase were sourced from the publicly available datasets via the UK Government website 'Find and check the performance of schools and colleges in England' (now renamed as 'Compare the performance of schools and colleges in England' https://www.gov.uk/school-performance-tables and

https://www.compare-school-performance.service.gov.uk/).

This includes data for 3125 mainstream state secondary schools[1] in England for the academic year 2021-2022.

I selected 13 variables from the data. These include the 6 headline performance measures and an additional 7 commonly used metrics of wider school information.

Table 9. Headline performance measures

Metric	Abbreviation used in this study
Attainment 8	Attain_8
Progress 8	Progress_8
% grade 5+ in English and maths ('Basics')	Basics
%EBacc entry	Percent_EBacc_Entry
EBacc Average Point Score	EBacc_APS
Pupil destinations	Destinations

These are designated by the Department for Education (DfE) as 'headline performance measures' to establish their greater importance regarding school accountability and to provide a basis for year-on-year comparison. They are selected and designed to provide an overview of school effectiveness, with a focus on student progress, attainment, curriculum breadth and post-16 transitions to education, employment or training. They are intended to enable a range of stakeholders to assess school performance, identify areas for improvement and support informed decision-making regarding education policy, school choice and resource allocation.

Table 10. Wider school information metrics

Metric	Abbreviation used in this study
KS2 Average Point Score	KS2_APS
% of pupils with Free School Meals ever in last six years	FSM6
% English as additional language	EAL
% Special educational needs	SEN

[1] All schools with complete data included; NAs and SUPP values removed following DfE guidelines.

school admission policy (selective or non-selective)	Selective
Whether the school has a religious character	Religious_Char
School gender of entry (mixed or single gender)	Single_Gender

These are the key items of information about the nature of schools and their intake included within the dataset. They are the main features and metrics related to schools and their student cohorts which are used to inform the context in which the school is working (Department for Education, 2023; Ofsted, 2025f, 2026). They feature as key items of wider information about schools which are used alongside school performance data in a wide range of settings, for example DfE Statistical First Releases, the DfE's 'Compare school and college performance in England' and 'Get Information about Schools' website, Ofsted Inspection Data Summary Reports and school inspection reports.

More information about the headline SPMS and Wider school information metrics can be found in Appendix 3. (Details of data elements used in quantitative analyses).

The initial dataset downloaded contains 5,768 records covering a wide range of educational establishments.

A series of inclusion and exclusion criteria were applied to isolate mainstream state secondary schools in England with complete Key Stage 4 (KS4) and destination data. The cleaning process proceeded as follows:

- Restriction to maintained schools only

Records not classified as maintained mainstream schools (e.g., special schools, local authority institutions, national institutions) were removed (n = 1,476), leaving 4,292 records.

- Removal of independent schools

Independent schools (n = 926) were excluded, reducing the dataset to 3,366 records.

- Exclusion of schools without KS4 outcomes

Schools with no Key Stage 4 results were removed (n = 121), resulting in 3,245 records.

- Matching with destination data

KS4 records were matched with pupil destination data. Schools without corresponding destination data were excluded (n = 106), leaving 3,139 records.

- Exclusion of Further Education Sector Institutions (FESI)

A final filter removed Further Education Sector Institutions (n = 14), producing a final analytic sample of 3,125 mainstream state secondary schools.

The final dataset comprises 3,125 mainstream state secondary schools in England, each with complete KS4 performance and pupil destination measures for the 2021–2022 academic year. All missing values and suppressed (SUPP) values were removed in line with Department for Education reporting guidance.

4.5.3 Phase 3: Qualitative Focus Groups with Stakeholders

The intention was to recruit the following participants to include voices from a range of different positions and perspectives with regard to SPMs:

- teachers with different periods of experience in schools
- researchers working in different fields related to SPMs
- policy makers
- school leaders and governors
- parents/carers of children of different ages

In designing the focus group sample, it was important to ensure diversity of experience and perspective, as recommended by the qualitative research literature, which highlights that focus groups work best when participants can offer contrasting yet relevant viewpoints on the focal issue (Krueger & Casey, 2015; Morgan, 1996). Given that discussion-based data emerges from interaction, sampling for maximum variation is widely recognised as strengthening the richness and credibility of the findings (Patton, 2015).

One early change that was necessary was to acknowledge that it was not going to be possible to recruit 'policy makers'. In initial communications with the Department for Education and Ofsted it was clear that they would not be able to participate. In an email received from a Senior Civil Servant in the Department for Education (personal communication, September 2024), it was noted that:

> "For propriety reasons, as civil servants, we would rather not feedback or comment on PhD work/findings, given it may have direct implications for and/or commentary on the work we are responsible for in DfE"

Such recruitment challenges are not uncommon; researchers note that access to elite or policy actors is often constrained by organisational gatekeeping and concerns about political sensitivity (Dexter, 2006; Mikecz, 2012). To better capture something of the breadth of ideas informed by close contact with the policy space, I recruited two officers from national teaching unions, a group recognised in the literature as a valuable proxy for policy-adjacent perspectives (Olson, 2010).

The primary approach to participant recruitment was purposive sampling. I directly approached some potential participants by email, using a standard invitation (Appendix 7. Advertisement for participants) attached to a brief email outlining the opportunity to contribute to research into school performance measures. These came from a variety of backgrounds and connections with SPMs. This approach was informed by research I conducted as preparation for this phase of the research project.

Purposive sampling is widely recommended for focus group research where the aim is depth, relevance and conceptual richness rather than representativeness (Palinkas et al., 2015). I also advertised for participants through social media (in particular Facebook and Twitter), focussing in particular on groups and pages which had specific interests (e.g. ECT Teachers, School Governors etc.) using the same advertisement. This therefore was not only purposive sampling but also self-selection sampling, a recognised strategy that can enhance engagement by attracting participants with a strong interest in the research topic (Robinson, 2014). I already had connections with some of the participants and so this also reflects the use of convenience sampling as part of this process, a pragmatic and commonly used approach in qualitative educational research where access and feasibility shape the sample (Etikan et al., 2016).

17 participants were recruited. A brief description of each participant is given in Table 11. I have included here not only their relationship to SPMs, but also listed whether they have children and what ages they are, as their connection with school through their children may also have informed their views of schools and SPMs. This aligns with the recommendation that researchers consider multiple positions and potential influences on participants' perspectives when interpreting focus group data (Barbour, 2018).

Table 11. Description of focus group participants

Participant No.	Role	Family	Location
Participant 1	Headteacher of secondary school (Comprehensive Academy within a MAT)	3 children (ages 5, 8 and 12)	South East, England
Participant 2	Early Career Teacher of mathematics (Comprehensive LA Maintained School)	No children	North East, England
Participant 3	Teacher of English (Comprehensive Academy)	2 children (ages 20+)	West Midlands, England
Participant 4	Researcher and statistician	No children	South East, England
Participant 5	School governor (Primary school)	4 children (ages 11, 16, 18 and 20)	South East, England
Participant 6	University researcher (economics)	3 children (ages 5, 7 and 9)	South West, England
Participant 7	University researcher (social statistics)	2 children (ages ? and ?)	South West, England

Participant No.	Role	Family	Location
Participant 8	Lecturer in education	No children	North West, England
Participant 9	Researcher (education)	1 child (aged 9)	South West, England
Participant 10	Assistant Headteacher of secondary school (Comprehensive Academy within MAT)	No children	East Midlands, England
Participant 11	Educator and Chair of Trustees of primary school (Academy)	2 children (ages ? and ?)	East Midlands, England
Participant 12	Education consultant (former CEO of MAT)	2 children (ages 30+)	Northern England
Participant 13	University lecturer in education and MAT trustee	2 children (aged 30+)	North East, England
Participant 14	Officer in teaching union	No children	South East, England
Participant 15	Officer in teaching union	No children	South East, England
Participant 16	Educator and social scientist	2 children (ages 8 and 10)	West Midlands, England
Participant 17	PhD student (Education)	1 child (aged 10)	Northern England

Contextual information about participants, including their professional roles, broad geographical region and (where applicable) the type of secondary school in which they worked, is provided in Table 11 to aid interpretation of the findings while preserving anonymity. Information regarding the performance status of participants' schools has not been included. Such classifications are themselves contingent upon the SPMs that form the subject of this research and would therefore risk importing contested assumptions about school performance into the description of participants and their contexts.

The composition of the focus groups is shown in Table 12. Focus groups were deliberately reconfigured between stages to expose individuals to a range of perspectives and experiences, while maintaining continuity across the three phases of the research. This approach was intended to enrich discussion by encouraging participants to engage with different viewpoints and to stimulate

reflection on ideas emerging from earlier sessions. The focus groups were not designed to be statistically representative. Instead, participants were brought together in varying combinations to maximise the exchange of perspectives and experiences relevant to the exploration of school performance measures and alternative approaches.

Additionally, feedback from participants indicated that they valued the opportunity to meet and discuss issues with individuals from different professional and stakeholder backgrounds. The changing composition of groups may also have contributed to participant retention across the study, which required a substantial commitment of approximately 4.5–5 hours per participant over the three phases.

Table 12. Participants included in each focus group

Focus group stage	Group number in stage	Participants
1	1	1, 3, 5
1	2	5, 8, 11,17
1	3	4, 6, 16
1	4	10, 12, 13, 15
1	5	2, 7, 9
2	1	4, 7, 16, 17
2	2	3, 5, 10
2	3	1, 6, 9, 14
2	4	2, 11,15
2	5	12, 13,18
3	1	7, 11,15
3	2	4, 6, 14
3	3	5, 9, 10, 12
3	4	1, 3, 13, 16
3	5	2, 8, 17

4.6 Data Collection Methods

4.6.1 Phase 1: Policy Documentation Analysis

Policy documents relevant to secondary school performance measures were systematically collected from publicly available sources, including the Department for Education (DfE), Ofsted and parliamentary archives. These included guidance, inspection frameworks, policy reports, legislative records, government white papers and consultation papers. Care was taken to collect documents over the full period since the introduction of SPM. Documents were selected based on their relevance to SPMs, with priority given to materials that explicitly addressed metric design, accountability frameworks or intended behavioural outcomes. All collected documents were organised and catalogued in a reference database to facilitate subsequent coding and analysis.

4.6.2 Phase 2: Quantitative Statistical Modelling

In Phase 2, quantitative analysis was conducted using national datasets relating to secondary school performance. These were downloaded from the UK Government's Data were cleaned and merged using the Unique Reference Number (URN) for each school to ensure accurate integration across sources. SPMs were standardised to facilitate comparability, while context variables — including % Free School Meals (FSM) eligibility, % English as an Additional Language (EAL) status, % Special Educational Needs (SEN) designation and prior attainment as indicated by the Key Stage 2 Average Point Score— were constructed to account for differences in student intake. All identified outliers were retained, as they did not represent extreme values, and some variability reflects the calculation of school performance measures. Schools with missing or suppressed data (mostly new schools or schools with very small cohorts) were excluded from the analyses. Together, these steps allowed for robust statistical modelling, enabling exploration of the extent to which observed outcomes reflect school provision and/or contextual factors.

4.6.3 Phase 3: Qualitative Focus Groups with Stakeholders

All of the focus group discussions took place online using the Google Meet platform. 90-minute meeting time was allocated for each session. Between five and six focus group meetings were conducted for each stage. The meetings were digitally recorded in two ways. I used Google Meet's video recording functionality and audio recordings were also made using a digital recording device. A transcript of each meeting was created using Google Meet. During the discussions, I also made some brief notes of pertinent points which may require further attention, either during the meeting or as part of the analysis stage.

4.7 Analytical Methods

4.7.1 Phase 1: Policy Documentation Analysis

This phase presents an analysis of policy documentation related SPMs. The analysis was conducted using thematic analysis, following the approach outlined by Braun and Clarke (2006), which allows for the systematic identification, analysis and reporting of patterns within qualitative data. This method is particularly suited to exploring the underlying assumptions, priorities and conceptual framings present in SPM policy documents.

4.7.1.1 Familiarization

The first stage of the thematic analysis involved closely reading the texts and annotating them, where appropriate, to begin to identify the key ideas which might help to address the research questions. Annotation was important at this stage because of the technical nature of the texts, the different contexts in which they were produced and the length of some of the documents. The detailed annotations to the texts both supported greater understanding of their content and were used when creating codes.

The texts were imported into the qualitative data analysis software NVivo. This software was used to further explore the texts using word frequency queries, the creation of word clouds, tree maps and cluster analyses (including 2- and 3-dimensional cluster maps, dendrograms and circle maps). These methods proved useful to gain an overview of the documentation. However, the specific outputs of these analyses were not ultimately used within the coding process. This was mainly due to their reliance upon word/phrase frequency. Word/phrase frequency means little in the context of this work due to the significantly different documents, contexts and purposes of the data used. This meant that instead close reading was vital to establish a strong understanding of the individual texts to allow identification of significant codes/themes within them.

4.7.1.2 Coding

Two approaches were used in the process of creating codes.

The primary approach was deductive. This was driven by the research questions I sought to examine. This includes both 1. the wider socio-political context of the production, publication and operationalisation of English SPMs and 2. the design and use of SPMs, including their stated aims and purposes.

Some codes were also created using a more inductive approach. This was achieved by closely examining the texts and identifying codes from the content, classifying meaningful and relevant concepts and coding the texts within the documents accordingly.

The creation of initial codes was followed by detailed refining, reviewing and revising of these codes and creation of new codes where necessary. This included beginning an exploration of the relationships between the codes as they were formed and the creation of subcodes where appropriate.

4.7.1.3 Generating, reviewing and defining themes

In the second stage of the analysis, the codes were then used to generate the themes. Due to the deductive approach employed, some ideas for themes had been identified even before the coding stage, drawing on insights from the existing literature.

The researcher's positionality is central to all forms of qualitative analysis (McCorkel & Meyers, 2003; Naples, 2003). It was important to recognise the personal and professional experience of the researcher and to distinguish between findings based solely on this experience and those which are based in the examination of the texts.

4.7.2 Phase 2: Quantitative Statistical Modelling

The methods used in this work are exploratory factor analysis (EFA), confirmatory factor analysis (CFA), structural equation modelling (SEM), correlation analysis and hierarchical cluster analysis approaches.

4.7.2.1 Exploratory Factor Analysis

Exploratory factor analysis (EFA) was first used to identify any latent constructs (or 'factors') underlying the set of metrics related to schools. I used the 'psych' and 'lavaan' R packages to do this.

The number of factors was identified using a range of techniques, including the visual examination of a scree plot using Kaiser's rule. A parallel analysis scree plot (Figure 2) and the outputs of a PCA conducted using all 13 variables (Table 12) suggested that there are either 3 or 4 latent variables which the 6 headline SPMs respond to.

The parallel scree plot (Figure 1) shows plots for three sets of data:

1. the real dataset being analysed,

2. a simulated dataset which was randomly generated to mimic the structure of the real dataset and so have a distribution of eigenvalues which that would be expected by chance (i.e. a distribution of eigenvalues which match the null hypothesis of there being no underlying factors in the data)
3. a resampled dataset made up of permutation samples from the real dataset.

There were eigenvalues of the factors from the actual data that were higher than those obtained from the simulated or resampled data (i.e. they exceeded the 95th percentile of the eigenvalue distribution from the simulated or resampled data). This indicated that the factors extracted from the actual data are likely meaningful and not simply due to chance. I used the point at which the eigenvalues from the actual data cross fall below the eigenvalue distribution from the simulated or resampled data to indicate the number of factors to retain. In this case it again indicated either 3 or 4 factors.

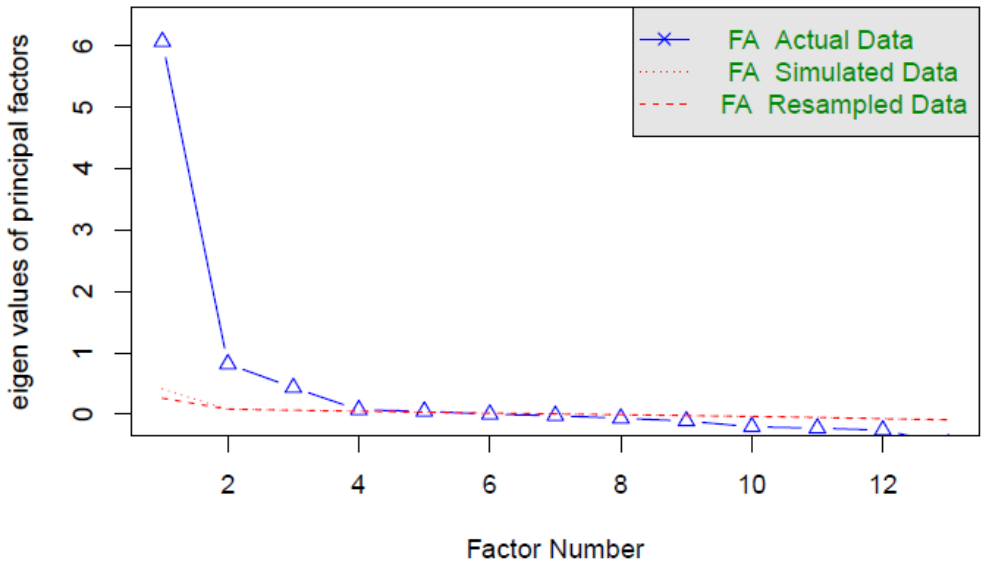


Figure 2. Parallel Analysis Scree Plot for dataset of 13 variables (6 headline SPMs and 7 school information metrics) for 3125 mainstream state secondary schools in England, academic year 2021-2022.

Table 13. Eigenvalues of the principal components arising from Principal Component Analysis (PCA) of the study dataset, indicating 3 principal components with eigenvalues greater than 1 (in green)

Principal Component	Eigenvalue
1	6.369
2	1.570
3	1.013
4	0.943
5	0.812
6	0.634
7	0.553
8	0.479
9	0.296
10	0.244
11	0.068
12	0.016
13	0.003

I conducted the EFA for the different numbers of factors in the range 2-5 and compared the fit measures (Comparative fit index (CFI), Tucker Lewis Index (TLI) and Root Mean Square Error of Approximation (RMSEA)). These collectively suggested the use of 3 or 4 factors within the EFA.

I then ran a PCA to gain eigenvalues so that they could serve as a criterion for selecting the factors which capture the underlying structure of the data. The output of the PCA (table 1.), using the Kaiser Criterion (Kaiser, 1960), suggested 3 factors, with 3 principal components having eigenvalues above 1. However, there was also a 4^{th} principal component with an eigenvalue very close to 1 (PC4 - 0.943). After running a 4 factor EFA I discovered that it resulted in a redundant factor (Factor 4) which had only very low factor loadings (Table 13).This therefore confirmed that 3 factors was most appropriate for this dataset.

Table 14. Factor loadings for 4 factor EFA using 6 headline SPM variables

	Factor 1	Factor 2	Factor 3	Factor 4
Attain_8	**0.679**	0.515	0.308	0.417
Progress_8	0.368	**0.837**	0.251	0.137
Percent_EBacc_Entry	0.236	0.205	0.**851**	0.107
EBacc_APS	**0.617**	0.459	0.495	0.399
Basics	**0.657**	0.485	0.316	0.379
Destinations	**0.609**	0.215	0.18	0.000
	Factor 1	**Factor 2**	**Factor 3**	**Factor 4**
SS loadings	1.836	1.5	1.259	0.51
Proportion Var	0.306	0.25	0.21	0.085
Cumulative Var	0.306	0.556	0.766	0.851

The EFA that I the conducted for data related to all 3125 schools in the dataset and for the 6 headline performance measures, allowed me to explore (i) the relationships between these metrics and their underlying factors and (ii) how they shape the concept of 'performance' within SPMs. I used varimax rotation in order to maximise the shared variance and so more specifically identify the factor upon which the data loaded.

4.7.2.2 Confirmatory Factor Analysis and Structural Equation Modelling

I used confirmatory factor analysis (CFA) and structural equation modelling (SEM) to explore and validate various models. These models had increasing levels of complexity in the relationships among some or all of the manifest variables and factors.

CFA was first used to test and confirm the underlying factor structure for the set of 13 observed variables, which was identified using the EFA techniques described above.

SEM was then used to extend the CFA by not only examining the relationships between observed variables and their underlying factors, but also allowing the simultaneous examination of models that incorporated numerous latent and observed variables and various types of relationships (e.g., direct, indirect, mediating and/or moderating).

A series of models was developed using CFA/SEM outputs and fit measures, alongside theory related to school metrics, to consider the extent to which SPMs represent elements beyond the influence of schools.

The primary limitation of SEM is its reliance on causal assumptions (Bollen & Pearl, 2013). While it can test relationships, it cannot definitively prove causality. As a consequence, the model is only as strong as its underlying theoretical rationale. I used a literature-based set of model fit indices (Hu &

Bentler, 1999; Kyndt & Onghena, 2014) to evaluate, compare, modify, refine and validate CFA and SEM models and the relationships between the manifest and latent variables that they represented, with acceptability thresholds set based on best practice in prior literature (Table 14).

Table 15. Model fit indices used in CFA/SEM analysis and development of models.

Model fit index	**Acceptable value for the purpose of this study**	**Description of model fit index and best practice literature referencing acceptability thresholds**
Goodness of fit index (GFI)	A value greater than 0.90	Measures the proportion of variance and covariance in the data explained by the model, with higher values (closer to 1) indicating better fit. Simulation studies (Cho et al., 2020; Sivo et al., 2006) suggest that the GFI can effectively distinguish between correct and mis-specified models with a cut-off value of 0.9
Non-Normed Fit Index (NNFI), also known as the Tucker Lewis index (TLI)	A value greater than 0.95	Compares the fit of the specified model to a null model, adjusting for model complexity; values closer to 1 indicate a better fit. Hu and Bentler (1999) have suggested NNFI ≥ 0.95 as pointing towards goo fit.
Comparative fit index (CFI)	A value greater than 0.90	Compares the specified model with an independent (null) model while accounting for sample size; values closer to 1 indicate a good fit. Studies have shown that a value greater than 0.90 (e.g., CFI ≥ 0.95) is needed in order to ensure that mis-specified models are not accepted (Hu & Bentler, 1999).
root mean square error of approximation (RMSEA)	Value Less than 0.10	Estimates how well the model fits the population covariance matrix per degree of freedom; lower values indicate a better fit. Browne and Cudeck (1993) recommended that RMSEA ≤ 0.05 would indicate a close fit of the model in relation to the degrees of freedom," and Hu and Bentler (1999) suggested that an RMSEA smaller than .06 and a CFI and TLI larger than .95 indicate relatively good model–data fit in general

These techniques allowed me to further develop an understanding of the complex interplay variables between the underlying factors and the manifest variables and to extend our conceptualisation of the factors.

4.7.2.3 Multiple regression analysis of school context and school characteristics variables

To further explore the relationship between school context and headline performance measures, an additional multivariable regression analysis was undertaken. This analysis was conducted as a robustness check for patterns identified elsewhere in the quantitative analysis.

Unlike the preceding SEM analyses, which sought to identify latent constructs underlying school performance measures, this analysis treated all contextual and school character variables as observed predictors entered simultaneously into the model. The aim was to estimate the unique association of each predictor with each performance measure while holding all other predictors constant.

The school context variables included are:

- KS2 Average Point Score
- % of pupils with Free School Meals ever in last six years
- % English as additional language
- % Special educational needs

The school character variables included are:

- school admission policy (selective or non-selective)
- Whether the school has a religious character
- School gender of entry (mixed or single gender)

Six headline outcome measures were modelled:

- Attainment 8
- Progress 8
- % grade 5+ in English and maths ('Basics')
- %EBacc entry
- EBacc Average Point Score
- Pupil destinations

The analysis was estimated within an SEM framework using observed variables only. This approach is algebraically equivalent to estimating multiple linear regressions simultaneously and permits estimation of relationships across multiple outcome variables while accounting for covariance among outcomes. Standardised coefficients are reported to facilitate comparison of effect sizes across measures.

This regression analysis was intended primarily as a robustness check and was subsequently compared with the cluster analysis to assess whether both approaches identified similar contextual patterns.

4.7.2.4 Correlation Analysis

Correlation analysis, in the form of correlation matrices, was used to examine the associations between SPMs. This provided some important insights into both a) the conceptualisation and b) the breadth and limitations of the concept of school 'performance' which is suggested by the selection of headline school metrics and the individual design of these metrics.

4.7.2.5 Cluster analysis

Agglomerative hierarchical clustering analysis was conducted on a sample of 3,125 mainstream state secondary schools in England. The analysis used z-score standardised (mean-centred and variance-scaled) versions of seven variables capturing wider school characteristics. The aim of this analysis was to assess the potential of unsupervised learning approaches to provide new insights into patterns in the school information and performance dataset.

I used Euclidean distance as the measure of similarity between data points and Ward's linkage to determine cluster proximity. Euclidean distance is useful because it provides an intuitive, geometry-based measure of similarity which works well with standardised continuous variables (Everitt et al., 2011). Ward's linkage builds on this by merging clusters in a way that minimises within-cluster variance, producing compact, well-separated and interpretable clusters (Kaufman & Rousseeuw, 2005; Murtagh & Legendre, 2014). Euclidean distance together with Ward's linkage is one of the most common combinations for hierarchical clustering. This increases reproducibility and ensures that any new insights are grounded in techniques which are consistently recognised as best practice, thus strengthening both credibility and potential policy relevance.

4.7.3 Phase 3: Qualitative Focus Groups with Stakeholders

This phase presents a qualitative analysis of focus group discussion transcripts with participants exploring school performance measures. The analysis was conducted using thematic analysis, following the approach outlined by Braun and Clarke (2006), which provides a systematic framework for identifying, analysing and reporting patterns within qualitative data. This method enables an in-depth exploration of participants' perspectives, experiences and interpretations, allowing the study to capture the nuanced ways in which SPMs are understood and discussed.

4.7.3.1 Familiarisation

The transcripts of the focus groups were carefully checked for accuracy using the recordings. After this stage the transcripts were anonymised and recordings were deleted, in line with the research data management plan for this project.

The transcripts were then imported into the qualitative data analysis software NVivo. A range of analytical tools in NVivo was used to explore the transcripts' content. This included the use of word frequency queries, the creation of word clouds, tree maps and cluster analyses. These methods proved useful to gain an overview of the content of the discussions. This established a solid understanding of the transcripts which in turn provided a basis upon which to identify significant codes/themes within them.

4.7.3.2 Coding

The process of generating codes and, from that, identifying themes used two main approaches.

The initial approach was deductive. This was driven by the specific research questions I sought to examine and the questions and prompts which shaped the discussions in the focus groups. For this first round of focus groups, the aim was to explore three aspects of SPMs. These are:

1. The purpose of SPMs,
2. The effectiveness of SPMs in meeting their intended purpose and
3. The wider effects of SPMs.

These concepts therefore form the primary codes of the responses from participants and the discussions which arose from them.

4.7.3.3 Generating, reviewing and defining themes

Subcodes, based around common themes in participant responses and conversations were created using an inductive approach. To achieve this I closely examined the transcripts and identified codes

from the content and context. These collected together important related concepts which were coded accordingly.

A further development of the codes and subcodes followed the process of refining, reviewing and revising these codes and the creation of new subcodes where appropriate.

These themes and subthemes will form the basis for the 2nd and 3rd rounds of focus groups in which participants will be encouraged to explore strategies to address the identified 'issues' with SPMs and plans for alternative approaches to them. An ideas document, designed to inform and stimulate discussion in the 3rd round based on the themes emerging from the 1st and 2nd rounds of focus groups, can be found in Appendix 5.

4.7.4 Mixed-Methods Integration

This study uses a mixed-methods design in which each phase informs, challenges and strengthens the others. The qualitative policy document analysis identifies the key ideas, assumptions and contextual factors underpinning SPMs, generating initial conceptual insights into how these measures are designed, justified and operationalised. The quantitative statistical modelling then tests these ideas empirically, using EFA, CFA, SEM, correlations and hierarchical clustering to examine the latent structure of SPMs, the relationships between metrics and the extent to which they reflect factors beyond schools' control. These analyses provide evidence about how 'performance' is constructed in practice and reveal the strengths and limitations of the metrics used. The qualitative focus groups add a further layer, capturing how stakeholders interpret, experience and evaluate SPMs and generating grounded insights into their perceived purpose, effectiveness and wider effects.

Together, these phases allow for triangulation: areas of convergence strengthen confidence in the findings while areas of tension prompt further investigation or refinement. The integration of qualitative interpretation with quantitative statistical modelling ensures that neither statistical patterns nor textual meanings are taken at face value but are understood in relation to each other. Overall, this approach produces a more comprehensive, credible and policy-relevant account of how SPMs function and, crucially, how they might be redesigned.

4.8 Ethical Considerations

This research involved multiple phases, each raising distinct ethical considerations. Across all phases, the study was guided by principles of reflexivity, transparency, confidentiality and respect for participants, ensuring that methodological and analytical decisions were ethically robust.

4.8.1 Phase 1: Policy Documentation Analysis

In the first phase, a reflexive approach was applied to the analysis of policy documentation. I considered my own positionality, including prior experiences, professional knowledge and preconceptions about school accountability systems, in interpreting the material. Reflexivity was particularly important given my extensive insider knowledge of the English school system, which could influence the selection, reading and interpretation of policy documents. All analysis was conducted on publicly available documents, so there were no direct participants involved and no identifiable personal data were used.

4.8.2 Phase 2: Quantitative Statistical Modelling

The second phase utilised administrative data on English secondary schools, drawn from publicly available datasets. Ethical considerations focused on transparency in analytical decisions, including variable selection, handling of missing data and modelling choices. All data were anonymized at the school level and no individual students or teachers were identifiable. This research adhered to principles of open and reproducible analysis, with documentation of all steps to allow scrutiny of methodological choices and to support integrity and accountability in data use.

4.8.3 Phase 3: Qualitative Focus Groups with Stakeholders

Phase three involved focus groups with school leaders, teachers and other educational stakeholders. The study followed standard ethical procedures for human subjects research. Ethical approval was granted according to the University of Sheffield procedure for research ethics, on 14 June 2023 (University of Sheffield Ethics Application Reference Number 052314). Participants were provided with full information about the study, including its aims, the voluntary nature of participation and the intended use of the findings. Informed consent was obtained prior to participation and participants were explicitly reminded of their right to withdraw at any stage without consequence. Measures were taken to ensure professional confidentiality, including pseudonymisation of transcripts and careful reporting of findings to prevent identification of individuals or schools.

In addition, I engaged in reflexive consideration of my own positionality, acknowledging how my prior experiences and professional perspectives might shape interactions with participants and

interpretations of qualitative data. Reflexivity was also applied to the facilitation of discussions, ensuring that power dynamics were managed sensitively and that participants felt able to express their perspectives freely.

4.9 Reliability, Validity and Trustworthiness

This section outlines the strategies employed to ensure reliability, validity and trustworthiness across the three phases of the research, as well as in the integration of mixed-methods findings. Each phase employed techniques appropriate to its methodological approach to support confidence in the findings.

4.9.1 Reflections on Research Quality and Rigour

Given the mixed methods nature of this study, considerations of research quality and rigour were addressed across both the qualitative and quantitative components of the research design. In qualitative inquiry, quality cannot be reduced solely to reliability or replicability in a positivist sense, but instead involves broader considerations of credibility, transparency, reflexivity, coherence, and contribution (Tracy, 2010).

This research sought to demonstrate methodological rigour in several ways. First, the study employed methodological triangulation through the integration of policy analysis, quantitative analysis of headline SPMs and qualitative focus groups exploring SPM effectiveness and effects and exploring alternative approaches to them. The use of multiple methods enabled the research questions to be examined from different perspectives and helped strengthen the credibility of the findings.

Second, transparency was supported through detailed documentation of the analytical processes used throughout the thesis. This included explicit discussion of data sources, variable selection, statistical modelling decisions, clustering procedures, and thematic analysis processes. Where interpretive decisions were made, these are described and justified within the relevant chapters.

Third, reflexivity was important given the politically contested nature of school accountability and performance measures in England. The thesis recognises that performance measures are not neutral technical instruments but are socially and politically constructed. Reflexive consideration was therefore given to the assumptions embedded within existing accountability systems, as well as to the researcher's own positioning and interest in exploring alternative approaches to evaluating school performance.

The study also aimed to achieve what Tracy (2010) describes as "meaningful coherence" between the research questions, theoretical framing, methodological choices, and interpretations. The combination of complexity theory, sociological perspectives on accountability, quantitative

modelling, and qualitative exploration was intended to provide a coherent framework for examining both the technical functioning and wider consequences of secondary SPMs in England.

On the quantitative side, research quality was also informed by established standards relating to statistical validity, reliability, and robustness. In particular, attention was given to issues of model specification, treatment of missing data, and the assumptions underpinning regression-based and multilevel approaches to educational data analysis (Gelman & Hill, 2007; Shadish, Cook, & Campbell). This is particularly pertinent in the analysis of school performance measures, where estimated school effects may be sensitive to compositional context, measurement error and modelling decisions, and where different specifications can yield substantively different.

Finally, the research sought to make a substantive contribution by combining empirical analysis with conceptual and methodological critique, while also exploring possible alternatives to existing accountability arrangements. In this sense, the study aimed not only to evaluate current performance measures, but also to contribute to wider debates about the meaning and purposes of school accountability.

4.9.2 Phase 1: Policy Documentation Analysis

In the policy documentation analysis, dependability was enhanced through a transparent and systematic coding process. Coding decisions were documented and justified to allow replication and review. Reflexivity was applied throughout, I explicitly reflected on my own positionality, prior knowledge and values, acknowledging how these could shape the interpretation of policy texts. Triangulation across multiple sources of policy documentation further strengthened trustworthiness.

4.9.3 Phase 2: Quantitative Statistical Modelling

In the quantitative phase, reliability and validity were supported through rigorous statistical procedures. For structural equation modelling (SEM), model fit indices (e.g., CFI, RMSEA, SRMR, see Table 14) were assessed to evaluate the adequacy of model specification. Robustness checks were conducted to examine sensitivity to analytical decisions, including alternative model specifications and variable operationalisations. Multicollinearity among predictors was assessed and mitigated to ensure stability of parameter estimates. Verification of clustering stability and additional sensitivity analyses are acknowledged as areas for potential further work, providing transparency about the limits and reliability of the current models.

4.9.4 Phase 3: Qualitative Focus Groups with Stakeholders

Trustworthiness in the qualitative phase was supported through several strategies. Credibility was enhanced via thick description, verbatim participant quotes and careful presentation of context, allowing readers to assess the plausibility of interpretations. Dependability was supported through a transparent coding and thematic analysis process, with documentation of coding decisions. Reflexivity was applied through continual reflection on the researcher's positionality, values and prior experiences and how these might shape interactions, interpretations and analytic judgments.

4.9.5 Mixed-Methods Integration

The integration of findings across phases addressed reliability and validity through the systematic examination of convergence and divergence between quantitative, qualitative and policy data. Inferences were drawn across levels and methods, highlighting where data aligned or diverged and reflecting critically on the implications of these patterns for understanding school performance measures. This integrative approach supports a more comprehensive, trustworthy understanding of SPMs in practice, policy and perception.

4.10 Limitations of the Methodology

This study was conducted in three sequential phases: Phase 1 -Policy Documentation Analysis ; Phase 2 - Quantitative Statistical Modelling; and Phase 3 - Qualitative Focus Groups with Stakeholders. As with any mixed-methods research design, a small number of limitations are acknowledged. These are outlined selectively below, focusing on those most relevant to the interpretation of findings.

4.10.1Phase 1: Policy Documentation Analysis

The thematic analysis of SPM documentation is necessarily shaped by the availability and accessibility of policy texts, consultation documents and official publications. While these sources provide a robust account of the formal development and stated rationales of secondary school performance measures in England, they offer only a partial view of the policy process. In particular, informal negotiations, internal deliberations and unpublished influences on policy formation are not directly observable through documentary analysis alone. This limitation is inherent to policy research of this kind and is mitigated, in part, by drawing on a wide range of sources across multiple time periods to identify patterns, continuities and shifts in policy discourse.

4.10.2Phase 2: Quantitative Statistical Modelling

The quantitative phase relies on secondary administrative datasets collected for accountability and operational purposes rather than for research. As a result, the analysis is constrained to variables that are routinely recorded, which may omit relevant aspects of school practice, organisational culture, or informal processes that contribute to educational outcomes. In addition, administrative data may contain reporting errors, suppressed values, or inconsistencies in recording practices across schools and over time, which may introduce bias into the findings.

More broadly, the use of school-level performance measures necessarily captures outcomes rather than the full range of processes that generate them. While modelling and statistical techniques allow for the examination of relationships between context and performance measures, they cannot fully disentangle causal mechanisms or account for unobserved factors operating within schools. These limitations are addressed through careful data cleaning, transparent modelling decisions and the interpretation of findings as indicative of structural patterns rather than definitive measures of school effectiveness.

4.10.3Phase 3: Qualitative Focus Groups with Stakeholders

This qualitative phase draws on focus groups to explore stakeholder perspectives on school performance measures and alternative approaches. Participants were not intended to be statistically representative of all stakeholder groups, and the findings should therefore not be interpreted as

generalisable in a probabilistic sense. Instead, the focus groups were designed to elicit depth, diversity of viewpoints and critical reflection, particularly in relation to experiences of accountability and performance measurement.

As with all qualitative research, the data reflect the perspectives of those who chose to participate and the dynamics of group discussion, which may shape how views are expressed. This limitation is mitigated through careful facilitation, reflexive thematic analysis and the integration of qualitative insights with findings from the documentary and quantitative phases. Together, the three phases provide complementary perspectives that strengthen the overall credibility and coherence of the study.

4.10.4Mixed-Methods Integration

Overall, while each phase has distinct methodological constraints, the mixed-methods design enables triangulation across historical, quantitative and qualitative evidence. This combination allows the study to balance breadth and depth, and to situate empirical findings within their wider policy and theoretical context.

4.11 Conclusion to Methodology

This chapter has outlined the theoretical and analytical foundations, methods and frameworks employed in this research. By combining policy logic analysis, quantitative modelling and qualitative focus groups, situated within the lenses of Critical Data Studies (CDS) and computational FATE research, the design enables a nuanced examination of English school performance measures. The methodology allows the study to interrogate both the empirical realities of SPMs and the policy assumptions, incentives and ethical implications embedded within them. By integrating multiple data sources and methods, the research is well-positioned to provide a rigorous, reflective and critical account of the operation and effects of SPMs across multiple levels of the educational system.

5. Findings from Phase 1: Thematic Analysis of Documents Related to Secondary School Performance Measures in England

In this analysis, various UK government documents related to English Secondary School Performance Measures (SPMs) were collected. These documents were then analysed thematically using the approach outlined by Braun and Clarke (2006).

This work aims to address Research Question 1 by providing insight into the socio-political context of the production and publication of secondary school performance measures in England. It also is important in providing insights into the explicit and implicit aims of SPMs, thereby providing a foundation upon which to begin to address Research Question 2. In order to focus sharply on these aims I have created 2 sub-research questions for this specific research phase. They are:

> a) What political, social and economic factors have influenced the development and implementation of secondary SPMs in England?
>
> b) How are the purposes and design features of secondary SPMs in England articulated and explained?

Addressing these questions directly supports this thesis's original contribution by enabling a holistic evaluation of SPMs that integrates the socio-political contexts from which they emerged with their policy logic, design features and stated purposes. By examining both the forces shaping SPMs and how their aims and design are articulated, this phase provides the necessary foundation for assessing the coherence between the ideas underpinning SPMs, their construction and their subsequent interpretation, use and effects. In doing so, this research responds to the current scarcity of empirically grounded accounts that trace how ideological, administrative and economic considerations have become embedded in the purposes, design and communication of SPMs and how these foundations continue to shape their operation within the English secondary school system.

Analysis of the documents revealed four main themes related to SPMs. These are:

(i) the ongoing influence of neoliberalism

(ii) the impact of traditional neo-conservative thinking

(iii) increased emphasis on Social Democracy

(iv) aims and purposes of SPMs

This last theme (iv) contributes to addressing Research Question 2, which examines the effectiveness of the production, publication and operationalisation of secondary school performance measures in England in meeting their stated aims within this context.

Appendix 1 details the coding frequency within documents details the coding frequency within documents all documents used in the analysis.

5.1 Socio-political context (Research Question 1)

5.1.1 The continuous influence of 'neoliberalism'

Neoliberalism appears as a central theme throughout all the analysed documents and it is the dominant theme concerning SPMs within them. Several subthemes related to neoliberalism in the context of school performance measures were identified and form a continuous thread across all the documents examined. The following section presents these ideas: standards, quasi-markets, economic focus and autonomy. They are used in reference to:

- the **purpose** of SPMs, e.g. "the crusade for higher standards a reality in every classroom" (Excellence in Schools, 1997, p.4)
- the **use** of SPMs, e.g. "have sound, consistent, national measures of pupil achievement for each school at each Key Stage" (Excellence in Schools, 1997, p.27)

While many researchers have identified the neoliberalisation of education policy (Ambrosio, 2013; Ordofa & Asgedom, 2022), there are no published studies that analyse the broad range of secondary SPM documents used here which span multiple policy bodies, legislative sources and government terms.

Existing research on neoliberalism in education often focuses on broad features of policy, such as the marketisation of schools, competition and accountability, or how neoliberal logics shape teacher professionalism and curriculum practices (e.g., Ball, 2003b; Ozga, 2009), but does *not* provide the same document-by-document, extended historical analysis of the specific texts that shape SPMs. For example, Ambrosio (2013) and others have shown how neoliberal rationales underpin accountability mechanisms generally, linking market principles to performance reporting, data use and competition. Similarly, studies of neoliberalism in education policy often describe overarching trends such as performativity, managerialism and market logic (e.g., Ball, 2003a; Ozga, 2009), yet their focus is on themes and mechanisms generally, not on the detailed analysis of policy texts establishing, legitimising and stabilising performance measures across decades.

Wilkins' (2018) work on neoliberal citizenship and education policy discourse analyses some policy texts to show how parents are positioned as 'consumer-citizens' within the broader neoliberalisation processes. However, this study, like most in the literature, remains relatively limited in scope, focusing on particular constructs or subsets of policy rather than the full suite of SPM-related documents from different government bodies, inspection agencies, legislative texts and periods of government over a 30+ year period.

This analysis addresses a distinct gap by systematically examining how neoliberal rationales (e.g., transparency, accountability, competition, marketisation and improvement discourse) are articulated, repeated, normalised and adapted across all key SPM policy texts, from primary White Papers and legislation to regulatory guidance and inspection frameworks. This comprehensive, longitudinal, thematic-analytic approach has not previously been applied to this policy domain and therefore constitutes a novel contribution to the literature on neoliberalism and education policy.

Importantly, the following findings also lay the groundwork for the original findings arising from Phase 2 quantitative analysis, Phase 3 focus groups discussions and directions for alternative approaches to SPMs that follow later in this thesis.

5.1.1.1 Standards

At the centre of the neoliberal education reforms is the principle of raising educational standards. These reforms emerged in political discourse from the mid to late 1970s (Cox & Boyson, 1975, 1977**)** and the 1988 'Great' Education Reform Act (ERA) (Lawton, 1992; Wragg, 1988) and, I will argue, continue to shape the education landscape today. We can see references to 'standards' in various forms throughout the documents analysed as part of this project (Fig. 1). Specifically, it mainly focuses on actions that will 'raise' or 'improve' standards. This is evident in references to the 1988 Education Reform Act as a framework "which will raise standards, extend choice and produce a better-educated Britain" (Education Reform Bill 2nd Reading 1 December 1987, column 771). Similarly, the term is used for policies and practices aimed at addressing 'low' or 'falling' standards; "Intervening where necessary to tackle failure and low standards." (Schools Achieving Success, 2001, p.6).

'Standards' and 'raising standards' also play a prominent role in each of the UK Government Education White Papers that formed the dataset for this study, except for the 2005 white paper '14-19 Education and Skills'. I explore the possible reasons for this in the discussion section below. References to both "standards" and raising standards' are especially prevalent in the 1997 UK Government White Paper 'Excellence in Schools'. For example, "To carry out the agenda for raising standards in education we shall need a new form of Government involvement" (Excellence in Schools, 1997, p.33) and "a sign of our commitment to education and our determination to deliver our standards agenda" (Excellence in Schools, 1997, p.4). It is perhaps unsurprising that this is the case, given that this White Paper laid the foundation for the 1998 School **Standards** and Framework Act [my emphasis].

It is important at this stage to emphasise that 'standards' here does not refer to 'standardisation'. While the former, especially the effort to raise standards by the actions of market forces, is

associated with neoliberal ideas, the latter, which involves more regulation, is generally seen as a pillar of neoconservative thought (Apple, 2006; Neumann et al., 2020). However, some (Jessop, 2002; Moran, 2003; Davies, 2014) have highlighted a distinction between the rhetoric of neoliberal ideology, which advocates for reducing state influence to promote free market competition and deregulation and the actual implementation of neoliberal ideas, including the role of the neoliberal regulatory state. Here, rather than retreating, the state can be viewed as restructuring its role, intervening to create and sustain the conditions for a society based on market principles. In education, this is reflected across various areas, including:

- accountability and performance management frameworks,
- the use of standardised tests to reform education as a commodified 'product' and
- the marketisation of school choice to the move to greater centralised control away from local, decentralised regulation of education provision and intervention.

'Standard' can be seen as synonymous with 'quality'. It denotes a uniform measure against which something can be evaluated or compared, a benchmark. 'Quality' is also used in relation to the impact of policy on the 'quality' of education and the role of SPMs in driving actions to improve educational provision, as well as offering mechanisms for monitoring and assessing the quality of education delivered by schools. The interchangeable use of 'standards' and 'quality' is evident in their reference to the role of government in the following quotes.

> "**Raising the quality of education** in our schools is the most important task for this Parliament" [my emphasis] (Education Reform Bill 2nd Reading 1 December 1987, column 771).
>
> "To carry out the agenda for **raising standards in education** we shall need a new form of Government involvement" [my emphasis] (Excellence in Schools, 1997, p.33).

Interestingly, the reference to raising standards, used in the 1997 Excellence in Schools White Paper and quoted above, is closely echoed in the phrase "high and rising standards" often used by the current Secretary of State for Education, Rt Hon Bridget Phillipson MP, as well as other ministers within the current government (Association of School and College, 2025, February 24), as the cornerstone of her mission to improve education provision (GOV.UK., 2025, February 3).

'Quality' is used in the quote above in relatively vague terms. In another example, "providing a high and improving quality of education" (Self-Government for Schools, 1996) the term 'quality' is used but provides no further details about how "high and improving quality of education" might be

identified. Indeed, some of the vagueness in the use of the term 'quality' may indicate a degree of flexibility/subjectivity in its use, as well as an acceptance that there is no single, universal definition for all stakeholders, given their diverse values, priorities and perspectives on 'high quality'. By contrast, 'standards' can often be found used alongside 'quality', e.g., "the central issue of quality and standards" (Hansard Education Reform Bill - Hansard - UK Parliament - 2nd Reading 1 December 1987, column 771), which hints at something more objectively determined, indicating a nuanced difference in their meanings.

Similarly, reference to 'target-setting' is initially recognised as one of the main themes of the 1996 education white paper 'Self-Government for Schools', where the metrics introduced four years earlier in the 1992 Education Act are seen to be becoming more widely adopted, better understood and utilised by schools to set targets for self-improvement.

> "Schools will need to put forward fresh development plans with rigorous new targets for raising their standards, showing how they intend to build on their achievements to date and share their experience with other local schools". (Self-Government for Schools, 1996, p.47)

The notion of 'target-setting' remained a central idea in the white papers published during the Labour Party's period of government between 1997 and 2010. This highlights mechanisms for using agreed metrics to reliably and meaningfully measure the 'quality of schools' or 'standards of education' as part of strategies for school improvement and for evaluating the impact and effectiveness of these strategies. At that time, these targets at a school level might have been partly seen as administrative processes, with such 'targets' also forming a key part of the school's relationship with their local authority or government department. However, we can see here the development of a central logic, from the introduction of performance measures as part of the Education Reform Act (ERA) of 1988 to the 'Self-Government for Schools' white paper of 1996, through which schools would have the freedom to set performance targets for improvement using SPMs as metrics and to be held accountable through the same measures.

5.1.1.2 Autonomy and accountability

The next neoliberal principles I examine are autonomy and accountability. 'Autonomy' and 'freedom' are frequently referenced throughout all documents. Of particular interest for this study is the reference to two different groups of actors, schools and parents: a) freedom/autonomy concerning the management and governance of schools and b) freedom of choice for parents within the school admission process. This is addressed later, in section 5.1.1.4, when considering the concept of 'choice'.

Regarding the freedom of schools to determine their management arrangements and to be self-determining, two main foci can be identified. First, 'freedom' is referred to as a value, something to be protected and increased as an end in itself. For example, in the 1987 parliamentary debate about the 1988 Education Reform Act, the Education Secretary, Kenneth Baker, appealed to members of parliament "who believe in freedom and democracy" (Hansard, 2nd reading of Education Reform Bill, 1987, Column 860) to vote for the bill and the principles of "freedom, choice and improving standards" (Hansard, 2nd reading of Education Reform Bill, 1987, Column 838). Second, there is the concept, associated with both neoliberal ideology and New Public Management practices, of deregulation and the freedom of schools to manage their own affairs. Not only is this seen as providing the best basis for context-appropriate decision-making, but crucially, it can also be considered to establish a foundation for robust accountability. Until schools have the freedom to decide how they will develop, what they will prioritise and how they will spend their budget, they cannot reasonably be held accountable for those decisions. Throughout this study's documents, we can observe the principle of the inseparability of autonomy and accountability, e.g., "Empowering great teachers and leaders – autonomy and accountability will better position people to succeed and provide more effective leadership structures" (Educational Excellence Everywhere, 2016, p. 54) and "autonomy and accountability go together: greater autonomy in decisions relating to curricula, assessments and resource allocation tend to be associated with better student performance, particularly when schools operate within a culture of accountability" (Educational Excellence Everywhere, 2016, p. 105). Contemporary education policy research conceptualises 'School Autonomy With Accountability' (SAWA) as a core reform model in which schools are granted substantive decision-making authority while being held responsible for performance outcomes through formal accountability mechanisms (e.g., standardised assessments, reporting systems and improvement requirements). This body of work suggests that autonomy and accountability are not simply co-occurring features but are interdependent policy levers that shape governance practices, educational equity and innovation and that their effective coupling is pivotal for realising intended outcomes (Verger et al., 2024).

As noted earlier, we observe repeated use of terms related to freedom and autonomy used in two distinct ways. Sometimes, it refers to the general principle and value of 'freedom', as expressed by UNESCO in "the principles of freedom, justice and tolerance" (1995, p.22). At other times, it points to specific examples such as 'academic freedom', 'freedom of choice' or 'freedom to make their own decisions'. This alternating usage and repetition can be seen as rhetorical devices aimed at establishing and embedding these basic neoliberal principles as part of the natural order of what

should be expected from the education system, rather than as aspects of policies with particular goals.

In the 2016 White Paper Educational Excellence Everywhere, there is a reference to “supported autonomy,” a principle that underpins all the reforms introduced in this document. This is described as a mechanism to build capacity within the school system, providing support for school leaders to ensure they can access necessary resources and expertise from within the system itself. This marks a departure from the usual view that autonomy is, in itself, a powerful tool for driving change. Instead, we see the government taking an active role, intervening to offer support where needed and granting autonomy only when it has been earned. This ‘support’ appears in different forms, including directed funding “aligning funding, control, responsibility and accountability in one place” (Educational Excellence Everywhere, 2016, p.4) to meet system needs and new infrastructure to provide ‘scaffolding and support’ (Educational Excellence Everywhere, 2016, p.18) for schools to improve. Here, the government’s role is more controlling and interventionist than might be expected from the ‘free market’ of neoliberal orthodoxy.

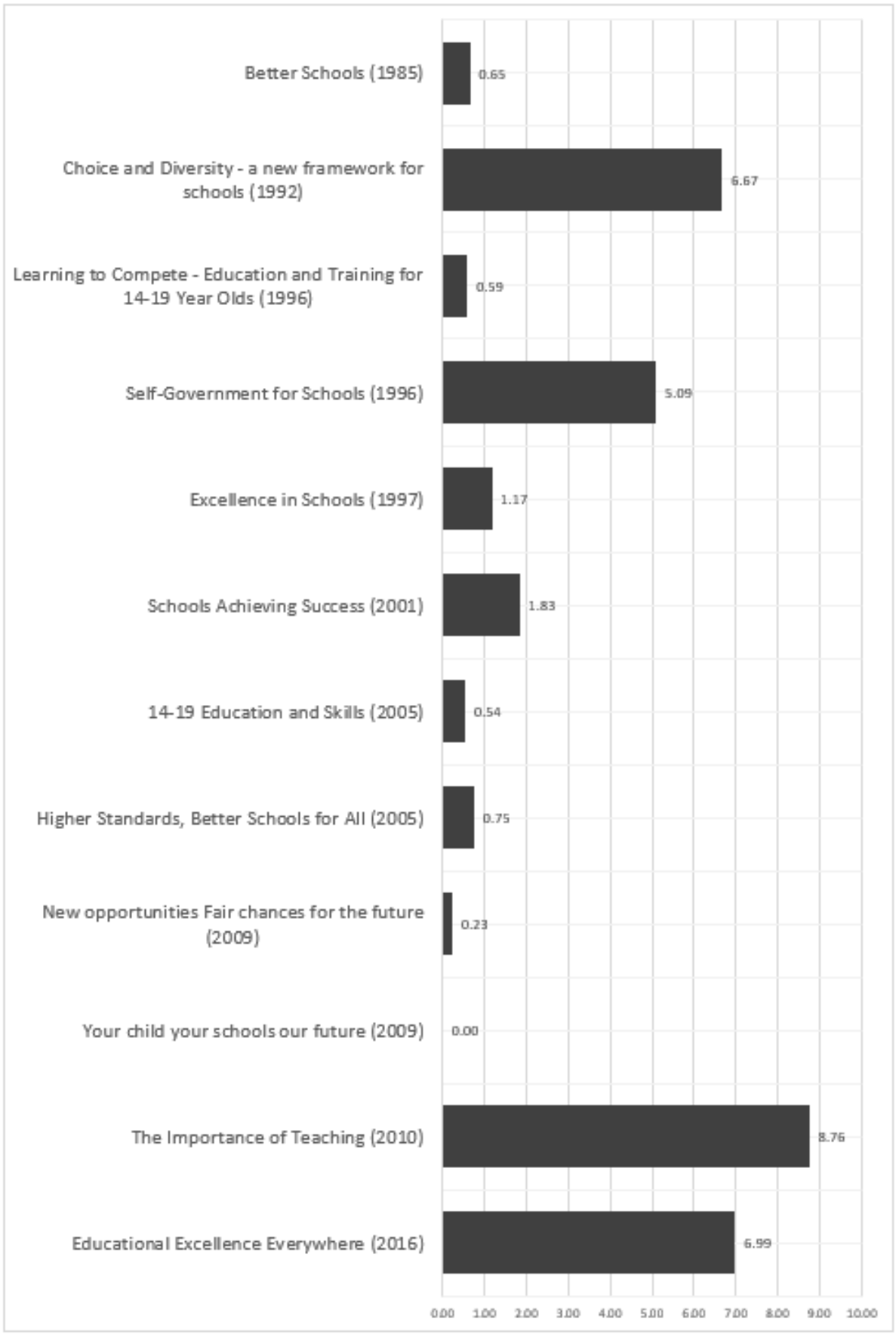


Figure 3. Frequency per 10000 words of the terms 'freedom' or 'autonomy'

5.1.1.3 Economic focus

SPMs were introduced as part of the Education Reform Act (ERA) of 1988. Alongside changes involving the creation of a national curriculum and standardised assessment to evaluate schools' educational outcomes and promote accountability, the ERA also introduced Local Management of Schools (LMS) and Grant-Maintained (GM) Schools. LMS provided much greater financial independence to individual schools, enabling them to manage their own delegated budgets, established through a system-wide regime of school-level management across the maintained sector.GM Schools legislation allowed schools to opt for GM status, meaning they would receive funding directly from central government, rather than through the local education authority. It is therefore not surprising that initiatives to introduce school performance measures and accountability systems are part of a broader programme rooted in neoliberal principles, aimed at improving educational quality and increasing financial efficiency. By making schools publicly comparable and reinforcing the market-oriented logics of choice and competition embedded within LMS, SPMs helped to legitimise and stabilise the extension of autonomy associated with GM status. In this sense, SPMs functioned less as direct regulatory instruments and more as part of the governance infrastructure through which differentiated forms of school autonomy became politically and operationally viable. For example, SPMs provide a basis for a public narrative of differential school quality. This provides a vehicle for claiming that some schools were “objectively better” than others, thereby justifying expanding autonomy to “successful” schools. Once again, we see autonomy—in this case, budgetary autonomy—working hand in hand with accountability as the mechanism chosen to bring about improvement and change.

The focus seen in the 1988 ERA on ‘economy’ was primarily about improving efficiency. From the white papers of 1996 onwards, there has been much greater emphasis on the role of education in promoting prosperity and meeting the requirements of the economy, both globally and locally (Ball, 2012). The value of education is expressed in terms of the financial benefits it can bring to individuals and their communities, as well as its contribution to making the UK economy competitive (Department for Education and Employment, 1996). This reflects neoliberal-influenced, econo-centric thinking regarding education.

Measures of student progress (value-added measures) have increasingly become a feature of SPMs since the 1990s, complementing earlier attainment-based indicators (Leckie & Goldstein, 2017a). Rather than simply measuring raw outcomes, value-added measures consider student starting points and, as such, are seen to provide a fairer basis upon which to evaluate a school’s performance. These measures and the associated methodology have evolved over time, including the introduction of Contextual Value Added (CVA) measures between 2006 and 2010. What is significant for the purpose

of this study is the use of the term 'value-added' in this context, a term originating from financial economics (English, 2005; Saunders, 1998) that refers to the amount by which the market value of a commodity or service is increased through its production. These terms were, to some extent, introduced by those advising the government on assessment and accountability. While some of the members of these panels, including the Task Group on Assessment and Testing, which reported to the Secretary of State for Education and informed the content of the ERA, were experts in education and cognitive psychology, they also included members who were economists, familiar with the use of measures to evaluate performance of products, services and markets (Department of Education and Science & Welsh Office, 1988; National Assessment and Testing: The TGAT Report, 1989). Here, we see again ideas associated with the logic of the market being applied to schools and school performance.

5.1.1.4 Choice, Competition and the Quasi-market

The neoliberal agenda fundamentally relies on a strong belief in the free market. This depends on choice and competition as its key mechanisms. Over the past three decades since the introduction of SPMs, we have observed a transformation in the frameworks within which schools operate, with the emergence, normalisation and further growth of quasi-markets for schools, aimed at enhancing both efficiency and effectiveness. Next, I will examine the related ideas of choice and competition.

'Choice', mainly referring to parental choice of school, remains a key theme throughout all documents examined in this study, for example the 1992 White Paper 'Choice and Diversity: A new framework for schools' which cited "more parental choice" (p.iii) as one of the principles underpinning reform, or the 2005 White Paper 'Higher Standards, Better Schools For All More choice for parents and pupils' which claimed that "parent choice can be a powerful driver of improved standards" (p.3). Regardless of the specific secretary of state or the political party or parties in government, choice is shown to be a central tenet of educational policy thinking. Not only is the idea of freedom of choice a vital element in the operation of the quasi-market, but it also reflects a change in the relationship between schools and parents. Choice positions parents as educational consumers, responsible for using available intelligence to choose the best school for their child and competing with other parents for what are perceived to be limited places. An example of this is in the 2005 White Paper 'Higher Standards, Better Schools For All - More choice for parents and pupils' which set out the government's plans for "greater choice so that parents can choose a school that suits their child's strengths and interests" (p.43) through a variety of measures, including the increase in places at those schools perceived to be highest performing and promoting diversity through the Specialist Schools Programme (Department for Education and Employment, 2006). The role of SPMs

here is, along with other information, to guide parents in making that choice. “Performance tables and inspections have given many parents the information that has enabled them to make objective judgements about a school’s performance and effectiveness” (White Paper ‘Higher Standards, Better Schools For All More choice for parents and pupils, 2005, p.3). Consequently, SPMs are central to the neoliberal project of placing schools within a quasi-marketplace.

Just as we saw with the closely related concept of ‘freedom’ earlier, ‘choice’ is often referred to as having value in and of itself. This is evident in the Foreword of the White Paper Choice and Diversity: A new framework for schools (1992) “Our reforms rest on common-sense principles - more parental choice; rigorous testing and external inspection of standards in schools; transfer of responsibility to individual schools and their governors; and, above all, an insistence that every pupil everywhere has the same opportunities through a good common grounding in key subjects. Few people would now argue with these principles.” (p.iii). Choice is expressed here as a principle in its own right, rather than a mechanism to achieve improvement through the operation of competition. In other documents, choice is expressed as an entitlement, describing it as the “Rhetoric of Choice” (Education and Skill Committee Report, 2005, page 31), in doing so emphasises parental empowerment. Kenneth Baker discusses parents on presenting the ERA to the House of Commons (1987, column 772) as “consumers of education” and as such have greater power to “expect and demand that all educational bodies do the best job” (House of Commons ERA 2nd reading, 1987, column 772), drawing us to a direct comparison to the way that customers might demand value for money for products or services that they have chosen to spend their money on.

However, in some documents, the role of choice is explicitly connected with placing schools in a competitive arena. Throughout all the documents related to legislation and policy (i.e., Hansard transcripts and White Papers) covering the entire 38-year period of the documents used in this study, ‘competition’ is seen to be a major subtheme related to the theme of schools acting in quasi-markets. Some of this presents itself in the relatively muted rhetoric of common phrases related to the concept of competition. For example, we see references to ‘competitive spirit’ or ‘healthy competition’ (Hansard 2nd reading ERA). These terms, which might be considered to be more acceptable, gesture toward market discipline without fully committing to market logic; their use perhaps indicates a caution in openly and fully embracing the language of the market at that stage. This perhaps is because such ideas required both political and moral legitimisation (Ball, 1993; Whitty et al., 1998). By contrast, however, by the time of the 1992 White Paper Choice and Diversity: A new framework for schools (1992) we can see a shift from tentative articulation to far greater consolidation, explicitly setting out aims to “maximise school autonomy, actual choice, diversity and competition.” (p.57). This pattern reflects a recognised phenomenon in policy reform. The

introduction of market-oriented mechanisms precedes their explicit articulation in policy discourse, with language lagging behind practice until such new governance arrangements become normalised and politically defensible (Ball, 1990; Schmidt, 2008). More recently, we can read that the further expansion of the Academies Programme 'provides real accountability, competitive pressure and choice – improving performance, enabling innovation and scaling success' (Educational Excellence Everywhere, 2016).

One aspect of competition between schools, especially with reference to SPMs, is the development of 'league tables'. It is interesting to note that in the parliamentary debates on the creation of school performance measures as part of the 1988 ERA and the 1991 Education (Information on School Examination Results) regulations, most comments express concerns around the 'risk' of introducing league tables. These specifically draw attention to the action of league tables in diverting scrutiny away from schools at the top or the bottom of the 'tables', to the risk of such tables being open to misinterpretation and to their failure to recognise "excellence and effort by schools in areas where attainment on entry to a school is low" (Baroness David, Hansard Education (Information On School Examination Results) (England) Regulations, 1991). However, between their first publication in 1992 and the Schools (Self-government) White Paper of 1996, secondary school performance tables were increasingly treated by central government and within policy discourse as having been established as a crucial element of informing parental choice when applying for secondary school places (Secretary of State for Education and Employment, 1996; UK Parliament, 1996). The tables were not simply expected to provide information about individual schools, but were explicitly designed to allow the comparison of performance between different schools:

> "it will be possible to compare the results they achieve and so measure more effectively the value added to pupils' education by individual schools" (Choice and Diversity: A new framework for schools, 1992, p. 17).

5.1.2 The influence of traditional neo-conservative thinking.

5.1.2.1 Traditional values

In addition to the subthemes linked to neoliberal economic principles, I also identify some subthemes that reflect more traditional, neo-conservative values within the selected documents. These subthemes may be somewhat limited in scope due to the focus of this work on specific references within the documentation to SPMs. Some documents, such as the White Papers, cover a broader range of topics. For example, the introduction of the national curriculum as part of the ERA might be viewed as typically conservative, in the sense of the government establishing consistency in the curriculum and addressing perceived declines in educational quality. The implementation of a

national curriculum is significant for understanding SPMs, but within these documents, it forms part of wider educational reforms that are not specifically related to SPMs.

Sections suggesting traditional conservative thinking include, for example, discussions on 'standards of behaviour' and 'discipline'. Such themes can be found in the documents introducing SPMs; for instance, Maureen Hicks, the Conservative Member of Parliament for Wolverhampton Northeast, stated at the reading of the ERA:

> "*By addressing ourselves to the task of raising standards and increasing parental choice, I hope that we can bring about a knock-on effect on the social problems so prevalent in our education system. How can we genuinely defend the present position, in which truancy is increasing, discipline is worsening and violent attacks on teachers are commonplace?*" (Hansard Education Reform Bill - Hansard - UK Parliament - 2nd Reading, 1987, column 802)

This is just one of many references that link the quality of education, measured by student attainment and achievement, to standards of student behaviour and attendance. It reflects the perceived role of the government in setting acceptable behaviour levels based on traditional morality. Here, a connection is made between the role of the market, such as 'parental choice' and the promotion of traditional values, like 'discipline'. This illustrates the tension at the core of neo-conservative, New Right thinking (Heywood, 2012), which can be seen as economically liberal but socially conservative. On one hand, there is a commitment to free-market neoliberal economics, which relies on market forces to bring about necessary improvements. On the other hand, there is a call for government intervention and state control to ensure schools are accountable to parents, children and communities for their work. 'New Right' in this context might be considered a distinct ideological current emerging from a blend of neoliberalism (economic liberalism advocating free markets, deregulation and reduced state intervention) and neoconservatism (social conservatism focusing on traditional values, law and order and moral authority).

I will return to these apparently conflicting approaches, which some have identified (Barry, 1987; Williams, 2024) as being a feature of the New Right agenda, in the discussion of these findings later.

5.1.2.2 Social mobility

Significantly, we observe an increased emphasis on 'disadvantage' and 'equality of opportunity' in relation to SPMs emerging during the years of the Blair/Brown government between 1997 and 2010 (Whitty & Anders, 2013). This focus on disadvantage and equality of opportunity also extended into the Coalition government period from May 2010 to May 2015 and beyond, where policy documents continued to address educational inequality and outcomes for disadvantaged pupils. From about 2010 onwards, however, there is a discernible shift in policy rhetoric instead toward 'social mobility'

as a central organising concept in education and wider social policy, with government strategy documents explicitly framing reforms in terms of improving social mobility for individuals regardless of their background (Cabinet Office, 2010; Department for Business, Innovation & Skills & Deputy Prime Minister's Office, 2015; Connolly, 2013). UK politicians have increasingly used this term across the political spectrum to signify a commitment to the principle of 'fair access' to opportunity, both in education and the workplace. Indeed, regarding SPMs and social mobility, interventions in schools resulting from poor 'performance' have enabled those schools to become "centres of excellence and engines of social mobility" (The Importance of Teaching, 2010, p. 51).

Within neoconservative thought, social mobility is not conceived as a redistributive or egalitarian project but as a moralised and selective process through which effort, discipline and adherence to dominant norms are rewarded. Inequality is rendered legitimate insofar as outcomes are framed as the result of fair competition within an ostensibly open system. In this framing, education functions as a key mechanism for identifying and elevating individual merit, while structural constraints are downplayed or reframed as deficits of aspiration, culture or behaviour. School performance measures are therefore well aligned with this conception of social mobility, operating as technologies that rank, differentiate and moralise educational outcomes while reinforcing the legitimacy of unequal results.

Central to the idea of schools playing a key role in promoting social mobility is the belief that high-quality education can help "children to overcome the accidents of birth and background to achieve much more than they may ever have imagined" (The Importance of Teaching, 2010, p. 6). It has been argued (Payne, 2012) that this reflects a trend in government communication from 2010 onwards, where the term 'social mobility' is used in an overly simplistic manner. This approach focuses on the individual and their need to make the most of personal 'opportunities' and 'life chances', two core neoliberal concepts, rather than emphasising societal structures that produce, sustain and exacerbate inequality (Payne, 2017), a tendency that also aligns with neoconservative traditional values of merit, responsibility and moral deservingness.

5.1.2.3 Standard

When discussing the influence of neoliberalism evident in the documents we previously referred to as 'standards', the word—often used in the plural—means 'quality'. However, there are also references to 'standard' in a subtly different sense, referring to a specific level of quality. This is particularly clear in references to "floor standard" (Importance of Teaching, 2010, p.13) or "minimum standard", (Ed Skills Comm Report School Accountability, 2010), p.24) both of which imply fixed levels of acceptable performance along with the associated processes for their measurement. This is

significant for understanding the role of SPMs in both defining and shaping expectations within schools in England, as well as in formulating what it means for the quality of education to be unacceptable. It also suggests a belief in a relatively fixed, externally assigned, acceptable standard or value—proactively decided by government—a perspective more closely aligned with neo-conservative rather than neoliberal thinking.

This analysis reveals that UK government education policy documents since 1985 have incorporated a combination of both neoliberal and neo-conservative ideas. While neoliberalism is evident in market-based reforms, such as school choice, autonomy and accountability, neo-conservative thinking is also apparent, particularly in the emphasis on discipline, standards and the promotion of social mobility. These themes emerged during the Conservative government era, spanning from 1985 to 1997, as representations of the New Right ideology of the Thatcher government. SPMs and the establishment of a wider school accountability system are outcomes of the ERA of 1988 and can be considered to be a representation of this thinking. This blend of both neoliberal and neoconservative ideas is also evident in the documents of the New Labour government, which was in power between 1997 and 2010. This, perhaps, reflects a politically pragmatic approach adopted by the New Labour government (Hay, 1999; Prideaux, 2005), as it aimed to move towards the centre ground to appeal to a broader electorate. Next, I will discuss the evidence for a turn towards a greater focus on social democratic principles during the New Labour government.

5.1.3 Increased focus on Social Democracy

In the documents produced by the Labour government between 1997 and 2010, we can see an increased focus on ideas associated with social democratic principles, particularly an emphasis on 'disadvantage', 'collaboration', 'social justice' and 'support'.

5.1.3.1 Disadvantage

While some fleeting references were made to disadvantage in the documents prior to 1997, these were typically within the wider context of parliamentary debate and were raised by members of parliament from both opposition parties and the governing Conservative Party. It is only from 1997 onwards that significantly greater focus was placed on addressing disadvantage and promoting the associated ideals of equality of opportunity and equity. We can see repeated references to the benefits of quality education in addressing the needs of economically and socially disadvantaged children. We see this focus emerge in the 1997 and 2001 White Papers (Excellence in Schools, 1997; Schools Achieving Success, 2001). Indeed, the first line of Excellence in Schools (1997) makes this clear:

> "This, the first White Paper of the new Government, is as much about equipping the people of this country for the challenge of the future as it is about the Government's core commitment to equality of opportunity and high standards for all". (Excellence in Schools, 1997, page 3)

However, it is not until the 2005 White Paper (Higher Standards, Better Schools for All, 2005) that this subtheme becomes the key factor in the proposed changes. In her introduction to this White Paper, the then Secretary of State for Education, Ruth Kelly, stated "we must deliver for all children, but particularly for those whose family background is most challenging" (Higher Standards, Better Schools for All, 2005, p. 5). It is also worth noting that this emphasis has, to a large degree, been sustained and can be seen, for example, in multiple references to 'disadvantaged pupils' peppered throughout the current Ofsted Framework documents.

5.1.3.2 Collaboration

We have previously discussed the ongoing influence of neoliberal ideology, particularly in the continued emphasis on competition between schools as a driver of improvements and standards. However, in the Secondary Education (2005) and School Accountability (2010) White Papers, we see a much-increased focus on collaboration between schools. This collaborative turn is also reflected in later documents, including discussions about the need for 'competition and collaboration' to "deliver the best possible outcomes" (Educational Excellence Everywhere, 2016, page 9). Throughout various periods of educational reform, numerous structural changes have been proposed to encourage collaboration, including informal partnerships, Teaching School Alliances, specialist school networks and partnerships in the delivery of qualifications (notably the 14-19 Diploma), as well as more recent developments such as Multi-Academy Trusts. However, at the same time, we observe no significant change in SPMs to reflect this shift in emphasis. I will revisit this point later, in the context of proposed alternative approaches to SPMs in section 8.5.4. It could be argued that calls for competition and collaboration between schools are inherently contradictory. Given that this is happening alongside an established accountability system centred on SPMs—which are designed to promote competition—such calls for collaboration are likely to have limited impact, at best. Therefore, calls for collaboration appear to lack substance and are, at best, well-meaning but rhetorical and at worst, deliberately misleading.

5.1.3.3 Social Justice

The concept of social justice is seen as a prominent theme in two education White papers produced by the Conservative-Liberal coalition government, The Importance of Teaching (2010) and

Educational Excellence Everywhere (2016). Interestingly, in the following quote, we see this placed alongside those same neo-conservative (e.g. 'civilised society') and neoliberal (e.g. 'economic growth') principles we have discussed previously:

> *"Education is the hallmark of a civilised society, the engine of social justice and economic growth, the foundation of our culture and the best investment we can make in the future of our country" (Educational Excellence Everywhere, 2016, page 5).*

The meaning of 'social justice' being used here is clear from another quote from the same White Paper written by the then Secretary of State for Education, Rt Hon Nicky Morgan MP, in the foreword:

> *"Education has the power to transform lives and, for me, is a matter of social justice – extending opportunity to every child, wherever they live and whatever their background" (Educational Excellence Everywhere, 2016, page 3).*

'Social justice', in this context, refers to extending opportunities to those previously perceived as being excluded and to addressing the inequalities associated with geography or 'background'. The use of 'background' here is used as a comprehensive term to include any social, economic, cultural or personal circumstances that might create disadvantage or limit a child's opportunities in education.

This analysis has identified three main socio-political themes related to SPMs. They demonstrate the influence of neoliberal and neoconservative ideas in shaping the school accountability system, including SPMs. This process began with the policies of the Thatcher government in the early 1990s and has continued through successive governments of different political persuasions—Conservative, Labour and the Conservative/Liberal Democrat Coalition—to the present day. During the 13 years of the New Labour government, from 1997 to 2010, there was a greater emphasis on social democratic principles, promoting a focus on how disadvantage affects educational outcomes and encouraging collaboration between schools, as well as principles of social justice. However, this social democratic shift occurred within the context of a commitment to the same neoliberal principles of choice, competition, autonomy and accountability. For this study, this is significant. It not only provides a historical and socio-political background for the introduction, development and ongoing use of SPMs but also offers valuable insight into their intended purpose. Next, I will examine the aims and purposes of SPMs in more detail, as revealed by the selected documents. This will, in turn, lay the groundwork for assessing the effectiveness of SPMs in relation to these stated aims.

5.2 Aims and purposes of SPMs (Research Question 2)

In this section of the study, themes regarding the aims and purposes of SPMs are developed through an analysis of the selected government documents. This analysis provided a foundation upon which to address Research Question 2 by questioning how effective the production, publication and operationalisation of secondary school performance measures in England are in meeting their stated aims.

The main themes identified in the documents are: a) Purpose of SPMs, b) Methodology of SPMs, c) Limitations of SPMs and d) Interpretation of SPMs.

5.2.1 Purpose of SPMs

In examining the purpose of SPMs within the selected documents, several subthemes were identified.

5.2.1.1 Transparency through publication

The first of these is the concept of 'publication'. While it could be argued that publishing is central to what we understand by SPMs, there is a clear notion that schools, as publicly funded organisations, should be open and transparent about their work. In this context, sharing school performance information with stakeholders through the publication of SPMs can be seen as a purpose in itself.

> "greater accountability and transparency through the publication of secondary school performance tables and school inspection reports, information for parent"'
> (Hansard Schools (Self-government) (25 June 1996), c 153)

However, the concept of transparency relies on more than just the publication of information about schools. Other aspects of the information are also critical, such as accessibility, understanding, utility, consistency and reliability. I will revisit these ideas when considering various alternative approaches to SPMs in section 6.1.4.2.

The selected documents reference SPMs being used by various stakeholders: Local Authorities, School Improvement Partners, governing bodies, trustees, school leaders (during self-evaluation and strategic improvement planning), government statisticians, Ofsted (which will be discussed later) and parents. Among these groups, publishing SPMs to inform parents stands out as the most prominent idea emerging from the documents. This is closely linked to the concept of choice and the functioning of schools in quasi-markets, as discussed earlier in section 4.1.1.4.

5.2.1.2 Accountability

As may be expected, there are repeated references to 'accountability' and schools being 'held to account' through the use of SPMs.

> "every school and college is held to account for the quality of what it delivers" (2005 White Paper 14-19 Education and Skills, p 26)

SPMs are considered a vital element of a broader school accountability system, which, although it has been reviewed and revised periodically since its inception in the 1988 ERA, has remained essentially unchanged. This accountability system comprises three main components. These are:

Identifying Underperformance – Triggering Intervention

SPMs are described as playing a vital role in identifying school 'underperformance' and which schools might be classified as 'failing'.

> "School level figures are used to identify schools where intervention may be needed to bring standards up to an acceptable level" (DfE Key stage 4 performance, Methodology – Explore education statistics – GOV.UK, p.2)

This, in turn, can trigger various forms of intervention, such as:

- Requiring schools to collaborate with higher-performing schools.
- Allocating resources and expertise to assist in 'turning around' such schools.
- Closing or replacing the school, or converting or brokering it to form a new academy or to become part of an established and successful Multi-Academy Trust (MAT)

Informing School Improvement

A subtheme which recurs through many of the documents is that schools are expected to use SPMs to inform strategic school improvement planning. There is little or no reference to this specific purpose for schools before 1997, the year that New Labour became the governing party. Equally, there is no reference to its use in school improvement in the documents after 2010, when they left government. That is not to say that this was not, a perhaps implied, aim of SPMs outside of this period. The concept of SPMs influencing school behaviours is evident throughout many of the documents analysed, as discussed below in section 4.2.1.3. However, it is explicitly set out in many of the New Labour policy documents.

> "Our vision is of a supporting infrastructure which includes an accountability framework and school improvement strategies underpinned by the new School Report Card" (21st Century Schools, 2008, p.10)

These set out a core purpose of SPMs to inform school improvement, through the use of performance targets, self-evaluation and work with school improvement partners.

Informing Ofsted inspection

SPMs are referred to as having an important part in the school inspection system, which has been carried out by the schools inspectorate and regulator, Ofsted, since 1992.

> "Used as part of the background information to inform school inspections. For example, it enables inspectors to identify areas of the curriculum or groups of pupils where a school appears to be doing less well". (DfE Key stage 4 performance, Methodology – Explore education statistics – GOV.UK, p.2)

This includes:

- The use of SPMs to monitor school performance. This can trigger an inspection if performance is indicated to have deteriorated significantly or dropped below an acceptable standard.
- SPMs being used prior to and during inspections to provide information and evidence to be used in forming judgements.

Informing parental choice

An important principle relating to SPMs is parental choice and the associated pressure on schools to be considered by parents as performing well.

> "Used to inform choice of a secondary school for their child". (DfE Key stage 4 performance, Methodology – Explore education statistics – GOV.UK, p.2)

This is pivotal to school accountability. Damage to a school's reputation due to perceived 'underperformance' or 'failure' in comparison to other neighbouring, competing schools can reduce admission numbers, leading to a decline in student-led funding**.**

5.2.1.3 Influencing school behaviours

While the core purposes of SPMs are to inform school choice, identify schools in need of intervention and guide inspections—all within an environment that encourages schools to be competitive—there are suggestions within the selected documents that a broader role for SPMs is to influence school behaviours. These behaviours might involve, for instance, promoting desirable actions or discouraging undesirable ones. For example, SPMs are described as being "designed to encourage secondary schools to teach a broad and balanced academic curriculum, with a strong focus on English and maths" (How to understand school performance – everything you need to know, 2022). They are also intended to address the risk of schools 'gaming' the system, which can include behaviours such as 'off-rolling' pupils or focusing narrowly on test preparation. Research suggests that publication of performance data creates strong incentives that can shape school behaviour and categorisation of pupils (Allen & Burgess, 2011; Goldstein & Leckie, 2011). Policymakers have responded by reviewing which qualifications count in performance tables and by investigating removal from rolls when pupils are likely to underperform, as undesirable practices linked to

accountability pressures (UK Government, 2018; Ofsted, 2019). Ultimately, performance measures exist to exert pressure on schools to improve and maintain acceptable standards, but this pressure can have unintended consequences on school decision-making (Goldstein & Leckie, 2011; McShane, 2020).

5.2.2 Methodology of SPMs

A considerable amount of detail related to the specific methodology for calculating school performance measures (SPMs) was observed, particularly in documents of a more technical nature. These documents provide us with details of the design of SPMs, for example, how student progress is measured between key stages (Department for Education, 2024). It also offers some context by explicitly stating what aspect of a school each measure aims to capture. In this way, we see the practical implementation of policy. The design and methodology involved in constructing SPMs can be observed to reflect the priorities of specific ministers. For example, Progress 8 aims to measure the progress a student makes during their five years in secondary education by comparing attainment at the end of key stage 2 (aged 10 or 11) with that at the end of key stage 4 (aged 15 or 16) (Department for Education, 2024). However, *"attainment"* at key stage 4, in the form of the SPM Attainment 8, is measured in a very specific way. This includes:

- a maximum of eight qualifications, all of which must be on the government's approved list of qualifications (Department for Education, 2024);
- qualifications in specified subject areas, including English and mathematics, as well as an additional three (or more) subjects from the English Baccalaureate (EBacc) subjects (Sciences, Humanities and Languages), demonstrating a government commitment to a strong academic core within the key stage 4 curriculum (Department for Education, 2024);
- English and mathematics are given double weighting, emphasising the government's priorities in these subjects (Department for Education, 2024).

Furthermore, Progress 8 deliberately remains non-contextual. This reflects a conscious policy shift away from more contextual SPMs, with the abolition of contextual value-added (CVA) measures around 2011 and replacement by simpler progress metrics, because ministers considered contextual adjustments as misaligned with expectations for all pupils (Leckie & Goldstein, 2017a; Department for Education, 2024). The CVA measure, which previously adjusted for pupil socio-economic and demographic characteristics, was discontinued and superseded by measures such as expected progress and Progress 8, which take pupils' prior attainment into account but not wider context (Leckie & Goldstein, 2017a). The now well-known phrase from then prime minister David Cameron

and deputy prime minister Nick Clegg that contextualised adjustments represented *"the soft bigotry of low expectations"* has been widely cited in policy commentary and media reporting of the White Paper *The Importance of Teaching* (Department for Education, 2010; Leckie & Goldstein, 2017a).

It is also worth noting, in relation to the communication of methodology in government SPM documentation, that there are national and international data ethics standards. These include the UK Government's *Data Ethics Framework* (UK Government, 2025), the *Algorithmic Transparency Recording Standard* (Government Digital Service, 2023), the OECD's *Principles on AI* (2019) and the EU's *Ethics Guidelines for Trustworthy AI* (European Commission, 2019). These standards stress principles such as proportionality, accountability, traceability, bias mitigation and stakeholder engagement. While these standards are rarely applied systematically to school accountability systems, they provide valuable and important established tools for evaluating the social, ethical and democratic implications of SPM design and implementation.

5.2.3 Limitations of SPMs

In several of the selected documents, we find references to the limitations of SPMs. For example, the reading of the Great Education Reform Bill, later known as the ERA, in 1988, raised a key issue: the risk associated with publishing SPMs, specifically the potential development of league tables. For instance, Baroness Blackstone sets out her fears about the introduction of SPMs based on national tests as part of the House of Lords debate of the Education Reform Bill:

> "They have very many dangers … they can be used as league tables on individual schools or on individual teachers. As we all know there is a terrible risk that once one sets out league tables of that kind they will be misinterpreted. Some schools which supposedly have very poor results may actually be doing extremely well if one takes into account the nature of their pupil intake. Other schools which may look as if they are doing rather well may be doing very badly because their pupil intake is such that one would have expected them to do very much better". (Hansard Education Reform Bill 1988-05-03, c 499)

The concern about schools' performance being compared through the use of league tables was also a primary concern for the Task Group on Assessment and Testing (TGAT), established to advise on the reforms comprising the ERA. In their report on the National Curriculum (Task Group on Assessment and Testing, 1987), they clearly express concerns about the publication of SPMs.

> "…there is a fear that results will be published in league tables of scores, leading to ill-informed and unfair comparisons between schools". (Task Group on Assessment and Testing, 1987)

These quotes both articulate concerns about the potential misinterpretation of SPMs if they are used to produce simple rank orders of schools, i.e. to form 'league tables'. The official 'school performance tables' were first published by the Department for Education in 1992. While the government referred to them as 'performance tables,' the media and the public could immediately take the raw data and organise it into ranked lists, which were dubbed 'league tables'. The term 'school league table' was quickly adopted by the media to describe these rankings (Leckie & Goldstein, 2017a) and it has stuck ever since (Butcher & Corfe, 2025, September 1; Dukes, 2025, September 21). The media's role has been to popularise and simplify the government's data in this 'league table' format. More recently, the online version of the government's 'Compare the performance of schools and colleges in England' website allows users to create lists of schools and then rank them based on various metrics covering performance, absence and pupil population. This, therefore, is designed to allow users to generate their own 'league tables' to compare schools.

Some documents offer more general insights into the limitations of SPMs. In addition to the clear encouragement to use SPMs to understand a school's performance and compare it with others, there are numerous references—both in the government's public-facing documents (e.g., Understanding school and college performance measures, 2023) and in more technical reports (e.g., Secondary Accountability Measures, 2023)—emphasising the importance of recognising the limitations of SPMs. Care has been taken to explicitly advise users to exercise caution when comparing schools or drawing conclusions about school performance based solely on SPMs. This is exemplified in a quote from Secondary Accountability Measures (2023, p.8):

> "We have always been clear that all users of educational school performance data need to consider this data alongside a range of other information about the school and its individual circumstances, for example by speaking to the school directly and that conclusions should not be drawn based on a single piece of data alone".

This is made especially clear in the advice provided alongside school data from 2021/2022, the year affected by the COVID-19 pandemic. A series of caveats was published alongside the data (Secondary Accountability Measures, 2023), warning of the "*uneven impact of the pandemic on pupils and schools" (p.8)* and emphasising the need to be *"sensitive in [the] use of it"* (p.9). Additionally, the name of the 'Find and compare school and college performance' website was changed in 2022 to 'Find School and College Performance Data' to *"reduce the emphasis on comparison between institutions,"* and the format in which the data was presented was altered, with coloured bandings removed to "discourage simplistic conclusions being drawn about a school or college's overall performance" (Secondary Accountability Measures, 2023, p.10). In 2024, the name was changed again, reinstating the term 'compare' as the DfE moved back toward a more standardised

accountability framework, with the website called 'Compare the performance of schools and colleges in England'.

While there are references to the limitations of SPMs found through this analysis, these were relatively few (n=70) and located in a small number (n=8) of documents. This is perhaps a reflection of the nature of the sample of documents used here. The references to the limitations of SPMs feature in a) parliamentary debates around the introduction of SPMs between 1988 and 1992, b) guidance documents, published more recently with the aim of supporting appropriate interpretation of the published SPMs, c) the Assessment Report published by the Office for Statistical Regulation in 2022 which reported on an evaluation of the key stage 4 performance statistics and d) 2 Reports to the Education and Skills select committee, which had taken expert advice from stakeholders and researchers on various aspects of SPMs to inform the report. The other documents (white papers, green papers, policy papers and acts/bills) all contribute to legislative processes and the development and implementation of government policy. As such, they are not intended to be in any way critical of SPMs and do not include the potential of alternative viewpoints or voices, which the other documents have.

When using the purposes of SPMs as a basis for evaluating their effectiveness, informed by discussions with various stakeholders in focus groups (section 6.1.2), I will revisit these stated limitations as an important consideration. Their inclusion in government documents regarding SPM methodology raises significant questions. What measures are evident to address the potentially harmful effects of these limitations? Is the interpretation and use of SPMs consistent with the cautions linked to their limitations? I will turn to the interpretation of SPMs next.

5.2.4 Interpretation of SPMs

Finally, we observe both advice, as outlined in government methodology documents and anxiety expressed, particularly in parliamentary debates, regarding the interpretation of SPMs. Concern about the risks of misinterpretation appears in many statements made during the ERA 2nd reading in 1987. Worries that the publication of SPMs might "invite unfair comparisons and give a partial picture of a school's work," which are "potentially misleading, particularly where significant decisions are based upon that information" (Education (Information On School Examination Results) (England) Regulations, 1991), are common in numerous government documents. In addition to calls for caution, discussed in the limitations section above, guidance is also published in government guides for stakeholders, such as 'Understanding school and college performance measures' (2023). This appears to aim at supporting the appropriate interpretation of SPMs, as published, by encouraging users of SPMs to consider factors that might impact school performance data. For example, alongside

information about the content of the Compare School and College website and the design and meaning of the different measures published there, advice is provided, for example, which encourages users to consider:

- looking at *"more than one performance measure when comparing schools and colleges to get a full picture of how it's performing"*
- *"how well each of a school's prior attainment groups are progressing as an indication of whether a school is equally effective at helping all of its pupils meet their potential"*
- *"the context of the school or college as there may be factors that can help explain its performance"*
- that it "*is not always possible to compare year-on-year results over time"*

However, what remains unclear from this analysis is how much of this advice is actually followed and how effective the guidance is in practice in reducing the misinterpretation of SPMs. Investigating this further is an important next step when assessing the effectiveness and broader impacts of SPMs through the focus groups with stakeholders (sections 6.1.2 and 6.1.3).

5.3 Conclusion to Thematic Analysis of Documents Related to Secondary School Performance Measures in England

This analysis demonstrates that the development and implementation of secondary school performance measures (SPMs) in England have been consistently shaped by political, social and economic factors, with neoliberal principles exerting a defining influence. Across nearly four decades of policy documents, themes of standards, autonomy, accountability, competition, quasi-markets and economic rationality recur as central logics, embedding market-oriented assumptions into the design and justification of SPMs.

The study also clarifies how the purposes and design features of SPMs are articulated in official documents. SPMs are intended to promote transparency, accountability, parental choice, school improvement and targeted intervention, while deliberately influencing school behaviour through embedded incentives. Methodological decisions—such as the structure of Attainment 8 and the non-contextual approach of Progress 8—reflect political priorities, revealing tensions between intended aims and practical limitations. The documents further highlight concerns about misinterpretation, the risks of league tables and the inherent constraints of SPMs, establishing what the measures are designed to achieve and what they cannot reliably capture.

In relation to the sub-questions for this research phase:

- ***What political, social and economic factors have influenced the development and implementation of secondary SPMs in England?*** The findings show a clear and sustained impact of neoliberalism and related ideologies in shaping SPM development, implementation and justification.
- ***How are the purposes and design features of secondary SPMs in England articulated and explained?*** Policy documents explicitly link SPMs to transparency, accountability and improvement while embedding incentive structures and acknowledging limitations, providing insight into the intended functions, priorities and risks of SPMs.

Taken together, these findings make an original contribution by moving beyond high-level (e.g. West, Mattei & Roberts, 2011) or episodic (e.g. Prior et al., 2021) accounts of SPM policy development. Rather than focusing on individual reforms or specific moments of change, this analysis systematically traces how these political and ideological influences are articulated, re-articulated and sustained across more than three decades of policy discourse. In doing so, it provides a more granular and longitudinal account of how secondary SPMs have been consistently framed, justified and defended within English education policy.

The thematic analysis of the 39 government documents related to SPMs reveals some important insights. This detailed analysis, which provides a basis for understanding the impact of socio-political ideologies, particularly the persistent influence of neoliberal thinking, on SPMs, not only provides historical and political contexts for their introduction and development. It also provides insight into the policy logic and intended effects of SPMs in driving higher educational standards through the operation of market forces, including school choice, which triggers interventions to address 'underperformance' and informs schools' strategic improvement planning. In addition, this analysis has allowed the identification of the stated aims of SPMs, as well as some of their specified limitations. This will be returned to in the discussion of the effectiveness and effects of SPMs later (sections 7.3 and 7.4).

The originality of this thematic analysis lies not in identifying these logics per se, which have been noted in previous literature (e.g. Ambrosio, 2013; Ordofa & Asgedom, 2022; Wragg, 1988), but in demonstrating how they are repeatedly mobilised and stabilised over time through policy discourse. By examining policy documents across a 30+ year period, this research shows how appeals to transparency, accountability and improvement are consistently used to legitimise the continued use and expansion of SPMs, even as their limitations are explicitly acknowledged within the same documents. This longitudinal perspective substantiates and deepens existing claims by revealing how these justificatory logics become normalised and taken-for-granted, contributing to the embedding

of SPMs as core policy instruments within the English school system. As SPMs become increasingly entrenched, they do not merely measure school performance but actively shape school behaviour, system structure and policy possibilities, thereby making substantive change to the accountability regime progressively more difficult.

Overall, these findings provide a strong foundation for understanding the ideological and policy context of SPMs, supporting later analyses of their effectiveness, wider effects and alignment with stated objectives. It also informs a better understanding of the conceptual and ethical foundations, aims, rationales, values, assumptions, conditions and processes of the SPM Policy Logic (section 3.3.3), which underpins a thorough evaluation of SPMs.

In the next chapter, I employ a variety of quantitative approaches to investigate the concept of 'performance' as defined within the headline SPMs, raising critical questions about the construct validity of SPMs and the associated implications for the ethics surrounding their use.

6. Findings from Phase 2: Analysis of the Structure of SPMs using Quantitative Analysis

Having established how secondary SPMs are constructed and justified within policy discourse, this section examines the underlying features of schools as represented in SPMs to investigate the breadth and meaning of the concept of 'performance' they operationalise.

This analysis seeks to address the following subsidiary research questions:

- What underlying latent constructs can I empirically identify and conceptualise as shaping 'performance' within SPMs?
- With reference to the current headline performance measures, what is the extent and breadth of the concept of 'performance'?
- To what extent do SPMs encapsulate elements beyond the influence of schools and can digital tools aid in assessing their influence on the stated policy goals of SPMs?

Together these contribute to analysis which focuses Research Question 2 (*to what extent is the production, publication and operationalisation of secondary school performance measures in England effective in meeting their stated aims, in this context?*) and Research Question 3 (*How can the findings of research and advances in technology be utilised to improve the communication and use of school data by stakeholders for school performance measurement purposes?*).

This research comprises a series of analyses to iteratively understand the construction and implications of SPM implementation, beginning with Factor Analysis. This provides a robust set of tools to unpack the headline SPMs and extract meaningful insights from the school performance measures dataset. These techniques are used to identify and better understand the latent factors which shape SPMs. These are the underlying, unobserved constructs for SPMs. Examining these allows me to explore the meaning of performance as represented by SPMs.

I first use Exploratory Factor Analysis (EFA) to empirically determine the number of potential factors underlying the 6 headline SPMs (section 5.1). After establishing the number of factors, I employ Confirmatory Factor Analysis (CFA) to interpret what these factors might represent (section 5.2). I then explore the relationships between the headline SPMs using correlation analysis (section 5.3). Next, I apply Structural Equation Modelling (SEM) to develop progressively more sophisticated and complex models of the latent factors, aiming to detect hidden patterns and gain a deeper understanding of their connections to SPMs, other school information variables and to each other (section 5.4). Finally, I conduct cluster analysis to identify groupings of schools based on variables

related to school context and character and to examine patterns associated with SPMs within these clusters (section 5.5). This combination of methods allows the SPMs to be closely examined, giving insights into the concept of performance as represented by School Performance Measures (SPMs). The findings of this sequence of analyses offer a robust empirical account of the breadth of SPMs, their validity and fairness as measures of school effectiveness and provide a basis for reconfiguring SPMs to make them more meaningful and fair.

6.1 Number of factors underlying headline SPMs

In this initial analysis, I employed Exploratory Factor Analysis on the SPMs dataset (detailed in section 3.6.2) to determine the number of latent variables present. This is used later to examine the nature of these factors in more detail. The eigenvalue analysis identified the number of latent factors describing the 6 headline SPMs by comparing two distinct datasets. As can be seen in the parallel analysis scree plot (Figure 4), these are:

- The **actual dataset** being analysed.
- A **resampled dataset**, created using permutation samples from the real data.

The eigenvalues from our actual data consistently surpassed those from the resampled data. In fact, they exceeded the 95th percentile of the eigenvalue distribution from this comparison set. This strongly suggests that the factors I extracted from the actual data are meaningful and not just random occurrences.

To determine the number of factors to retain, I identified the point at which the eigenvalues from the actual data dropped below the eigenvalue distribution from the resampled data. This analysis indicated that I should retain either three or four factors.

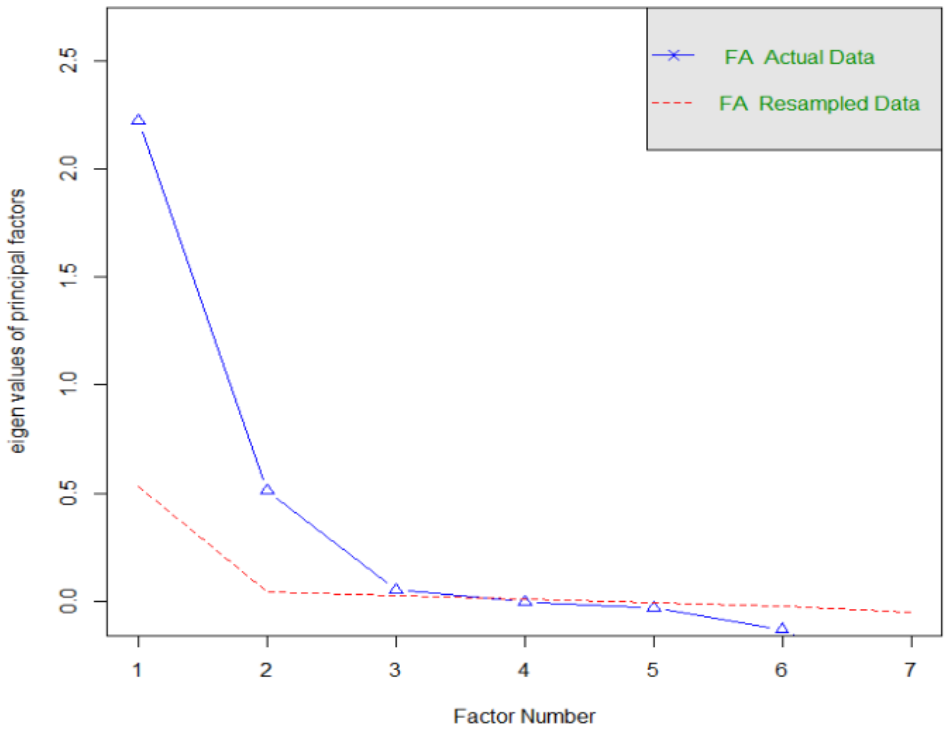


Figure 4. Parallel Analysis Scree Plot for dataset of 6 headline SPMs for 3125 mainstream state secondary schools in England, academic year 2021-2022.

A comparison of the fit measures from the EFA using three and four factors suggests that a three-factor model is preferable.

The factor loadings for the 4th factor of a 4-factor CFA are observed to be very low, confirming that 3 factors are most suitable for this dataset.

Table 16. Eigenvalues of the factor arising from Exploratory Factor Analysis of the study dataset, indicating 3 factors with eigenvalues greater than 1 (in green)

Factor	Eigenvalue
1	2.633
2	1.352
3	0.949
4	0.834
5	0.556
6	0.461

Summary descriptions of the underlying, latent variables for SPMs are developed using the 3-factor EFA factor loadings (Table 16 and Figure 5).

These latent variables are conceptualised as:

- Factor 1 = '**Attainment (+ Destinations)**' - Attainment 8, Basics, EBacc APS and Destinations
- Factor 2 = '**Progress**' - Progress 8
- Factor 3 = '**Curriculum**' - % EBacc entry

Table 17. Factor loadings for the 3-factor EFA using the 6 headline SPM variables.

	Attainment (+ Destinations)	**Progress**	**Curriculum**
Attain_8	**0.857**	0.413	0.299
Progress_8	0.456	**0.854**	0.239
Percent_EBacc_Entry	0.291	0.183	**0.831**
EBacc_APS	**0.783**	0.37	0.491
Basics	**0.819**	0.389	0.307
Destinations	**0.548**	0.192	0.182
	MR1	MR3	MR2
SS loadings	2.612	1.259	1.205
Proportion Var	0.435	0.21	0.201
Cumulative Var	0.435	0.645	0.846

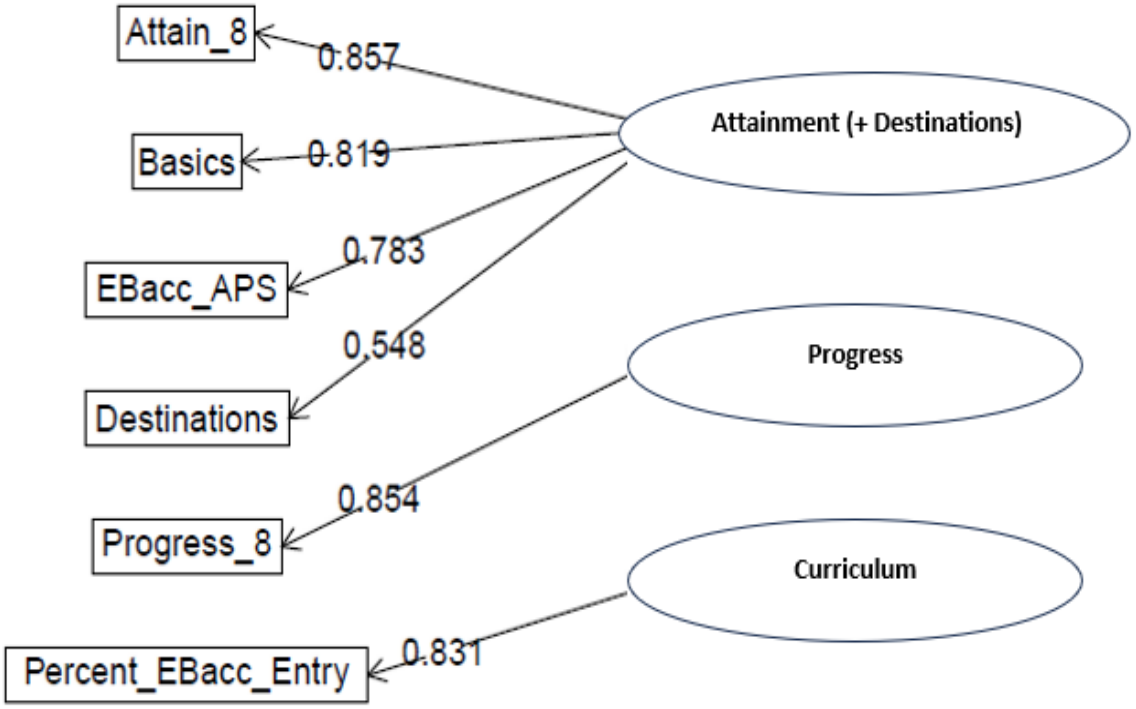


Figure 5. Factor Analysis Diagram showing factor loadings

Most notably, 3 out of the 6 headline SPMs (Attainment 8, Basics and EBacc APS) load heavily onto a single factor, highlighting redundancy among these measures and revealing limitations in capturing the full scope of 'performance' in the published metrics. This suggests that the concept of school

‘performance’ is predominantly driven by one core factor, namely Attainment/Destinations. I will later, in section 5.4.2, explore in greater detail the nature of this factor and the variables influencing it, which are beyond schools' control, thereby questioning its validity as a measure of school effectiveness.

Progress 8 and % EBacc Entry load heavily on their own factors (‘Progress’ and ‘Curriculum’ respectively), indicating that these indeed capture different aspects of school performance. However, their loading is not complete, suggesting that other factors may also be affecting these measures. In practical terms, this means that Progress 8 and EBacc entry each largely measure what they are intended to measure—student progress and curriculum choices—but they do not measure only those things. A school’s Progress 8 score, for example, reflects not just how much progress pupils make, but is also influenced by other factors. Similarly, EBacc entry rates reflect curriculum decisions, but are also shaped by other as yet unspecified factors. As a result, these measures may provide useful but incomplete pictures of school performance and should not be interpreted as standalone indicators of school effectiveness.

‘Destinations’ has a considerably lower factor loading than the other performance measures, indicating limited correlation with any of the underlying factors. However, I do not observe Destinations to load more strongly on another, fourth factor that might reflect additional breadth in performance. I propose two possible explanations of this:

First, the Destinations measure is based on the previous academic year’s cohort of students. The other headline SPMs are based on the current year. Consequently, these measures apply to different cohorts of students. As a result, statistical fluctuations can be expected based on these different sets of students’ individual interests, aspirations, influences and opportunities.

Second, the mean student cohort size for the analysed dataset was 177, meaning that the annual sample size is likely too small to smooth such variation statistically. In addition, the low loading may suggest the presence of other variables (e.g., availability of employment, education and training in the school’s locality) which are not included in the SPM-only model.

Overall, the EFA indicates that headline SPMs reflect three latent dimensions of performance, with attainment-related measures accounting for the majority of explained variance. Progress and curriculum measures capture distinct but incomplete aspects of performance, while the Destinations measure appears weakly aligned with the existing factor structure, suggesting limited additional breadth within the current SPM framework.

6.2 Factors underlying SPMs

Building on the insights gained from the Exploratory Factor Analysis (EFA), I construct models for expected factors related to school performance data and broader school information. These models are built around the descriptions of the headline SPMs gained from the analysis of SPM government documents conducted in Phase 1, in particular those relating to methodology as outlined in section 4.2.2. I then test and refine these models using Confirmatory Factor Analysis (CFA) before combining them using Structural Equation Modelling (SEM) (section 5.4.2). For school performance, I compare a shared 'Performance' construct with the 'Attainment,' 'Progress,' 'Curriculum,' and 'Destinations' constructs identified in EFA, separating out 'Destinations' due to its distinctive characteristics compared to the other SPMs. For wider school information, I analyse two theory-driven constructs: 'School Context' and 'School Character'. Taken together, these factors provide a valuable basis for evaluating policy around school accountability, the interpretation of SPMs and the complex meanings and values they represent.

6.2.1 SPM factors ('Perform', 'Attain', 'Progress', 'Curriculum' and 'Destinations')

To compare with the data-driven latent variable model constructed through EFA, a 'performance' model is created (Figure 6), which includes all 6 headline secondary SPMs to examine how well these six measures represent a shared 'performance' construct.

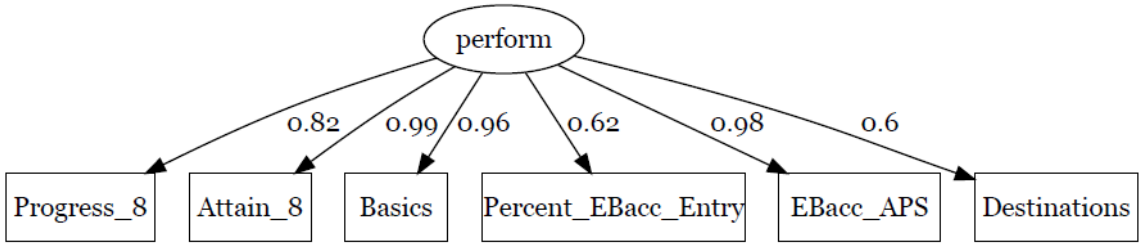


Figure 6. Model diagram for Initial SPM model

Table 18. Model fit indices for Initial SPM Model

GFI	NNFI	CFI	RMSEA
0.824	0.813	0.888	0.321

This model has relatively low fit measures (Table 17), suggesting a single shared construct is not sufficient to reflect the behaviour of the 6 SPMs. This is compared to a model derived from the EFA

analysis, in which three variables, Attain_8, Basics and EBAcc_APS all load strongly onto 1 factor, conceptualised as 'Attain' (i.e. attainment). The other latent constructs, 'Progress', 'Curriculum' and 'Destinations', were each represented by single manifest variables (Progress 8, Percentage EBacc Entry and Destinations SPMs respectively). Due to the low loading of Destinations during EFA and the distinct nature of this measure, it is appropriate to assign it to its own construct for CFA analysis. 'Progress', 'Curriculum' and 'Destinations' are therefore degenerate factors, i.e. they each represent only one of the SPMs. However, in the development of more complex models, these are presented as factors so that we can distinguish between a) the variables and b) the underlying factors, within these models.

The Final SPM Model includes the four SPM factors and all six SPMs (Figure 7). The fit measures (Table 18) suggest that this model does not have a significantly better fit than the Initial SPM Model with a single 'performance' construct. However, this is still considered to be an improved model in that it represents, through the four factors, the four different types of outcome measure included in the SPMs. The latent factor of 'Attainment' is a more useful representation as indicated by the very high factor loadings for Attain_8, Basics and the EBacc_APS. Together they constitute a very strong model for measuring 'Attainment'.

Crucially, the strong loadings of these three SPMs onto the shared 'Attainment' construct also brings into question whether it is either necessary or appropriate to use 3 out of the 6 headline school performance measures to measure this single latent variable, when one metric might have been sufficient and why this specific aspect of school performance has been prioritised over others within the headline performance measures. This question becomes even more important following the announcement by the Department for Education (2025a) that no Progress 8 measures will be published for 2024/25 and 2025/26 as a consequence of the cancellation of primary tests and assessment in 2019/20 and 2020/21 due to disruption caused by COVID-19. Three of the remaining 5 metrics realistically reflect the exact same aspect of school 'performance' and so the breadth of the concept of school 'performance' is even further constrained.

In practical terms, therefore, the headline measures place disproportionate weight on students' examination attainment, repeating the same message across multiple metrics while sidelining other dimensions of school performance. The suspension of Progress 8 in 2025 and 2026 further intensifies this imbalance, reducing the breadth of what counts as "performance" in published data.

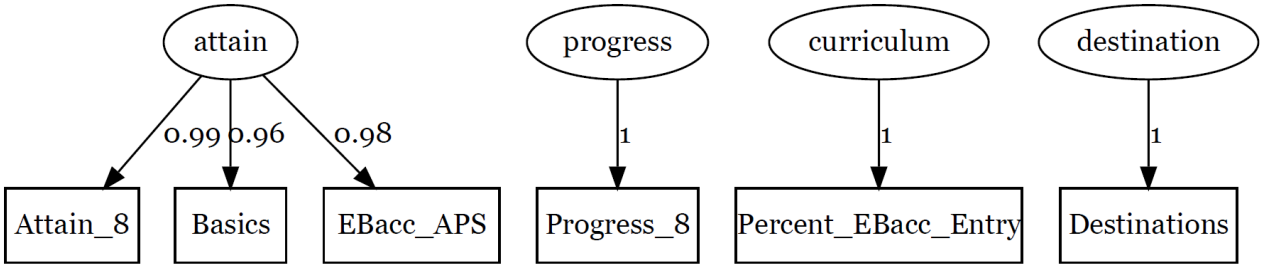


Figure 7. Model diagram for Final SPM model

Table 19. Model fit indices for Final SPM Model

GFI	NNFI	CFI	RMSEA
0.827	0.722	0.889	0.391

6.3 Correlation analysis

I next examine the relationships between each of the 6 headline SPMs using correlation analysis. This aims to explore the breadth of the concept of 'performance' as represented by this collection of performance metrics.

The correlation matrix for combinations of pairs of the headline SPMs (Figure 8. Correlation matrix of all current headline SPMs.) features four main aspects. First, it displays the distribution of each variable on the diagonal. Next, below the diagonal, there are bivariate scatter plots for each pair of SPMs. Third, above the diagonal, the Pearson correlation values are shown for each pair. Finally, each correlation value is accompanied by its corresponding significance level. For each value, the significance is represented by three stars, meaning that all correlations are statistically significant at the 0.1% level (i.e. the p-values for all correlations are less than 0.001). The very strong associations between Attainment 8, Basics and EBacc APS seen in the previous analysis are also suggested by this correlation analysis. While this is relatively unsurprising, given the common basis of each of these metrics, it is perhaps the strength of this correlation (Table 19) that is particularly notable.

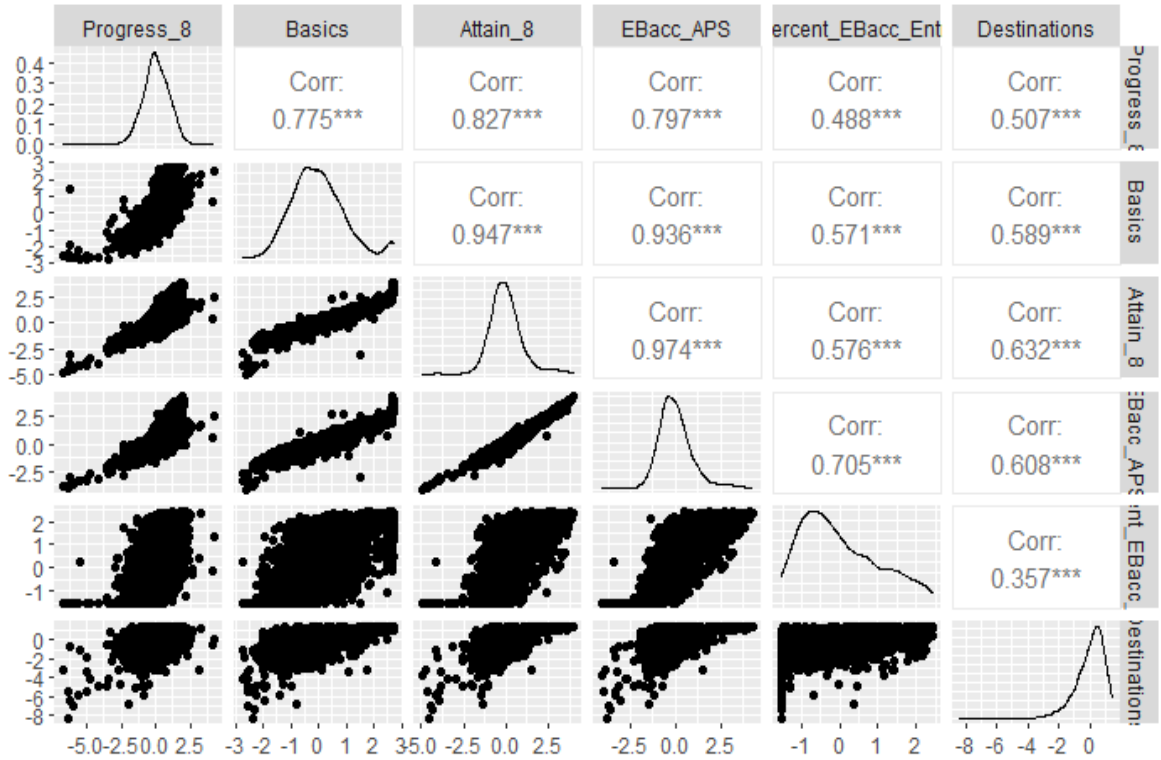


Figure 8. Correlation matrix of all current headline SPMs.

Inspection of the KS4 destinations variable indicated a highly non-normal and left-skewed distribution, as seen in Figure 8, driven by a small number of extreme negative values. Observations below -3 standard deviations (n = 35) were removed and the distribution was then re-examined.

This cleaning procedure reduced the minimum value from -6.705 to -2.742 and reduced skewness from -1.321 to -0.791. The mean also shifted slightly from 0.020 to 0.063, however the median (0.289) and interquartile range (1.166) remained unchanged, indicating that the central mass of the distribution was largely unaffected by the removal of the outliers and that this adjustment primarily removed extreme lower-tail values.

Even after trimming, the distribution remains substantially skewed and highly compressed at the upper end. This reflects a very strong ceiling effect in the underlying destinations measure (mean = 93.76% in the original scale), which limits variability between schools. Consequently, the restricted range is also likely to attenuate observed correlations with other performance measures. This is not resolved through standardisation, as z-score transformation does not alter the underlying distributional constraints of the variable. This limitation is therefore retained as a substantive feature of the measure and considered in the interpretation of subsequent analyses.

Table 20. Correlation between Attainment 8, EBacc APS and 'Basics' SPMs.

Headline SPMs		**Correlation**
Attainment 8	EBacc Average Points Score	0.974
Attainment 8	% grade 5+ in English and maths ('Basics')	0.947
% grade 5+ in English and maths ('Basics')	EBacc Average Points Score	0.936

As suggested by the findings of the 3-factor exploratory factor analysis above, the strong correlation between these three headline SPMs indicates a substantial narrowing of the concept of 'performance' within the 6 headline SPMs. In turn this might suggest some degree of scope for policy to replace one or two of these measures with alternative measures. This would retain the focus on performance in a core of academic subjects, including English and maths, while also introducing greater breadth to the concept of school 'performance'. I will return to these ideas in greater depth in the discussion chapter.

6.4 Relationship between the concept of 'performance' and factors underlying wider school information

Next, I use Structural Equation Modelling (SEM) to create increasingly sophisticated and complex models of the latent factors, aiming to identify hidden patterns and gain a deeper understanding of their associations with SPMs, other school information variables and with each other.

6.4.1 School Information constructs ('School Context' and 'School Character')

A model for school information, encompassing all seven variables (as detailed in section 3.4.2), can be categorised into two distinct groups. First, there are metrics that measure the schools' pupil populations. These are the continuous variables of:

- Key Stage 2 Average Point Score (KS2_APS),
- % of pupils with Free School Meals ever in last six years (FSM6),
- % English as additional language (EAL) and
- % Special educational needs (SEN),

which, for the purpose of this study, I define as the 'School Context' latent variable.

Second, there are those variables which are associated with what can be described as the character of the school rather than its cohort of students. These are the binary variables of:

- school admission policy (selective or non-selective) (Selective),
- whether the school has a religious character (Religious_Char) and
- School gender of entry (mixed or single gender) (Single_Gender),

which are defined here as the 'School Character' latent variable. The Initial School Information Model includes these two latent variables and all seven school information variables (Figure 9).

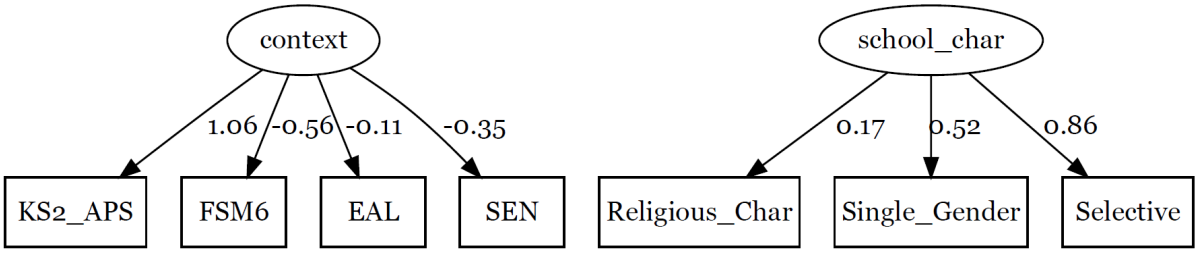


Figure 9. Model diagram for Initial School Information Model.

Table 21. Model fit indices for Initial School Information Model

GFI	NNFI	CFI	RMSEA
0.928	0.749	0.845	0.149

While the GFI for this model (0.928) suggested a good model fit, the other model fit indices bring into question some aspects of this model (Table 20). This, combined with the relatively low factor loadings for three variables, SEN. EAL and Religious_Char, suggests the need to test further models (see Appendix A). The intention here is to refine the model, by iteratively removing those variables with low factor loadings and then observing the effect of the change on fit measures. By following this process over three iterations I arrive at the Final School Information Model (Figure 10).

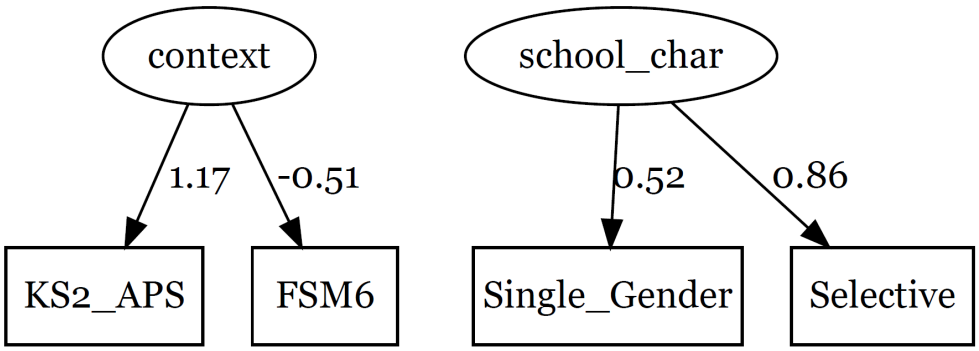


Figure 10. Model diagram for Final School Information Model.

Table 22. Model fit indices for Final School Information Model.

GFI	NNFI	CFI	RMSEA
0.999	0.995	0.999	0.034

The variables which were removed in these steps were EAL, SEN and Religious_Char. One possible explanation for this is the very nature of these variables. SEN, the percentage of students with a special education need, does not, in itself, give a clear indication of the implication of this, with regard to potential consequences for the school's SPMs, for a school and its students. Instead that would largely depend upon the nature, complexity and the severity of the needs of individual students, which are not indicated by this variable (Hanushek et al., 2002; Parsons & Platt, 2014). Similarly, the variable for percentage of students who have English as an additional language, does not provide any direct indication of the effect of this for the school.

Other factors would likely affect this, including, for example, length of residency in England, individual student's English ability, first language (Hutchinson, 2018; Strand & Hessel, 2018) as well as an interaction with other factors e.g., ethnicity, socio-economic disadvantage (Strand et al., 2015) or special educational needs (Parsons & Demie, 2019). These data are, on the whole, currently routinely collected through standard school census returns and made available as part of the National Pupil Database e.g., SEN provision, language, ethnicity, Income Deprivation Affecting Children Indices (IDACI), length of residency in England (which can be deduced from the available data). However, any analysis which sought to fully incorporate these data into the production of school performance metrics, including the interactions of these various factors together, would require the utilisation of such data at a student level.

Whether the cost of this far more granular approach to the processing of school data would be justified in terms of its potential benefits would need to be the subject of further study. However, the constant and rapid advances in computational approaches, in terms of processing power, innovative methodologies and volumes of data, for example, make such a prospect increasingly feasible. Religious_Char was the third variable removed to form the Final School Information Model. There is evidence to suggest that whether a school has a religious character or not is not sufficient to indicate implications for work of schools (Education Policy Institute, 2018). To address this, it would be necessary to consider factors including the specific nature of the religious character, the communities served and interactions with other features for the school e.g., levels of deprivation and special educational needs.

While these data are publicly available, they are largely self-reported by schools and published via the Department for Education's 'Get Information About Schools' (GIAS) website. The accuracy of the information, therefore, depends on schools maintaining and updating their own records, and the Department for Education itself notes that outdated or incorrect entries may result in schools being misrepresented across linked services. As a result, there are reasonable grounds to treat some of the information held within GIAS with caution, particularly given that these data are not routinely subject to independent validation or widespread scrutiny.

6.4.2 SEM: Relationships between latent and observed school factors

Next I use a structural equation modelling approach to construct a model (Figure 11) made up of the 2 final factor models for 'School Character' and 'School Context' and the 4 SPM factors 'Attainment', 'Destinations', 'Curriculum' and 'Progress'. This provides a framework for estimating the strength and significance of the complex relationships between the latent factors and manifest variables and for assessing model fit to further iteratively adapt and refine the model.

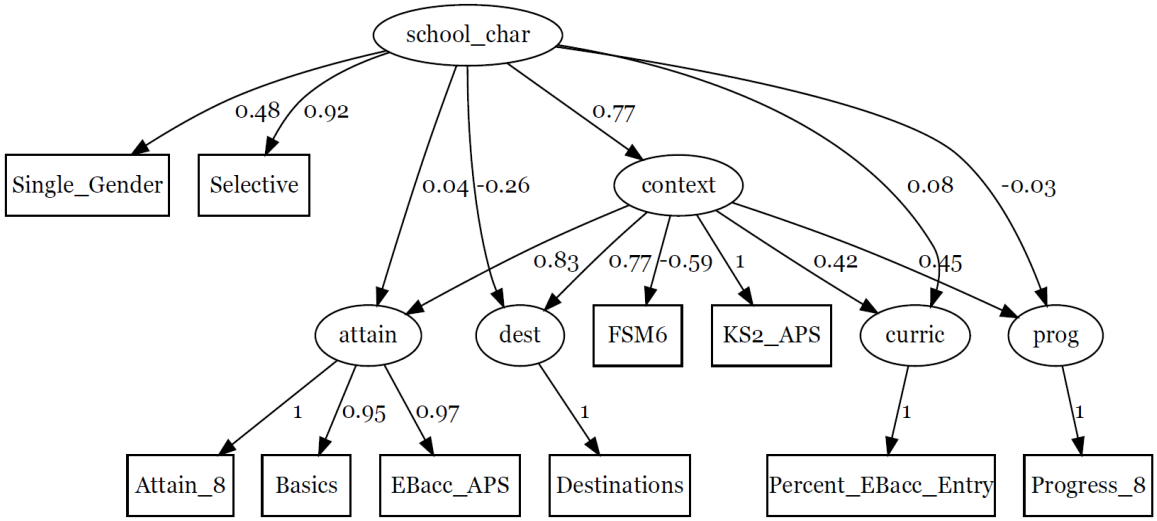


Figure 11. Initial SPM Model – initial SEM model incorporating all four SPM factors as responding to both 'School Character' and 'School Context' factors and 'School Context' responding to 'School Character'.

Table 23. Model fit indices for initial SPM Model 1.

GFI	NNFI	CFI	RMSEA
0.727	0.571	0.687	0.287

Two things are evident from the output of this model. First, the fit measures for the initial SPM Model (Table 22) do not indicate a strong model fit. Therefore, while this model may prove to be a useful starting point, further work is necessary to iteratively move towards a model with a fit to the data which is satisfactory. Second, there were some factor loadings within this model with very low values. In particular, the factor loading for 'School Character' on 'Attainment' (0.04) and 'Curriculum' (-0.03). Therefore, for a refined SPM Model, these regressions of 'Attainment' and 'Curriculum' on 'School Character' are removed (Figure 12).

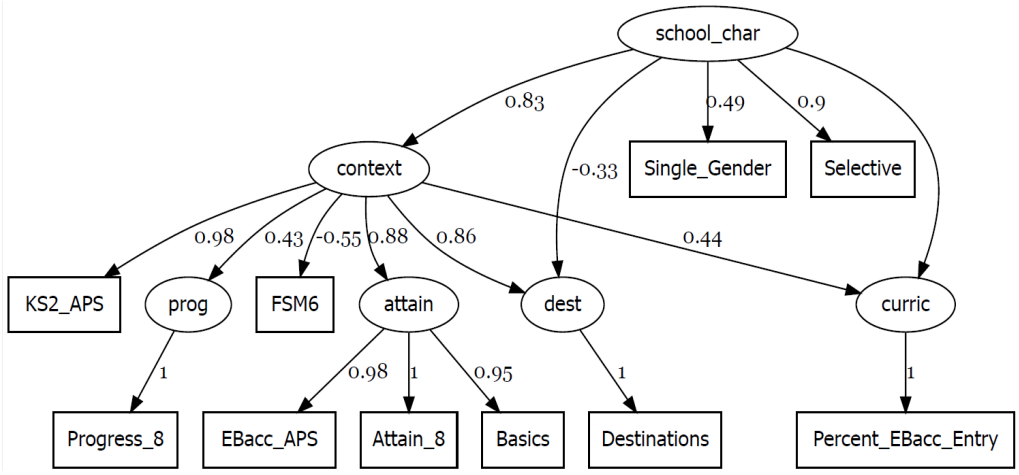


Figure 12. Refined SPM Model – SEM model incorporating all four SPM factors as responding to 'School Context' and only 'Destinations', 'Curriculum' and 'School Context' responding to 'School Character'.

Table 24. Model fit indices for refined SPM Model.

GFI	NNFI	CFI	RMSEA
0.840	0.831	0.906	0.232

The fit measures for the refined SPM Model (Table 23) are stronger than those for the Initial SPM Model (Table 22) but still do not indicate a good model fit. Using the same strategy as employed in the previous iteration, the lowest factor loadings can be identified and then removing from the model. 'School Character' and 'Destinations' (factor loading of -0.33) and 'School Character' and 'Curriculum' (factor loading 0.06), are removed from the refined SPM Model to form a final SPM Model for all SPM latent variables (Figure 13). This final SPM Model for all SPM latent variables has stronger fit measures (Table 24) compared to its predecessor and is more parsimonious. This suggested that, for this dataset, 'School Character', as defined in section 6.4.1, by whether a school is single gender and/or whether it has a selective admissions policy, shares substantial variance with the features of schools' student bodies (with a factor loading of 0.83), in particular the average prior attainment of students at the end of key stage 2 (KS2_APS) and the proportion of students who are disadvantaged, as indicated by the proportion who have been eligible for free school meals at some point in the last six years (FSM6).

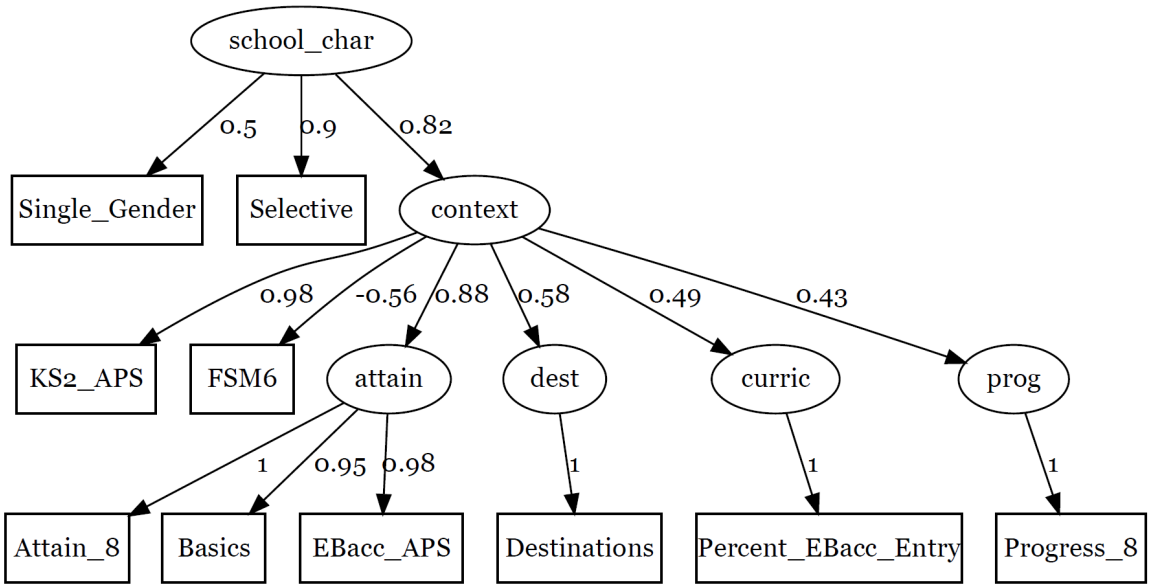


Figure 13. Final SPM Model for all SPM latent variables - incorporating latent variables of 'School Character' and 'School Context'' and all four SPM latent variables ('Attainment', 'Destinations', 'Curriculum' and 'Progress').

Table 25. Model fit indices for final SPM Model for all SPM latent variables.

GFI	NNFI	CFI	RMSEA
0.831	0.839	0.904	0.226

The final SPM Model for only the 'Attainment' SPM latent variable (Figure 14) is a further refinement of the final SPM Model for all SPM latent variables (Figure 13). This model allows me to focus solely on attainment within the 4 'performance' factors. It emphasises the relationship between the 'School Character' and 'School Context' factors and the primary SPM latent variable, 'Attainment', with only RMSEA slightly outside the recommended threshold for a good model fit (Table 25), indicating a very good fit overall. This suggests that (a) whether schools are single gender and (b) whether they have a selective admissions policy (here captured under the single factor, 'School Character') has a strong statistical association with both the levels of prior attainment and the proportion of disadvantaged students within schools (i.e., 'School Context'). 'School Context', in turn, is a strong indicator of school 'Performance', as reflected by 'Destinations', 'Progress' and 'Curriculum' and is a very strong indicator of the 'performance' factor 'Attainment'. Most notably, factors beyond the school's provision are shown to have a very strong impact on the metrics currently used as indicators of school 'Performance'. It appears, therefore, that SPMs may well measure something meaningful about the schools. However, since the factors described here as 'School Character' and 'School Context' have no clear, direct relationship to the quality of school

provision, this raises serious questions about the extent to which the metrics underlying the six headline SPMs truly measure school 'performance'. This is crucial, as SPMs are intended to act as proxies for school effectiveness, informing parental choice, inspections, interventions and school improvement planning.

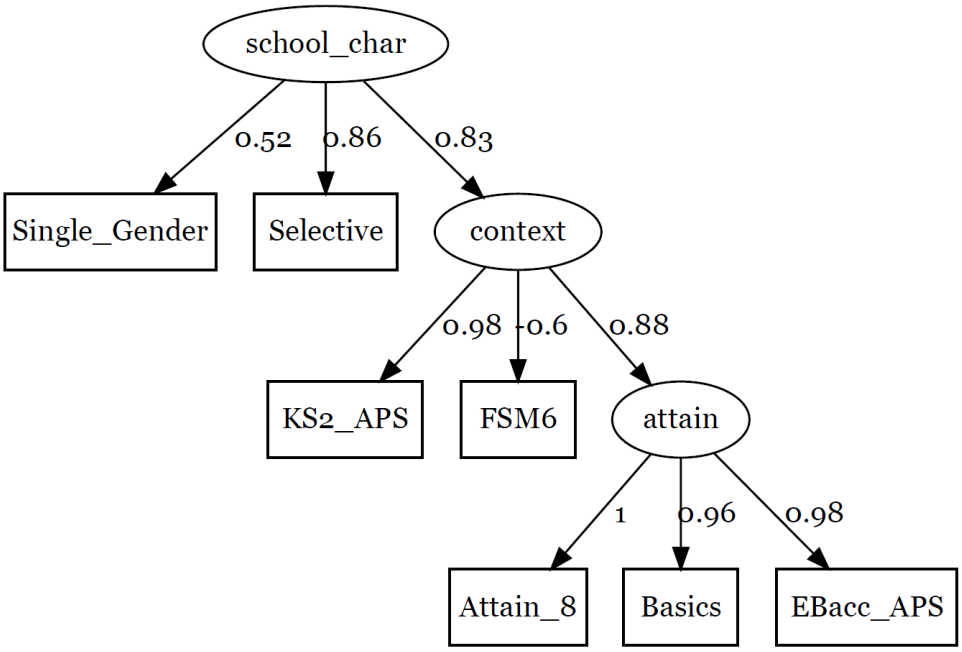


Figure 14. Final SPM Model for only the 'Attainment' SPM latent variable - incorporating latent variables of 'School Character' and 'School Context'

Table 26. Model fit indices for final SPM Model for only the 'Attainment' SPM latent variable

GFI	NNFI	CFI	RMSEA
0.964	0.973	0.985	0.103

6.5 Multiple regression analysis of school context and school characteristics variables

A series of multivariable regression models were estimated to examine the independent associations between school context and school character variables and six headline performance measures. The analysis was undertaken as a robustness check and to assess whether the relationships identified in other analyses were consistent when all predictors were considered simultaneously.

The standardised regression coefficients (ß) for SPMs can be found in Table 26

Table 27. Standardised regression coefficients (ß) for SPMs

Predictor	Attain 8	Progress 8	Basics	EBacc APS	% EBacc Entry	Destinations
KS2 APS	0.726***	0.225***	0.71***	0.699***	0.445***	0.395***
FSM6	-0.202***	-0.376***	-0.218***	-0.199***	-0.075***	-0.432***
EAL	0.219***	0.408***	0.238***	0.255***	0.295***	0.112***
SEN	-0.034***	-0.077***	-0.033***	-0.031***	-0.021	-0.021
Selective	0.007	-0.08***	-0.034**	0.025*	-0.026	-0.149***
Religious Character	0.009	0.012	-0.003	0.006	-0.001	0.031*
Single Gender	0.064***	0.121***	0.068***	0.07***	0.095***	0.067***

6.5.1 School Context effects

Across all outcome measures, prior attainment (KS2_APS) was the strongest and most consistent predictor. Standardised effects were particularly large for Attainment 8 (β = 0.726), Basics (β = 0.710) and EBacc APS (β = 0.699), and remained substantial for Destinations (β = 0.395) and Progress 8 (β = 0.225).

Free School Meal eligibility (FSM6) was consistently negatively associated with all outcomes, with the strongest effects observed for Destinations (β = -0.432) and Progress 8 (β = -0.376), indicating a persistent socioeconomic gradient across all SPMs.

In contrast, English as an Additional Language (EAL) showed a consistent positive association across all outcomes after controlling for KS2 attainment and FSM6. Effects were strongest for Progress 8 (β = 0.408) and EBacc APS (β = 0.255), and remained positive across all six measures, including Attainment 8 (β = 0.219) and EBacc entry (β = 0.295). Special educational needs (SEN) showed small and consistently negative associations across outcomes, with effects close to zero for EBacc entry and Destinations.

6.5.2 School character effects

School character variables showed more limited and variable associations with performance measures.

Selective schools showed no meaningful association with Attainment 8 (β = 0.007) and a negative association with Progress 8 (β = -0.080) and Destinations (β = -0.149) after controlling for intake characteristics. Religious character showed no consistent pattern, with small effects that were generally non-significant across attainment measures.

Single-sex schools, by contrast, showed small but consistent positive associations across all outcomes, including Attainment 8 (β = 0.064), Progress 8 (β = 0.121), EBacc APS (β = 0.070) and EBacc entry (β = 0.095).

6.5.3 Patterns across headline SPMs

A clear pattern emerges in the relative strength of associations across outcomes. Attainment-based measures (Attainment 8, Basics, and EBacc APS) were most strongly associated with KS2_APS, whereas Progress 8 showed a comparatively weaker association with prior attainment and stronger associations with FSM6 and EAL.

Destinations were most strongly associated with FSM6 (β = -0.432) and showed comparatively weaker associations with attainment-related predictors, suggesting that post-16 progression outcomes are more sensitive to socioeconomic composition than exam-based measures.

6.5.4 Residual structure among headline SPMs

Despite controlling for contextual and organisational variables, substantial residual correlations remained among headline SPMs, indicating strong shared structure beyond observed predictors.

Two distinct patterns are evident. First, a cluster of attainment-based measures—Attainment 8, Basics, and EBacc APS—exhibited consistently high residual correlations ($r \approx 0.79–0.90$), suggesting substantial overlap in the variance captured by these indicators even after accounting for intake and

school characteristics. This pattern is consistent with these measures reflecting a closely aligned underlying attainment construct.

Second, an exceptionally strong residual association was observed between Attainment 8 and Progress 8 ($r \approx 0.99$). This indicates near-complete alignment between attainment and value-added performance measures at the school level, even after adjustment for contextual and organisational factors. This coupling suggests that distinctions between attainment-based and progress-based accountability measures may be more limited in practice than their conceptual definitions imply.

Overall, these residual structures indicate that a substantial proportion of shared variance among headline performance measures is not explained by observed school characteristics, and instead reflects deeper structural relationships among the performance indicators themselves.

Table 28. Key residual correlations among headline SPMs (SEM model)

Pair	Residual correlation (r)
Attainment 8 – Basics	0.84
Attainment 8 – EBacc APS	0.90
Basics – EBacc APS	0.79
Attainment 8 – Progress 8	0.99

The persistence of strong residual associations among headline performance measures suggests that school outcomes retain a substantial degree of shared structure even after accounting for differences in intake and organisational characteristics. In particular, the close alignment between attainment-based measures and the near-equivalence of Attainment 8 and Progress 8 indicate that schools tend to vary along a limited number of underlying performance dimensions rather than across fully independent indicators.

This residual interdependence raises the possibility that schools may cluster into relatively coherent profiles of performance rather than exhibiting independent variation across each headline measure. Cluster analysis is therefore used in the following section to explore whether such underlying structure is reflected in empirically identifiable groupings of schools, and to examine how these groupings relate to differences in school context and composition.

6.6 Cluster Analysis

I have applied cluster analysis to explore groupings of schools based on variables related to school context and character and to examine patterns around SPMs for any clusters discovered. I aim to investigate how grouping schools based on their common features might offer new insights into any inherent advantages and disadvantages for schools in terms of their performance metrics. This cluster analysis extends the preceding phases of the study by moving from the structure and meaning of school performance measures to the distribution of schools within that structure. While the thematic analysis established how SPMs are defined and justified in policy discourse and the factor analysis and structural equation modelling identified the latent dimensions underpinning these measures, cluster analysis groups schools according to shared contextual and organisational characteristics. In doing so, it provides a means of examining how the operation and consequences of SPMs vary across different types of schools and whether patterns of advantage and disadvantage are systematically associated with particular school contexts rather than school effectiveness alone.

In my explanation of the findings from this Cluster Analysis, I will first describe the structure of the identified clusters. I will then discuss the features of the clusters, followed by some approaches to visualise these cluster features. Finally, I consider an evaluation of the quality of the clusters.

6.6.1 Agglomerative hierarchical clustering

Agglomerative hierarchical clustering analysis is used with the 3125 mainstream state secondary schools in England, using standardised versions of the seven variables related to wider school information (KS2_APS, FSM6, EAL, SEN, Selective, Religious_Char and Single_Gender). I use Euclidean distance as the measure of similarity between data points and Ward's linkage.

The initial analysis reveals 4 distinct clusters of schools. However, one cluster is found to be almost entirely composed of selective schools (99.4%). Since there is one binary variable which indicated whether the schools had selective admissions policies, the cluster analysis is repeated but with the 'selective' variable removed to test whether the 'selective' cluster is the direct result of the presence of this variable. This leaves only the 6 variables for wider school information (KS2 Average Point Score, % of pupils with Free School Meals ever in the last six years, % English as additional language, % Special educational needs, whether the school has a religious character and school gender of entry - mixed or single gender). Again 4 clusters are identified. In line with hierarchical clustering methodology, this can be adjusted, to form more or less clusters. In this case 4 clusters are chosen as the parallel scree plot (Figure 15), using the result of a PCA , shows variances for each principal component suggesting an elbow between 3 or 4 and the point that the eigenvalues from the actual

data cross fell below the eigenvalue distribution from the resampled data was also 4. In addition, the % of variance explained by 4 dimensions was satisfactory at 92.3% (Figure 16). Finally, a silhouette width plot (Figure 17) identified 4 as the optimal number of clusters.

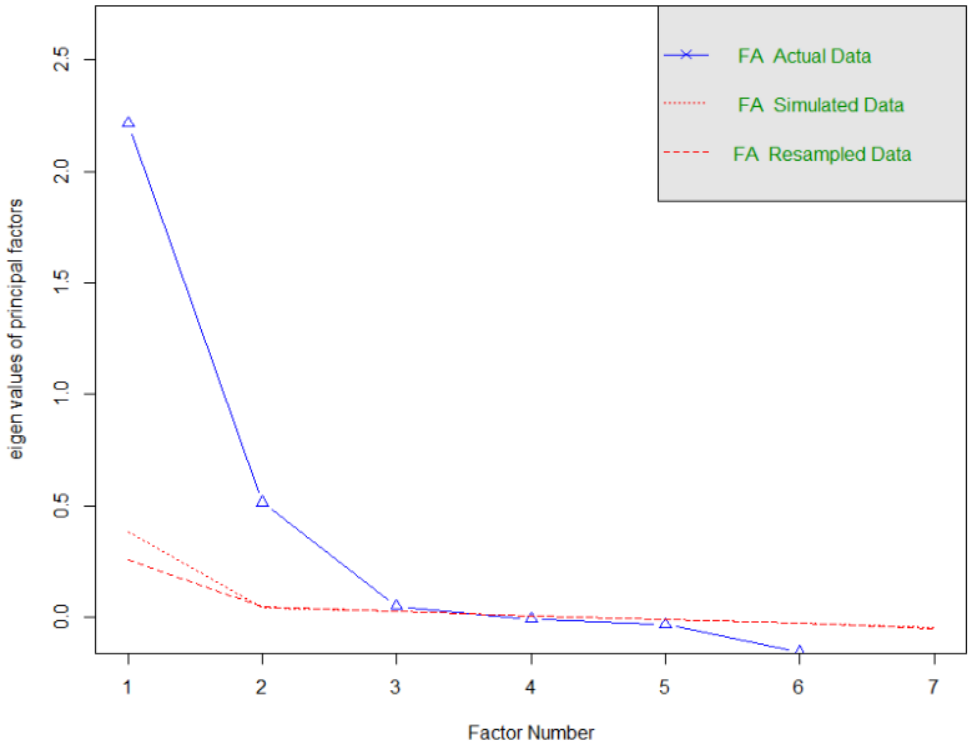


Figure 15. Parallel scree plot.

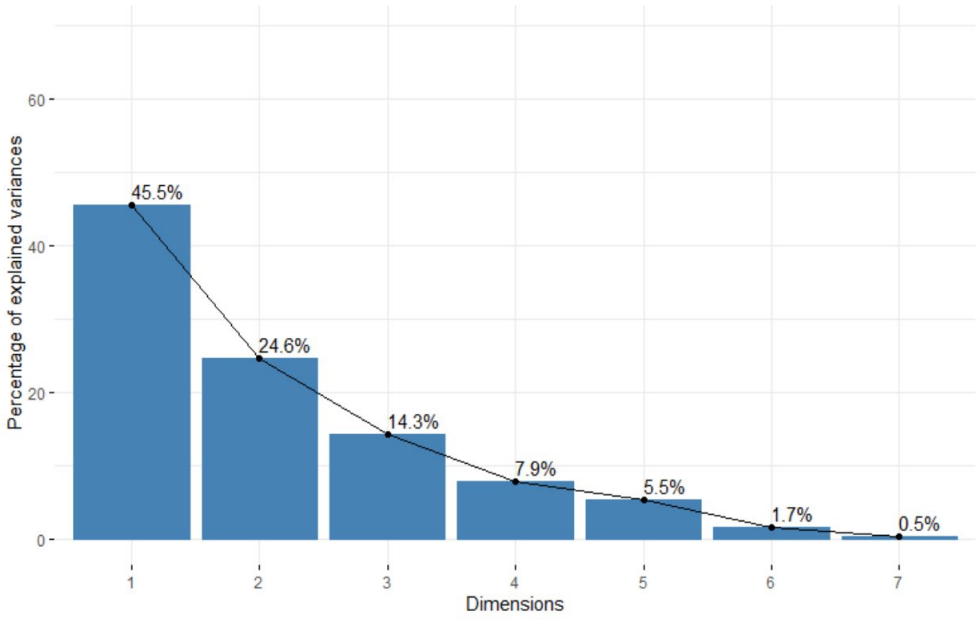


Figure 16. Percentage of variance explained by each dimension.

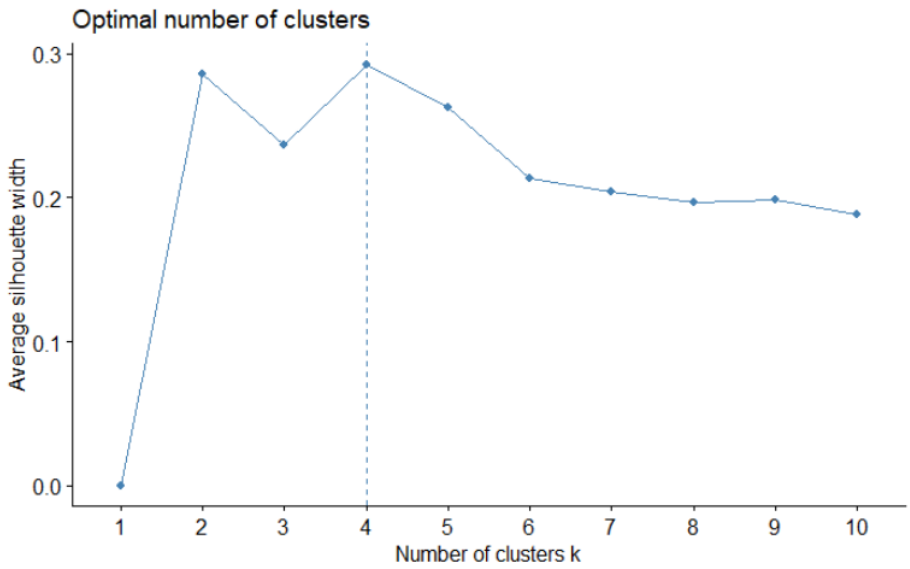


Figure 17. Average silhouette width plot - indicating optimal number of clusters to be 4.

6.6.2 Summarising the features of the clusters.

After running the hierarchical clustering for 6 school information variables, the mean score for each variable for each cluster is gathered for further analysis and interpretation (Table 27). This is then summarised (Table 28).

Table 29. Number of schools in each cluster.

Cluster	1	2	3	4
Number of schools	1415	1292	255	163

Table 30. Mean value of each variable for clusters 1 to 4 (overall mean is micro-averaged, i.e. it is calculated over schools rather than clusters and so is weighted to incorporate the different number of schools in each cluster).

		Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**	***Overall mean***
Wider school information variables	KS2_APS	102.15	104.78	102.14	112.75	*103.79*
	FSM6	34.32	17.44	41.93	6.98	*26.54*
	EAL	13.28	10.79	66.28	16.35	*16.74*
	SEN	18.17	11.67	13.53	6.16	*14.48*
	Selective	0.00	0.15	0.00	98.77	*5.22*
	Religious_Char	51.24	54.64	49.41	83.44	*54.18*
	Single_Gender	2.19	11.92	20.00	71.17	*11.26*
SPMs	Attain_8	43.85	52.65	48.88	74.19	*49.48*
	Progress_8	-0.25	0.13	0.26	0.57	*-0.01*
	Percent_EBacc_Entry	29.37	42.26	50.95	77.70	*38.98*
	EBacc_APS	3.73	4.64	4.38	6.99	*4.33*
	Basics	39.05	56.79	50.21	94.01	*50.16*
	Destinations	91.92	95.58	92.44	99.06	*93.85*

Table 31. Summary of each variable for clusters 1 to 4.

		Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**
Wider school information variables	KS2_APS	low	average	low	very high
	FSM6	very high	low	very high	very low
	EAL	average	low	very high	average
	SEN	high	low	average	very low
	Selective	no	no	no	yes
	Religious_Char	average	average	average	high
	Single_Gender	low	average	average	very high
SPMs	Attain_8	low	high	average	very high
	Progress_8	low	high	high	very high
	Percent_EBacc_Entry	low	high	high	very high
	EBacc_APS	low	high	average	very high
	Basics	low	high	average	very high
	Destinations	low	high	low	very high

These clusters can be summarised as:

Cluster 1 – high % disadvantaged, low prior attainment, high % SEND – **in general these schools are below average all SPMs**

Cluster 2 – low % advantaged, higher prior attainment, low % SEND and EAL - **in general these schools are above average all SPMs**

Cluster 3 – high % disadvantaged, high % EAL, low % SEND, low prior attainment – **in general these schools are broadly average for 4 SPMs, with % EBacc Entry and EBacc APS very significantly above national average**

Cluster 4 - Selective, advantaged, very high prior attainment, low %SEND – **in general these schools are significantly above average for all SPMs**

It is perhaps worth emphasising at this point an important feature of this analysis. This is something which will be discussed in greater depth in the Discussion chapter. The 4 clusters being considered here are formed from the similarities in the variables related to school intake (i.e. 'School Context')

and the nature of the school (i.e. 'School Character'). However, further analysis reveals additional patterns, within this identified cluster arrangement, in the schools' headline SPMs. This suggests statistical links between the school information variables and school performance measures, something I have explored previously using structural equation modelling techniques (section 5.4.2). This is further supported by the regression analysis, which indicates that %EAL retains an independent association with headline performance measures over and above other contextual variables. These analyses have revealed the strong association of context and character with the 3 attainment SPMs (Attainment 8, %Basics and Ebacc APS). I will consider, in the discussion chapter, the implications of this for both the concept of 'performance' and for the question of fairness in the existing SPMs when judging and comparing schools with different contexts and characters.

6.6.3 Visualising the cluster features

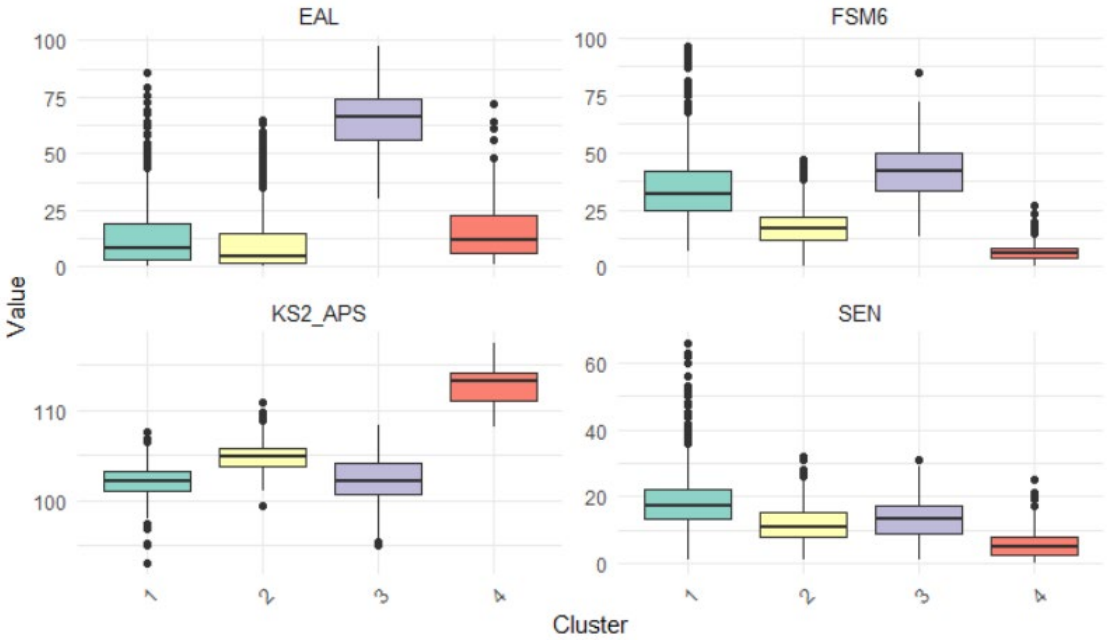


Figure 18. Boxplots to Show Distribution of Variables by Cluster.

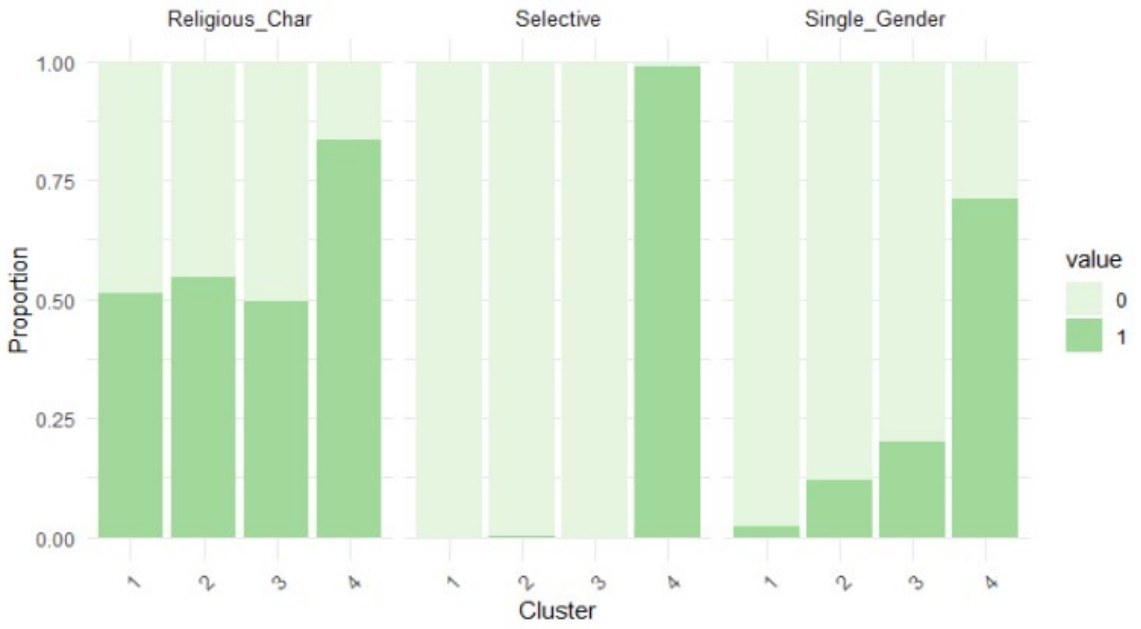


Figure 19. Stacked Bar Charts to Show Distribution of Variables by Cluster.

Visualisations of the distributions of all of the thirteen variables across the four clusters are created using boxplots (Figure 18 and Figure 20). These included crucial information about the range, the median, the inter-quartile range and any outliers and summarise visually the differences in the clusters, so aiding in their interpretation.

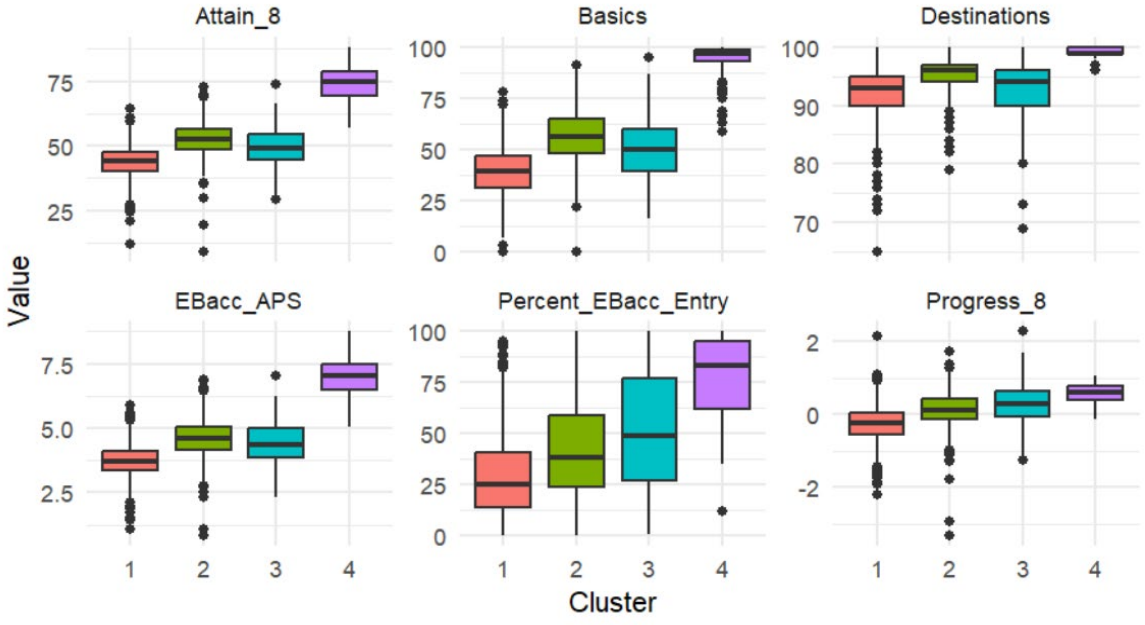


Figure 20. Boxplots of the distributions of mean scaled values of the 6 SPM variables for clusters 1 to 4.

An additional radar plot visualisation of the clusters is created (Figure 21). This can be used to both illustrate the complex interplay of the variables in each of the four clusters as well as providing a

means of identifying their distinguishing features, both in terms of the school information variables which were used to create them and in the SPMs for schools in each cluster. This visualisation enables direct comparison of the clusters across multiple dimensions simultaneously, making patterns of similarity and difference between groups of schools immediately visible in a way that is not easily captured through tables or univariate summaries.

By displaying the relative profiles of clusters rather than individual measures in isolation, the radar plot helps to reveal how combinations of school contexts, characteristics and performance measures co-occur within different clusters of schools. These patterns provide a basis for interpreting how schools occupying different structural and contextual positions are differentially represented within the SPM framework and suggest potential policy implications regarding the fairness, interpretation and use of performance measures across heterogeneous school contexts.

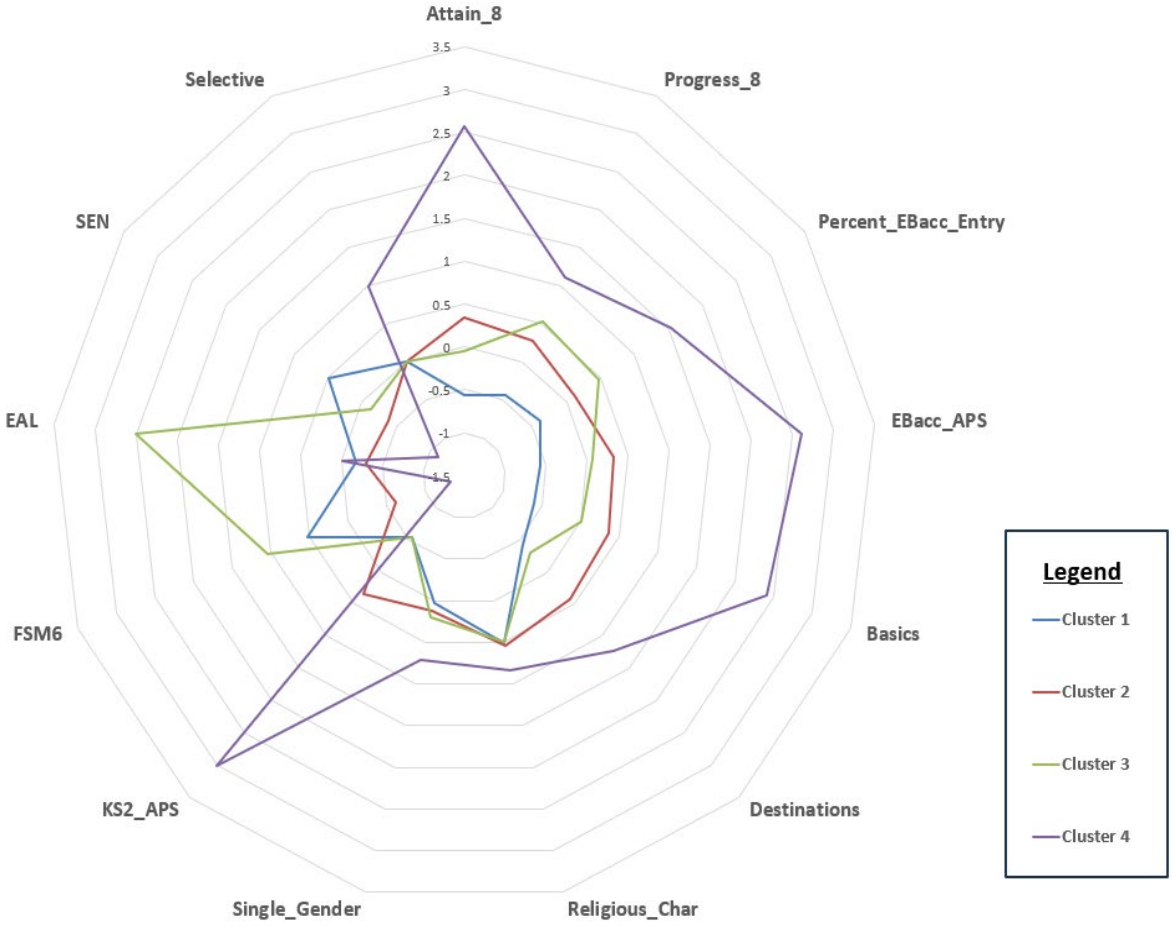


Figure 21. Radar plot of the mean scaled values for each of the thirteen variables for Cluster 1 -4 formed using hierarchical clustering of the seven school information variables.

6.6.4 Evaluating the clusters

A range of approaches is used to evaluate the quality of the clusters in this analysis. Evaluating the quality of clusters is necessary to assess whether the identified groupings represent meaningful and robust structures in the data rather than artefacts of a particular algorithm or parameter choice. Using a range of evaluation approaches allows different aspects of cluster quality to be examined—such as internal cohesion, separation between clusters and stability—providing a more comprehensive and reliable assessment than reliance on a single metric. This is important since clustering is an exploratory technique with no single 'correct' solution. This triangulation strengthens confidence that the resulting clusters capture substantive differences between schools that are relevant for interpretation and policy analysis.

The average silhouette width is found to be 0.21. Since this is a measure of how similar each point is to its own cluster compared to other clusters, this, on its own, might suggest that there is no substantial structure in the clustering. The silhouette plot (Figure 22) also suggests some potential issues with assignment in cluster 1 and, to a lesser extent, cluster 2, with significant numbers of schools with negative silhouette coefficients.

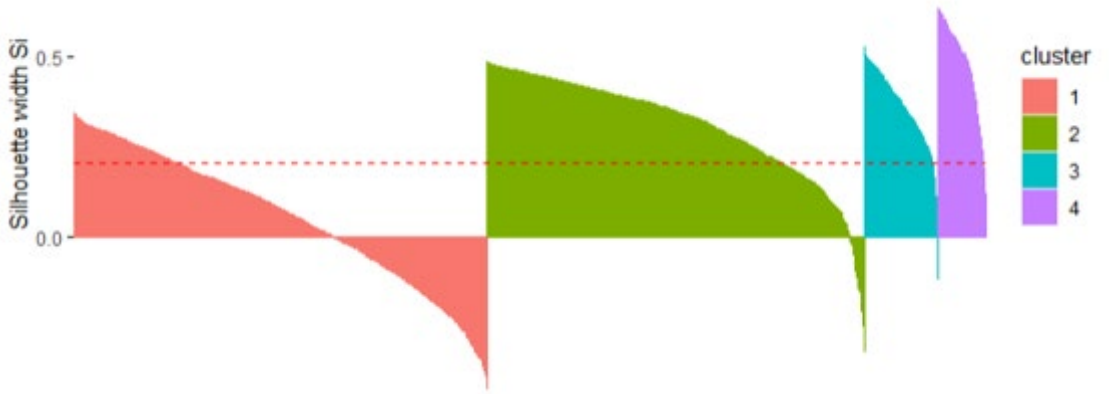


Figure 22. Silhouette Plot for agglomerative hierarchical clustering using Ward's minimum variance method for all English state schools and based on six variables of 'wider school information'.

The Dunn Index, which measures the ratio of the smallest distance between observations not in the same cluster to the largest intra-cluster distance, was relatively low at 0.01733. This again suggests that clusters may not be very separated or compact. However, the Calinski–Harabasz index indicated a relatively high value (890). This suggested, contrary to the other metrics above, that the clusters are both compact within themselves and also distinct from one another.

Finally, t-Distributed Stochastic Neighbour Embedding (t-SNE) is used as a method of dimension reduction and then visualization. The t-SNE visualisation (Figure 23) suggests some degree of distinct structure for the four clusters generated by agglomerative hierarchical clustering. At the same time, it raises some potential questions around the structure of cluster 2. The relatively higher proportion of schools with negative silhouette widths for this cluster indicates that some of these schools may be misclassified or lie at the boundary between clusters i.e. where the distinction is less clear. This suggests a potential area for further investigation.

These different approaches to clustering evaluation did not provide a consistent assessment of the quality of the clusters formed. I therefore needed to accept that, while this analysis suggests some potential in the use of these clustering techniques to further explore the relationships between SPMs and wider 'School Character' and 'School Context' metrics, further work would be required to refine and evaluate these techniques. I will return to a discussion of the benefits and limitations of these clustering evaluation approaches below.

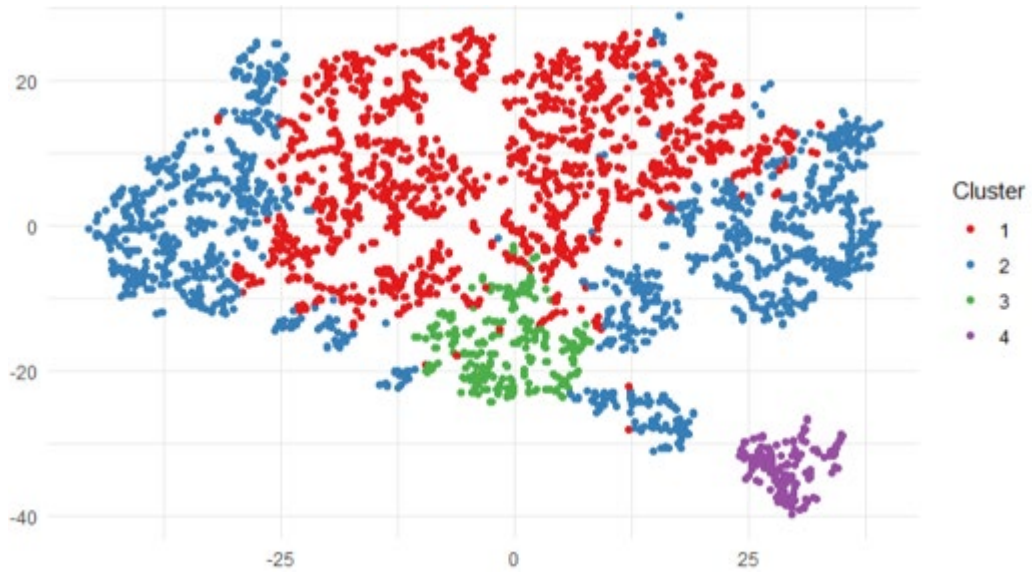


Figure 23. t_SNE Visualisation for agglomerative hierarchical clustering using Ward's minimum variance method for all English state schools and based on 6 variables of 'wider school information'.

6.7 Conclusion to Analysing the Structure of SPMs using Quantitative Analysis

The analyses reveal that headline secondary school performance measures (SPMs) capture a narrower conception of 'performance' than intended. Exploratory and confirmatory factor analyses indicate that attainment dominates the SPMs, with Progress 8, EBacc entry and Destinations contributing separately but weakly, highlighting redundancy and raising questions about the validity of using these metrics to represent overall school effectiveness. This study contributes originally by empirically constructing a latent factor model of school performance, quantifying the relative contribution of each headline measure and revealing how redundancy reinforces the priority of attainment in policy. Correlation analysis further confirms that attainment-related measures largely overlap, narrowing the operationalised definition of 'performance'. Exploratory and confirmatory factor analyses indicate that attainment dominates the SPMs, with Progress 8, EBacc entry and Destinations contributing separately but weakly, highlighting redundancy and raising questions about the validity of using these metrics to represent overall school effectiveness.

This finding contributes to ongoing debates in the educational measurement literature that critically examine the statistical properties and limitations of value-added and accountability metrics, especially Progress 8, in capturing school effectiveness beyond student intake characteristics (Prior et al., 2021; Leckie & Goldstein, 2018b). Research has documented that measures like Progress 8, while adjusting for prior attainment, still leave substantial influence from contextual factors such as socio-economic background, social composition and school intake, complicating their interpretation as indicators of school quality (Leckie & Goldstein, 2018b; Gill, 2024). Building on this, my analysis offers a novel empirical demonstration of how these latent constructs are dominated by attainment, with non-attainment dimensions contributing minimally, providing a clearer, quantified picture of the narrow conceptualisation of performance embedded in SPMs. In this context, the current analysis extends these insights by empirically demonstrating the dominance of attainment-related constructs within the latent structure of headline SPMs and showing how contextual factors overwhelmingly shape these outcomes.

Structural equation modelling shows that school context, primarily prior attainment and socio-economic disadvantage, strongly drives all SPM outcomes, while school character (selectivity, single-gender status) has a limited direct effect. This underscores that headline SPMs predominantly reflect student intake rather than the quality of school provision, challenging the assumption that they measure school effectiveness. The original contribution here lies in applying SEM to school-level performance data to confirm that contextual effects dominate SPMs, offering an alternative to

student-level multilevel modelling used in prior research and demonstrating that SPMs themselves encode contextual bias. Cluster analysis reinforces these findings, with distinct school groups aligning closely with SPM outcomes, further demonstrating how context shapes performance patterns. By linking clusters of schools to SPM outcomes, this work provides new empirical evidence for patterned disadvantage and suggests the potential for context-sensitive interpretation of SPMs. This would constitute a novel approach within the literature.

In relation to the research sub-questions:

- ***What underlying latent constructs can I empirically identify and conceptualise as shaping 'performance' within SPMs?*** Four factors—Attainment, Progress, Curriculum and Destinations—emerge as underlying dimensions, with attainment overwhelmingly dominant.
- ***With reference to the current headline performance measures, what is the extent and breadth of the concept of 'performance'?*** Current SPMs largely capture attainment, providing a narrow representation that omits many broader aspects of school performance.
- ***To what extent do SPMs encapsulate elements beyond the influence of schools and can statistical techniques aid in assessing their influence on the stated policy goals of SPMs?*** Analyses show that student intake and other contextual factors are strongly associated with SPMs, highlighting the limitations of these measures. The digital tools used here, such as factor analysis, SEM and cluster analysis, indicate potential future approaches for disentangling these potential influences and assessing alignment with policy goals.

Overall, the findings indicate that headline SPMs inadequately capture the multidimensional nature of school performance, are heavily context-dependent and risk misrepresenting school effectiveness. These results provide a strong empirical basis for considering more nuanced, contextualised and holistic measures in future SPM policy and accountability systems.

7. Findings from Phase 3: Thematic analysis of focus group discussions on the subject of school performance measures

The focus of much research on SPMs has been on the statistical methods used to generate these metrics. In this research phase, I instead aim to consider the purpose of SPMs. From this, I then examine their effectiveness as well as their wider effects. To do this, I conduct a series of focus groups with various stakeholders. Focus groups are carried out in small groups (between 2 and 4 participants), with each participant contributing to each of the 3 rounds of focus groups. I then carry out a qualitative, thematic analysis of the transcriptions of the discussions.

The only comparable data and associated analysis, which seeks to better understand stakeholder opinions of the English school accountability system, is in the form of national consultations, typically conducted by or on behalf of a government body, eg DfE or Ofsted (Department for Education, 2025d; Ofsted, 2021). In September 2024, Ofsted published the findings of what they called the Big Listen, a national consultation on the work of Ofsted (Bessant et al., 2024). The DfE and Ofsted consultations have some common ground with this analysis. These consultations and this analysis all consider the general principles of school accountability. However, the DfE consultation also focused on the development of school profiles and on approaches to school improvement, while the Ofsted consultation covered suggested improvements to the school inspection process. Neither of these specifically addressed SPMs, the primary focus of this work.

In seeking to address the overarching research question for this thesis - **How well do secondary school performance measures in England serve their stated purposes and what does empirical analysis reveal about how they might be improved?** – I have generated 3 sub-research questions specifically for this research phase involving stakeholder focus groups. These are:

a) What do school stakeholders understand to be the purposes of secondary SPMs in England?

b) How do stakeholders perceive the effectiveness and wider effects of SPMs in practice?

c) What alternative approaches and improvements do stakeholders propose for the design, communication and use of SPMs?

7.1 Overarching themes from thematic analysis of focus group transcripts

This section synthesises the thematic analysis of the focus group discussions to identify overarching themes related to the purposes, effectiveness, effects and alternative approaches to school performance measures.

The analysis of the transcripts of the focus group meetings resulted in the definition of 4 overarching themes. These are:

Overarching Theme 1 Purposes, Principles and Priorities of SPMs

Overarching Theme 2 The Fundamental Flaws of Current Performance Measurement

Overarching Theme 3 The Detrimental Impact on Equity and Practice

Overarching Theme 4 Towards Holistic, Supportive and Contextualised Accountability

This analysis aims to address two research questions. First, there is Research Question 2: To what extent is the production, publication and operationalisation of secondary school performance measures in England effective in meeting their stated aims in this context? Discussions were initiated with focus group participants regarding the purpose, effectiveness and effects of SPMs in this regard. Second, there is Research Question 3: How can the findings of research and advances in technology be utilised to improve the communication and use of school data by stakeholders for school performance measurement purposes? Participants were encouraged to consider alternative approaches to the existing SPMs, which might help address the issues they have raised.

Overarching Theme 1 focuses on discussions aimed at uncovering the fundamental purposes of SPMs. This builds on and complements the findings from phase 1 (chapter 4) and establishes a foundation for assessing the effectiveness and effects of the current SPMs, which underpin Overarching Themes 2 and 3 and build upon the findings from phase 2 (chapter 5). Then, Overarching Theme 4 extends this by suggesting possible steps forward to address the issues with SPMs that have been identified. These 4 overarching themes are further subdivided into subthemes.

These themes and their subthemes will now be considered in turn.

7.1.1 Overarching Theme 1: Purposes, Principles and Priorities of SPMs

This first group of subthemes relate to the explicit purposes, underlying principles and stated priorities of SPMs, but, crucially, also begins to suggest how issues around these ideas might be

ethically and practically addressed. The discussions which relate to this theme are informed by and build upon the thematic analysis of government documentation (section 4.2). In these thematic analyses, I seek to address Research Question 2. By identifying the stated objectives of SPMs, I can then assess the effectiveness of the production, publication and operationalisation of secondary school performance measures in England in achieving these aims.

7.1.1.1 Accountability (general)

Some participants initially associate SPMs with the wider concept of school accountability. For example, Participant 6 describes SPMs as "*some kind of accountability for schools*" and Participant 7 says that "*we have got big themes of accountability*". After further discussion and prompts, the meaning of 'accountability' is expanded and clarified. Some of these clarifications will form the other subthemes below.

The main aspect of 'accountability' discussed is the inspection of schools as part of a "*continuous cycle of monitoring from Ofsted*" (Participant 10). This role for Ofsted, the schools' inspectorate for England, is challenged by Participant 15, who notes that, based on the existing framework (note: a new inspection framework is scheduled for introduction in November 2025), "we have an inspection system now that doesn't rely on data", with the focus since 2019 for judging the quality of education being on "inspecting the substance of education" (Ofsted, 2019a) through a focus on the "Intent, Implementation and Impact" (Ofsted, 2019a, p.10) of the curriculum. Participant 16 also adds that SPMs and Ofsted are both parts of a broader school accountability system, with SPMs "*contextualised as part of an inspection report*".

To explore accountability in more detail, the following sections examine two interrelated subthemes: 'Democratic good (transparency)', which emphasises the importance of making information about schools publicly accessible and open to scrutiny, and 'Part of the culture of schools', which considers how accountability is embedded in everyday practice.

Democratic good (transparency)

Some discussions about the purpose of SPMs highlight their publication as part of a broader classification of a 'democratic good'. From this perspective, accountability is positioned as a democratic good, premised on transparency and the public availability of information, with SPMs functioning as a mechanism through which schools are rendered visible and open to scrutiny. It also relates to the process through which citizens are afforded a right to information about "*publicly funded institutions"* (Participant 11) and the fact that schools are "*ultimately accountable for the use of taxpayers' money*" (Participant 10). One participant summarises this as being a "*Freedom of Information type thing"* (Participant 16). These comments suggest that the publication of SPMs arises

from a broader principle of transparency and aims to inform the public and local community about schools' work, regardless of other related purposes discussed below. In these discussions, little attention is given to the wider aspects, beyond publication, of data transparency practice. This might typically involve a more comprehensive approach to data governance and communication, ensuring data are not only available but also understandable, trustworthy and usable (Meijer, 2014).

Part of the culture of schools

Some participants reflect on how SPMs have shaped the general culture of schools by embedding accountability as a taken-for-granted feature of schooling, influencing how schools understand their purpose, priorities and responsibilities. Much of this debate is around the perception that SPMs have had an increasing influence on schools over their history, from being first introduced in 1992 to today and that their existence and to some extent their influence, was seldom questioned.

> *"It feels very common sense to us now that we hold schools accountable to kind of headline performance measures made up of GCSEs etc. We are used to it and many have never known schools without them. There was a time when that was actually unthinkable and was seen as completely beyond the role of government to do that"*. Participant 9

This speaks of the embedding of practices in schools that are based around SPMs and their perceived growth in influence.

7.1.1.2 Inform parents & choice

The role of SPMs in informing parents and their school preference through the school admissions process is considered by participants to be a central purpose. Participant 3 stated, for example, "*they are mainly for parents"* while Participant 10 said "*the primary purpose is to inform parents*" and they have a role in "*providing parents with the relevant information that they need to be able to send their child to school that they feel is best for their child*". Participant 6 spoke of their personal experience and the information provided by SPMs with regard to school diversity of approaches:

> "*where we live we could have chosen and got into about five schools. And they all have really different approaches. I personally valued that choice".*

Some comments, however, raise questions about the use of SPMs in informing parents. For example, Participant 10 stated "*I don't think a lot of parents look at them*" and "*not many people would look at league tables*", adding "*I think they are far more likely to read an Ofsted report*".

The suggested reasons for this are discussed in later sections regarding the effectiveness of SPMs.

Participant 11 goes beyond questioning the role of SPMs in informing school choice and challenges the very concept of school choice for parents. They state that parents *"don't really have much choice*

in reality" and that "*the lack of choice is very, very true",* highlighting their perception of the limited availability of schools and the complexity of school admission processes in some areas. Some participants argue that this situation is further worsened for some by the *"lack of appropriate and/or affordable school transport"* (Participant 16) and the current school admissions process. This process allows parents to express a preference, but then school admissions authorities apply criteria that may or may not reflect the parents' choice, something Participant 5 refers to as an "*illusion of choice".*

> *"I think that the reality for most people is that there isn't a choice really anyway. Parents can express a preference for a particular school, if they have a preference. But the reality for most people is there's not much choice".* Participant 5

Participant 11 describes the notion that "*performance measures provide us with information to make informed choices*" as a *"cultural deceit"*.

The concept of the 'illusion' of parental choice is contested, however, by Participant 6. For them, there is strong evidence of real parental choice operating within the school system:

> *"If you look at the percentage choosing their 'catchment school' [i.e. nearest school] as first choice it's around 60%. So, in my view, parental choice is very real".* Participant 6

The principle of parental choice is also closely connected in the discussions to the concept of 'competition' (a notion that I will return to in section 8.4.2.5 below) and 'market forces' for schools operating in quasi-markets. This link is illustrated by Participant 16 in describing a metaphor for school leaders acting as "*entrepreneurs in a competitive market trying to retain [their] market share*". Participant 9 sums up much of the thinking on how this is intended to operate:

> "*the bigger picture is kind of this idea of competition between schools, this kind of market logic kind of competition which drives improvement and it drives improving standards*' through the idea of *parents acting as consumers in choosing between different schools*".

The reference to parents acting as 'consumers' points again to the influence of neoliberal ideology in the operation of SPMs as explored in section 5.1.1.

Within the discussions around the purpose of SPMs informing parental choice, an interesting concept is developed. While the reality of parental choice with regard to the operation of the school admissions process may be contested, there is at least on the surface, parents applying for school places order potential schools by preference. These preferences may be based on a range of available information, including SPMs.

While accepting that SPMs are not ideal, Participant 6 argues that it is still better for parents to have them. They make the case that if SPMs were not provided to parents then they would be more likely to base their preferences on limited available information, including local 'reputation', which might then be less reliable than, admittedly flawed, SPMs.

> *"the more information that parents have - the better they can make their decision".* Participant 6

The argument here is that if SPMs were removed, for example due to their flawed nature, the vacuum created would inevitably be filled with less accurate or reliable information, some of which might result in existing inequalities being exacerbated. A crucial element of this argument is the extent to which SPMs provide useful, relatively accurate information to parents and the scope of the flaws in their production, communication and interpretation. This raises an important question: Are SPMs providing parents with useful, if limited, information about schools or are they so fundamentally flawed that the information that they communicate actually serves to 'mislead' rather than to 'inform'?

7.1.1.3 School improvement

The subject of school improvement in relation to SPMs forms a major theme for the discussions for all focus groups. This is relatively unsurprising given that this is one of the stated aims of SPMs within government documentation (as discussed in section 4.1.2.1).

Participant 17 stated:

> "*for me the general purpose of the school measures … is to improve education quality in general. Because if we don't have such standard measures how we're going to evaluate performance".*

Here they make the case that standardised measures provide a basis upon which schools might evaluate the quality of education provided as part of the process of school improvement.

Three main subthemes emerge from discussions around school improvement and SPMs.

First there is the concept of SPMs "*informing schools"*, to "*highlight the positives of the school*" (Participant 3), to inform schools about "*where [they] are as an institution right now*" (Participant 11), to "*enable [them] to track improvement*" (Participant 12) and to inform schools about priority areas for improvement, for example "*if the school has a massive SEND Gap or PP Gap or whatever"* (Participant 10).

However, some participants highlight the limitations of performance data in explaining the underlying drivers of the differences in SPMs between schools. They express the view that, for school

improvement purposes, information around the processes and practices which are considered to bring about certain outcomes would also be extremely valuable:

> *"the thing with a lot of this data on school differences in student outcomes is that you see differences but the data that we collect doesn't really allow us to understand or explain **why** one school has better outcomes than the other and it doesn't really tell schools what they might do to improve".* Participant 7

This critique argues that outcomes data, when divorced from contextual information about underlying processes, can create a 'black box' effect, hindering schools' ability to genuinely improve. This suggests that, for SPMs to be truly useful, they need to move beyond mere metrics - to provide more meaningful insights into the complex realities of teaching and learning.

The second school improvement subtheme is the notion of SPMs providing 'motivation' for schools to improve through the setting of 'expectations and standards'. For example, SPMs provide a basis upon which to "*fix in place a conception about what good standards are and thereby motivate schools to follow*" (Participant 16). Participant 12 states that, from a historical point of view about introduction of SPMs to the English education system, "*It was about getting to a standard. It was about getting to a standard over time*" with SPMs providing a "*measurement of success*" (Participant 3)

Finally, the potential of SPMs to allow the 'sharing of best practice' was expressed by some participants. One comment stated that this was happening now in their school, "*my school's part of an academy trust and so we are always sharing good practice across different schools*" (Participant 3). Another comment, however, described how this might be developed further:

> "*I know what academies are doing within our own trust, but actually someone might say 'there is this school up in Derbyshire say that they've got no people premium gap. So actually let's go up and have a look at what they're doing' ".* Participant 10

Another participant (15) pointed to the Independent Schools Inspection Framework as an interesting example of how inspection practice can "*highlight successes*" and provide a basis upon which "*schools could learn from the work of others*".

It would be fair to also point to the same limitations referred to above about explaining the underlying drivers of the differences in SPMs between schools with reference to the sharing of best practice. While SPMs might provide a starting point for the identification of potential exemplary practices, further exploratory work would be required for schools to learn from this and how it might influence practices in other settings.

One aspect of the principles underpinning the operation of SPMs is the process of publishing them and in particular the consequences for schools. In particular this relates to the reputation of schools, seen to be directly related to the professional reputation of school leaders, especially headteachers and principles.

> *"Why publish the measures? We'll be having a very different conversation if we were producing the measures but putting them in a sealed brown envelope and sending them to schools. But we're putting them on the internet for all to see".* Participant 16

> *"…it's the consequences of what happens to a school or a college or in institution when those measures are then published - that's the problem".* Participant 15

For some participants, in addition to the production of SPMs it is the impact of their publication which imbues them with significant power to influence schools.

7.1.1.4 Competition and collaboration between schools

Much of the discussion about SPMs is around the principles of competition and how there is a significant tension between a) competition resulting from the publication of SPMs and the operation of school choice and b) the type of collaboration which is considered by many to be valuable in school improvement. This tension raises critical questions about accountability: to whom are schools truly accountable under a competitive model and does this model foster an ethical environment for educational improvement? One participant articulates this tension particularly strongly:

> *"What I would want to start with, I think, in any redesign now is asking 'what behaviours do I want to incentivise?' So, if you think about the past 35 years, what we've had is all about collaboration - and that collaboration creates school improvement across the system. And I don't think anybody really disagrees with that. But the problem is every single measure disincentivises collaboration and incentivises competition".* Participant 12

This highlights how the design of SPMs, as tools of governance, can inadvertently create perverse incentives that undermine desired systemic behaviours. It also emphasised that metrics are performative: they shape the very realities they purport to measure (Kitchin, 2014). A key theme which emerges from the discussions is an almost complete lack of belief that competition is an effective mechanism to drive higher standards in schools, underscoring concerns about the fairness of a system that inherently creates 'winners and losers' rather than fostering collective improvement.

> *"The model was introduced back in the 1980s, when every headteacher was encouraged to run their school as if it was a branch of Tesco's and that is the truth. It*

> *just became a complete acceptance that competition between schools would raise standards. But it's absolute nonsense. It's rooted in competition, which creates winners - and losers. And so, it doesn't work. It doesn't work".* Participant 12

This 'Tesco's' analogy vividly illustrates how data, when deployed for market-driven competition, can decontextualise complex educational realities, reducing schools to comparable market units and in doing so, obscuring the diverse factors which influence their performance. On the other hand, alongside the voices of distrust in competition is an acceptance of the ways in which collaboration between schools can be effective in bringing about positive system change. One participant uses the outcomes of collaboration as part of the London Challenge, widely recognised as being a success story (Ghuman, 2018; Kidson & Norris, 2014; Ogden, 2008):

> *"I believe a big element of the London Challenge and one of the things that made it so successful was data sharing and collaboration. They got groups of schools that had similar characteristics sharing data - and learning from each other".* Participant 13

This points towards an alternative, more ethical use of data – one that prioritises shared learning and collective responsibility over individualistic competition between schools.

7.1.1.5 Enacting government policy

While each of the above essentially represent the aims of SPMs as stated in government guidance documentation (Department for Education, 2022b) of a) informing parents, b) informing inspections and potential interventions and c) informing schools for self-improvement, this subtheme relates to the perceived use of SPMs as a mechanism of implementing wider government policy.

One comment describes the potential benefits for policy makers in deploying policy through the use of SPMs, "*soft policy movement via monitoring and assessment of data. It's an easy way of making change without having to court public opinion*" (Participant 11). This is understood as shaping the decision making behaviour of schools to "*nudge behaviour of schools in a particular direction*" (Participant 11).

Much of the focus of discussion in this area relates to the influence of SPMs on curriculum design. For example, the priority time allocation given to English and maths is discussed. These subjects feature strongly in the secondary headline SPMs, being double weighted in Progress 8, for example. An illustration provided by Participant 16 is in reference to the focus placed on specific parts of the curriculum, where "*a school can't say 'I don't think English or maths are all that important. I'm not going to bother' "*. They elaborate on this by making a distinction between the concepts of the 'autonomy of means' and the 'autonomy of ends':

> "*We're giving them autonomy of means but not ends. We're saying the end is fixed by government. You have to pass these exams. You don't choose what the endpoint of this education is. You have no autonomy for that. But we're going to give you full, or at least a lot of, autonomy over the means you use to get there".* Participant 16

Another example was the introduction of the EBacc (English Baccalaureate, a specific combination of academic qualifications forming the basis of many of the secondary headline SPMs, which brought a greater emphasis on the inclusion of languages and humanities within the curriculum, especially at keystage 4. Participant 12 said that "*forcing all kids to do EBacc*" is not the same as "*schools complying with performance measures which have EBacc as their focus*", emphasising the notion of the soft policy implementation possible through the design and publication of SPMs rather than through any legislative process.

The majority of the discussion around the use of SPMs in the roll out of government policy suggests a degree of negativity towards this practice, with some expressing views that it appears to be shaping decision making behaviour in a deceitful, dishonest way. However, one participant makes an interesting counterpoint to this, while still expressing the manner in which SPMs shape behaviour as an issue.

> "*Really the problem is that what they do is they incentivise behaviour. What I would want to start with, I think, in any redesign now, is what behaviours do I want to incentivise?"* Participant 12

The point that Participant 12 is making is that there may be scope, within the design of SPMs, to consider desirable behaviours which they would want to incentivise and to design SPMs around this specific goal.

Link with regulatory framework and legislation

One participant raises the issue of the direct link between SPMs (and the wider accountability and regulatory framework) and legislation around school organisation, in particular the existing legal duty for the Secretary of State to issue an academy order for 'failing schools' under the Academies Act 2010 as amended by the Education and Adoption Act 2016 — a duty that the *Children's Wellbeing and Schools Bill* aims to replace with a discretionary power, with implementation expected around September 2026 as part of wider reforms.

> *"I think one of the changes in terms of purpose is the degree to which performance measures now feature in regulatory functions and particularly in legislation".* Participant 15

This points to potential obstacles to making any changes to the existing SPMs due to the manner in which they are embedded not only in school practices but also in existing legislation.

7.1.2 Overarching Theme 2: The Fundamental Flaws of Current Performance Measurement

This overarching theme relates primarily to the effectiveness of SPMs and captures participants' general dissatisfaction with the current school performance measures, highlighting their inherent inaccuracies and misleading nature. It points to a system that struggles to genuinely assess school quality due often to an oversimplification of complex educational realities. This theme also highlights the ethical dilemmas and potential for unfairness inherent in such simplifications.

7.1.2.1 The Misleading Nature of Current Measures

Most participants generally perceive current measures as inaccurate and of limited usefulness, which creates a gap between rhetoric and reality. This prompts important questions about whether these measures misrepresent school effectiveness. By emphasising the misleading nature of the current SPMs, participants' views align closely with the findings from the SPM quantitative analysis conducted in section 5.

Unfair or biased

The main issue raised in discussions about the SPMs themselves and their effectiveness is related to questions about their design and potential biases within the metrics. One participant (16) puts this especially strongly:

> *"This is as close as you can get to a completely uncontested finding in social science – and you don't get many of these things. As soon as the datasets emerged on this it was immediately obvious that schools intake was the biggest determinant of school performance".*

Participants make it clear that existing SPMs conflate student demographics (input) with school effectiveness (output). This aligns closely with the findings from Phase 2 (section 6.4.2).

> *"So, a big piece of work that we have been working on for some time is really saying that league tables don't measure school effectiveness they actually pretty much only measure the intake of a school".* Participant 4

This conflation highlights a lack of fairness and the potential for data to be misused and misleading.

Another participant said:

> *"there are years and years' worth of work looking at the kind of relationships within the school contexts and it's always been the case that disadvantaged schools have done less well and it's remained so".* Participant 4

Two potential solutions were suggested for the bias caused by the influence of school context upon SPMs. The first was to *'create a contextualised progress measure'* (Participant 16). It was suggested that this might be published alongside the non-contextualised version. The second was that SPMs should not be published in the current manner (as a set of national statistics which anyone can access with little or no explicit interpretation from the government's website) but instead *"performance measures should be contextualised as part of a school's inspection report"* (Participant 6). These ideas will be examined in more detail below in sections 6.1.2.3, 6.1.4.4.1 and 6.1.4.2.1.

Appearance vs. Reality

Participants across all focus groups express the view that SPMs are a facade, designed to uphold a political narrative rather than accurately assess quality.

> *"I mean, we find the same when we look at performance measures for all schools - literally, there is no statistical difference between the 5th percentile and the 95th percentile. They're all within the same margin of error - with a few outliers at the top and the bottom that might actually be meaningful".* Participant 9

> *"So obviously in some cases identifying that there are areas of deprivation and that that might have a terrible effect on the education of people in that area is not something that there would be a political will to support".* Participant 5

Participants express deep scepticism about the authenticity and usefulness of current measures, with one participant drawing parallels to the famous literary folktale and idiom 'Emperor's New Clothes', underscoring the political dimension of SPM production and interpretation.

> *"Actually, the whole thing feels a little bit like Emperor's New Clothes - we're bothered more about the perception than the reality".* Participant 3

Not focusing on the right things

Many comments from participants highlighted that SPMs tend to focus narrowly on academic outcomes, neglecting other potentially important aspects of schools. Participant 11 stated, *"We have all these data and as somebody who works with schools professionally to try and support help them overcome and find solutions to challenges, the last thing I look at is school performance measures. I want my children to be happy because if they're happy, the learning will come".*

Participant 16 added "*it's about breaking the problem down into a few steps and asking are we providing the right sort of information*". This comment points to potential strategies to address the

issues raised with SPMs, which I will discuss later (section 6.1.4) in considering participants' ideas for moving to a more holistic, supportive and contextualised accountability system.

Out of date

One participant makes the case that, by definition, when SPMs are published they are 'out of date': "*it's just not reliable, up to date information*" (Participant 16). They also pointed to research (Goldstein & Leckie, 2008) which shows that the gap between SPMs "*being published and a child that then joins a school in year 7 the following year sitting their examinations can be 7 or 9 years apart"*.

This comment challenges some of the underpinning assumptions of SPMs, that school performance, as represented by SPMs, is a reliable proxy for school 'quality' and that this is a stable, fixed and objectively measurable feature of schools. On the contrary the work of Goldstein and Leckie (2008) suggest instead that "exam performance now is a poor guide to performance in 6 year's time".

Oversimplified

Participant 16 states, of SPMs: "*the way they are used at the moment is oversimplified and maybe doesn't provide the kind of quality needed as a result*". They argue that SPMs, to be effective, need to "*be simple without being simplistic*". Participant 6 questions the reasoning for always presenting SPMs using school average values "*why do we need to present these things as a mean? Why do we have to come up with this single score?"* They expand on this by suggesting a way forward, which would involve publishing the distribution of student outcomes, rather than distilling it down to a single mean score, an idea I will return to in section 8.5.3 below.

Too complex

Apparently contradicting the challenges expressed above with regard to the 'oversimplified' nature of SPMs, some participants raise concerns about the complex nature of SPMs. Three main points are made in this regard.

The first is that while there are currently only six headline SPMs, the performance tables include many more outcomes for students, specific subjects or groups of subjects and for different student subgroups. For academic year 2023-2024 there are currently 476 different outcome metrics for each secondary school which can be downloaded from the government's Compare school and college performance in England website:

> "*how can we expect parents to wade through all this information and know the intricacies of school performance measures*". Participant 16

The second issue is in understanding the meaning of the grading system for examinations in England. This is in part due to the introduction of numbered grades (9 to 1 numbered grades for GCSE results

replaced the A* to G grading system in 2017) and difficulty that parents and employers have in understanding the meaning of these grades: "*they don't really know what a grade seven looks like*" (Participant 12). It is also due to the complexity of the system of adjusting grades (i.e. 'comparative outcomes') which uses data from a National Reference Test to ensure comparability of grades across years. "*Somebody could get a seven one year but not be as good as somebody who gets a six the following*" (Participant 12).

Finally, it is the degree of complexity within the design of some of the SPMs which is considered to be a barrier for them to be well understood by parents:

> "*Look at Progress 8 – how many parents are going to be able to know how to work that out?*" Participant 10

and by teachers:

> "*…and then you've got to divide it by 10 and you work out the average grade - teachers just don't understand that!*" Participant 10

It is worth pointing out that at this stage that, while above I made reference to an apparent contradiction between concerns about a system which is 'oversimplified' and 'too complex', the two are actually not mutually exclusive. They refer to issues with different aspects of the same system. It could be considered to be consistent to share concerns about the inherent complexity in publishing a large array of school performance metrics, the vast majority of which are based on poorly understood examination outcome grades, some of which are composite measures formed using complicated calculations but then expressed in an oversimplified mean for an entire school cohort.

Zero sum game

One participant makes the point that the way that SPMs are designed and communicated, e.g. in the Progress 8 measure (an in-year measure of relative progress of students) or with 'league tables', it is a 'zero sum game'.

> *"You've got a system set up to create winners and losers - and the most vulnerable kids are always going to be in those losing schools".* Participant 12

This comment appears to question the utility of measures like Progress 8 which, due to their design, do not allow for a meaningful systemwide comparison year-on-year in order to monitor any potential improvements or decline in standards.

7.1.2.2 Over-reliance on Ofsted and External Control

Participants express serious concerns about Ofsted's dominant influence which, they argue, dictates school priorities, often overshadowing governmental directives and student wellbeing. This, it is

suggested, drives compliance rather than genuine improvement, thus raising significant concerns regarding the operation, accountability and ethical oversight of SPMs.

Data Interpretation and Perception

Comments from participants seek to make it clear that School leaders view performance data primarily through the lens of how Ofsted will interpret it.

> *"So, in my role at [organisation name], thinking about stuff like the EBacc measure, people don't say, 'what's the government going to think about our EBacc entry rates?' They say, 'What are Ofsted going to think about it?' every single time and I think that's deeply problematic".* Participant 15

This underscores how the perceived end-use of data shapes both its collection and interpretation. It also raises critical questions about who controls the power to influence school behaviours to such a degree and whether this raises any further questions about the roles and responsibilities of and relationships between, for example, the Department for Education, the Secretary of State for Education, His Majesty's Chief Inspector and Ofsted.

Ofsted vs. Government Authority - An Accountability Imbalance?

Participants express a perception of a lack of direct public accountability for Ofsted's chief inspector. A number of comments from participants demonstrate a concern that Ofsted's influence exceeds that of the government in setting school standards. This raises critical democratic accountability concerns.

> *"It all comes down to the fact that actually Ofsted is an independent inspectorate and the public can't get rid of the chief inspector in the way that they can get rid of the prime minister (and therefore the education secretary). I think there's a fundamental point about democracy - that what we expect schools to do should be set by the democratically elected government, not by the inspectorate".* Participant 15

This dynamic highlights issues of power and control within the SPM data governance frameworks and how the operation of democratic accountability operates differently for Ofsted and the Department for Education (DfE). While the DfE sets the legal and policy framework for what schools must do, Ofsted independently checks schools' compliance with this and defines and then judges 'quality' for inspection purposes through its Education Inspection Framework (EIF). The DfE's policies and other statutory guidance set the fundamental expectations for schools, while Ofsted's EIF provides the detailed methodology and criteria that inspectors use to evaluate schools against these standards. Ofsted's framework is informed by research and professional dialogue, but its final contents are ultimately decided by His Majesty's Chief Inspector (HMCI).

Both Ofsted and the Department for Education are democratically accountable, but in importantly different ways.

The DfE is a ministerial government department, led by the Secretary of State (SoS) for Education which makes it directly accountable to Parliament and the public (Department for Education & Ofsted, 2025). The SoS for Education is a Member of Parliament (MP) and is accountable to the Prime Minister and the Cabinet. They must answer to the House of Commons for the DfE's policies and actions (Pope & Tymms, 2017). The DfE's policies are also subject to debate, questioning and scrutiny by the Education Committee, a cross-party body of MPs. Finally, changes to education policy are often enacted through Acts of Parliament, which must be debated and voted on by elected representatives in both the House of Commons and the House of Lords.

Ofsted is a non-ministerial government department, which gives it a degree of operational independence from the DfE. This is designed to ensure its inspections are impartial. HMCI is appointed by the King on the advice of the Prime Minister and reports directly to Parliament, not the Secretary of State for Education (Ofsted, 2022). HMCI presents an annual report (Ofsted, 2025a) on the state of education and care to Parliament. HMCI and other Ofsted officials are routinely called to give evidence to the Education Committee (House of Commons Education Committee, 2017), which scrutinises their work, policies and effectiveness. Ofsted also regularly conducts public consultations, such as the 'Big Listen' consultation in 2024, to gather feedback from parents, teachers and the wider public on its inspection framework and practices.

The concerns expressed by participants are around the democratic accountability of Ofsted if Parliament and the Select Committee are only able to exert indirect influence. For example, they cannot directly remove the Chief Inspector from their position, instead they are typically limited to applying pressure through select committee recommendations and through parliamentary debate and scrutiny. In more extreme scenarios Parliament could intervene by exercising legislative power by for example, changing Ofsted's remit (Long, 2023; National Audit Office, 2018), changing the HMCI appointment process or by reducing funding to Ofsted (HM Treasury, 2022).

Assumed lack of internal motivation to improve

Comments from participants highlight that the school accountability system operates from a position of distrust towards schools, thus signalling an ethic of surveillance.

> *"And it brings you back to kind of the basic question of why we have school performance measures in the first place and whether is there what level of trust is there that schools are going to innovate because they're wanting to do the best that they can for their community because there's an expectation I think in*

> *lots of approaches to school performance measures and school accountability that left to their own devices, schools would choose the easy option and would let their communities down." Participant 3*
>
> *"If we're working on the principle that, if left to our own devices we would defraud the government and the communities we serve, then we've already lost before we start".* Participant 11

Participants are suggesting here that the systems and policies designed to hold schools accountable are fundamentally built on a foundation of distrust in schools and implicitly in teachers and school leaders. It appears to presume that, without rigorous external control, educators will either fail to perform effectively or intentionally act against the best interests of students. Such underlying principles can, in turn, be used to justify strong external regulation with high-stakes consequences which is focussed on quantifiable metrics, since these are perceived to be the only reliable indicators of a school quality and that schools might seek to 'hide' issues if there was not such a rigid framework.

This underlying distrust can create a vicious cycle and may go some way to explain some of the other issues around effectiveness and the wider effects of SPMs.

> *"There's a real issue around the 'high stakes' nature of accountability and how it's interpreted. So, it's almost you need to get to shifting the system away from that 'high stakes' nature first".* Participant 14

When school leaders feel they're being continually monitored and judged, there may be a tendency to become risk-averse, focusing instead on 'teaching to the test' and ensuring compliance rather than innovating and/or addressing students' individual needs. This could result in a culture of fear, stress and low morale, cultivating an ethos of defensiveness and competition through a punitive accountability system which is seen as threatening, rather than encouraging collaboration and professional development, which ultimately might hinder a school's ability to improve. Despite expressing concerns around this, one participant also points to how easy it might be to address this to shift the narrative and to begin to transform the overall system:

> *"It'd be very easy just to change the language used around it to shift things so that it was more of a system which was supportive rather than punitive".* Participant 3

7.1.2.3 The Contextual Blind Spot of Current Measures

There is much discussion throughout the focus group sessions around how the current measures fail to account for diverse socio-economic and contextual factors and that this leads to an inaccurate portrayal of effectiveness and unfair punishment or benefit based on student intake demographics.

Participants feel that the system largely ignores crucial socio-economic backgrounds, SEND and other external factors. This underscores the limitations of quantitative data and the current metrics and echoes the findings of the quantitative analysis discussed in section 4.2. It also speaks of an implicit call for a more holistic, ethically informed approach to measurement within SPMs.

> *"…the thing about context is so important and, I know you've talked about CVA [Contextual Value Added] and the moves away from that. But the other thing that I heard from a teacher who's a leader in an academy trust recently was that Ofsted had told his trust they were no longer allowed to refer to the context when explaining their children's outcomes. This kind of denial of context is so damaging, at all levels. And it's the same thing that Michael Gove said when he got rid of CVA, it's this 'no excuses' culture, mistaking context for excuses".* Participant 9

What Participant 9 is expressing here is some degree of exasperation with a system which includes a 'contextual blind spot', a deliberate avoidance of metrics which, in an attempt to more fairly and accurately capture school effectiveness, control for student context (e.g. CVA). CVA was discontinued in 2010 by the then Secretary of State for Education, Michael Gove. In the 2010 white paper which heralded these changes, 'The Importance of Teaching', he makes it clear that, in his opinion, the use of contextualised metrics signals the broader acceptance of an academic achievement gap between advantaged and disadvantaged students (Mulholland, 2012). This is a concern recently shared by the current Secretary of State for Education, Rt Hon Bridget Phillipson in a speech to the Centre for Social Justice (GOV.UK., 2025, February 3). In her speech she makes reference to "the soft bigotry of low expectations", a phrase used by Michael Gove in the 2010 white paper, referring to the consequences of schools having lower expectations for students from disadvantaged households compared to their advantaged peers.

In challenging the lack of contextualisation in the existing SPMs, participants are not only calling for greater fairness for those schools with greater proportions of disadvantaged students. They also express issues with a privileged intake advantage, in which schools with affluent intakes can appear successful with less effort. This highlights how SPM data have the potential to both perpetuate and obscure existing inequalities.

> *"So some schools can hide behind what the kids are like. All of those kids that are attending the selective schools would have had a high key stage two, for example. So it's pretty easy for their schools to be seen to be performing well because the students are motivated and whatever else to be, they don't have to put the same sort of effort into teaching".* Participant 14

There is a suggestion in the comment from Participant 14 that the existing SPMs might also be failing to identify some apparently 'high performing' schools where effectiveness falls below expectations.

These schools, as a result of their relatively affluent and academically able intakes and their resulting strong performance metrics, then fall below the radar.

There is a high degree of agreement amongst all participants that there is mismatch between SPMs and actual school effectiveness. They express concerns that SPMs often reflect student demographics rather than the impact of schools, leading to a statistical illusion, which serves to mislead rather than providing valuable information to stakeholders. This closely aligns with the findings of Phase 2 (section 6.4.2).

> *"I think there is research that shows that a very large portion (I think it is around 80%) of the variation in attainment can be attributed to student intake demographics, not the school's actions".* Participant 4

As a response to these concerns, participants call for more sophisticated models that adjust performance measures based on a school's specific context and student demographics. The aim of this is to enhance both fairness and accuracy.

> *"Finding ways to contextualise this kind of measure seems essential to me".* Participant 9

7.1.2.4 Different audiences have different needs and uses of the measures

Some participants' comments make reference to the notion that SPMs have multiple purposes and are intended for a variety of audiences and purposes. The issue here is that, for this to be the case, compromises are inevitable in the type of information provided and how that is communicated.

Participant 16 states that we "*actually need to accept that there are different purposes at play for these measures and there are different levels of understanding of what's there*". They go on to refer to some potential compromises necessary. For example, one purpose is for parents to be able to say that they are "*getting the information that [they] need in order to make a school choice and understand the school*". At the same time it would be hoped that researchers might be able to say: "*I agree that this is a fair representation of the data that we have that it has suitable health warnings that it makes the best use of the data that we have*". In addition, Participant 16 referred to the needs of policymakers so that they can be "*happy this information is steering schools in the right direction*".

In addition to highlighting the different audiences and purposes of the use of SPMs, participants commented on the potential differences in how 'success' might be conceptualised by these diverse users.

“school performance measures give you the cheat sheet: if you do this and you do it successfully your school will be regarded as an excellent school. The problem is that definition of success doesn't tally with everyone - and neither should it”. Participant 11

7.1.3 Overarching Theme 3: The Detrimental Impact on Equity and Practice

This overarching theme captures the wider effects and negative consequences of current SPMs on the broader education system, particularly how they exacerbate social inequalities and distort educational priorities and practices within schools.

7.1.3.1 The Stratifying Effects of 'Choice' and Measures

Participants express concerns that the current system, especially parental choice and data publication, inadvertently exacerbates social stratification and inequality among schools, potentially reinforcing existing disparities. The sharing of these perceptions raises some fundamental questions about the fairness and broader ethical implications of data use in shaping educational opportunities.

Some participants perceive that SPMs used to inform parental choice has a stratifying effect, thereby entrenching and widening social disparities.

> *"Any kind of accountability measure is going to drive segregation". Participant 6*

> *"Who gets to choose? So, who's in the position of being able to choose? We've also got within my own city, which is a very divided city of [city name], if you live in the north, which is the wealthier part of the city, you've got lots of great schools you can choose from. If you live in the south, you've got very limited choice unless you are either a member of a specific church or you pay".* Participant 9

This highlights how data, through their influence on opinion and choice, can contribute to and exacerbate systemic inequities. Participants refer to the effect of SPMs on the wider community with schools, which are considered, through their SPMs, to be 'successful', being able to attract children from families which are less disadvantaged and therefore more likely to be able to choose to live close to such schools.

> *"those schools pull on a neighbourhood and this exacerbates social clustering and inequalities".* Participant 16

Another participant described the impact of SPMs in managing a school 'market', raising questions about how this affects the lowest 'performing' schools.

> *"If you create a market you're going to create winners and losers. The winners get more kids and so they get more money. Now that's all very well and good. But what happens to the losers now".* Participant 12

An important consideration here is that the current Secretary of State for Education, Rt Hon Bridget Phillipson, has made clear her ambition to address the "scourge of inequality" (GOV.UK., 2025, August 13, page 4) and to "break down the barriers to opportunity" (GOV.UK., 2024, December 5, page 35) through education strategy. If the operation of SPMs is acting to exacerbate the very

inequalities that education policy seeks to address this should have a crucial bearing on future education policy thinking.

Some participants comment on a more general point around the cultural entrenchment of the concepts of 'better' and 'worse'. In the language used around the publication and use, for example in local and national media, where the idea of league tables continues to hold sway, of SPMs there is a persisting societal perception of schools being 'ranked'. One participant makes reference to the influence of the culture associated with market choice in shaping these conceptions.

> *"I feel like the culture that we're in of, market choice, already presupposes that there are better and worse choices. And I don't necessarily think that parents are in a position to step outside that system".* Participant 9

This speaks of an underlying assumption of SPMs i.e. that they provide information about which are the 'best' and the 'worst' schools. The 'best' schools might then provide a model for others to follow in school improvement planning, while the worst may be identified as requiring 'intervention'. What participants begin to explore here is the question of whether SPMs are intended to provide information about the 'qualities' or about the 'quality' of a school. 'Qualities' might refer to information about a school's offer, its characteristics, priorities, practices, ethos and specialisms. This might be done in a non-judgemental way, providing information to parents, for example, so that they are able to better assess whether there is a good fit for them and what they are looking for in a school. Alternatively, 'quality' assumes that it is possible and useful, to arrive at an overall judgement of a school and to rank them from best to worst. Whether SPMs are intended to provide information about school 'quality' or school 'qualities', or indeed both of these, is a crucial consideration in their design, communication and interpretation.

One participant, a school leader, speaks of unintended specialisation as a result of parental perceptions of the school, in part based on SPMs alongside wider school information. They talk about their schools becoming overburdened by a specific, high-need student population due to parents' use of available information about the school. Their comment implies a misinterpretation of the data having unintended consequences over time.

> *"We are a relatively small secondary school, so … we have an awful lot of parents of SEND children wanting to send their children here. SEND students do well with us, but it's also because we are a small school so therefore we 'must be more accommodating' and 'it will be a calmer environment' for them with 'less sensory overload'. So I can see it being very easy to get pigeonholed". Participant 10*

While this was a view of only one participant, who was speaking specifically from personal experience, it does illustrate the power of SPMs to not only be shaped by, but also to shape, the student population.

7.1.3.2 Perverse Incentives and Unintended Consequences

Existing performance measures can create counterproductive incentives and negative consequences leading to situations where schools might benefit from negative labels or adopt behaviours which manipulate their data in a high-stakes environment.

Gaming the System

The idea that high stakes incentivise schools to prioritise metrics over broader educational goals is a theme we return to numerous times across all focus groups. Participants share examples of practices which raise questions around the ethical implications of data-driven pressure and how it can compromise educational and professional integrity. Such 'gaming' behaviours are widely recognised and researched (Burgess & Greaves, 2021; Burgess & Thomson, 2023; Figlio & Getzler, 2006).

> *"That plays to the issue of 'is a school inclusive'? Is its ethos about the principles of actually meeting the needs of local children or is it about a gaming strategy where they'll focus on whoever's in the school then, but you don't want the more challenging child in the school in the first place".* Participant 14

The presence of SPMs provides a framework for schools to work within when understanding what is seen to be 'strong performance', 'effective' or to represent 'high quality provision'. SPMs, directly and indirectly, give schools targets to aim for. Participants make reference to a number of ways that SPMs affect the behaviour of schools wanting to maximise their SPMs and therefore their standing within performance tables. These include:

(i) the practice of 'off-rolling', i.e. removing students who would have a detrimental effect on the schools' SPMs from the school roll.

> *"There's always the perverse incentive to lose kids off the school roll".* Participant 4

(ii) 'teaching to the test' i.e. schools choosing to prioritise preparing students for specific assessments rather than providing them with a broad and balanced education.

> *"too many schools just teach to the test".* Participant 15

(iii) 'Narrowing the curriculum', limiting the breadth of the curriculum offer and prioritising qualifications which feature strongly in SPMs, rather than focusing on the interests and aptitudes of the students.

> *"Schools dropping the 'hard' subjects so students couldn't choose them and giving more and more time to the core [English and mathematics] subjects".* Participant 16

These suggest some degree of concern about the 'perverse incentives' which arise from SPMs, rewarding 'gaming' behaviours which have potentially harmful impact upon the quality of education provided to students when judged more holistically rather than through the lens of performance as defined by SPMs.

While participants generally expressed unease with gaming behaviours and their potential consequences, one participant shared that with careful design SPMs and the behaviours which they drive, could be used to define what is considered to be desirable within educational provision.

> *"So that what was being measured was designed in such a way that we actually want people to "game" it, because it ensures some degree of consistency about the quality and content of the education".* Participant 7

This comment hints at potential alternative approaches which exploit the power of data to drive certain behaviours but to design the measures in a way to encourage behaviours which bring about beneficial outcomes rather than the harmful consequences of SPMs which participants have shared.

Disincentive for Transparency

Participants express concerns that schools may avoid acknowledging operational difficulties for fear of negative labels.

> *"You have this perverse idea that nobody wants to fall into the bottom category because actually it screams that you aren't doing our job properly. In actual fact, it probably means that the school is working incredibly hard with poor resources serving a community that will never meet standards that are expected. But rather than ask for help and risk exposure, schools keep silent".* Participant 11

This brings attention to the challenge of achieving genuine openness in a system where data might be used to punish rather than to provide support.

Stifling Local Innovation

Participants speak of standardised external measures preventing schools from adapting to their unique community needs.

> *"I think that part of the problem is if those 'standards' or measures are externally set, you miss out on the innovation that takes place at grassroots level and you miss out on the opportunity for a wide number of people who have contributed to voices and things that should be present within this dialogue feeding into a*

> *mechanism that continually evolves and improves, for the benefit of the school's community".* Participant 11

This reflects a perspective on how centralised, standardised data systems can compromise localised knowledge and agency. This indicates how SPMs can lead to a reduction in effective school autonomy, incentivising work which focuses on improving a narrow set of, typically academic attainment, metrics to the detriment of other areas of work which might better serve the needs of the school community.

'Failure' as a Funding Mechanism

One participant speaks of a scenario where negative judgments, based on inspection informed by SPMs, can ironically lead to increased financial support.

> *"In actual fact, it is potentially in a school's best interest for long long-term development to fall into a 'category'[2] that is undesirable because of what might arise in terms of additional capacity made available through funding because the school has been identified as inadequate".* Participant 11

This reveals a theoretically paradoxical incentive structure, where schools might deliberately seek what otherwise would be considered undesirable judgments in order to access funding required for future development. An alternative interpretation of this scenario might be that this is a more appropriate mechanism of allocating school funding, i.e. the provision of additional resource to schools where it is needed, where improvement is urgently required. It also might be questioned whether schools would, or even could, deliberately 'choose' an inadequate judgement in the way described in this hypothetical scenario.

This does, however, underscore the highly complex relationship between SPMs, school choice, inspection judgments and resource allocation through the National Schools Funding Formula (GOV.UK., 2024, 28 November).

7.1.3.3 Stifling Inclusivity and Diverse Curricula

A significant flaw, expressed by participants, in current measures is their inability to adequately capture or incentivise inclusive practices and the provision of diverse curricula, hindering effective support for all children.

[2] meaning that Ofsted has judged the school's overall effectiveness or specific key areas to be 'inadequate' potentially leading to 'serious weaknesses' or requiring 'special measures' to improve

Traditional Qualification Bias

The overemphasis on traditional, academic qualifications which leads to a neglect of other valuable outcomes is a subject which is returned to many times across the focus groups. This is seen as a consequence of the design of the SPMs themselves. In the 5 examination based headline SPMs we can see this academic qualification emphasis clearly. Three of these SPMs are made up of metrics only related to academic qualifications (Basics, EBacc APS and %EBacc Entry). The other two (Attainment 8 and Progress 8) are very heavily dominated by academic qualifications, with English and mathematics being double counted to deliberately emphasise their value within these measures and each contains a further section which excludes any qualification other than those included in the EBacc (sciences, computer science, history, geography and languages).

> *"There are issues around whether the data that's available captures what a school does to support a child who perhaps isn't going to achieve or for whom the curriculum and qualifications that are offered or judged you is not going to be appropriate. And I think there's far too much focus on the "traditional" qualifications".* Participant 14

This suggests that the narrow scope of collected data can lead to incomplete and, as a result, potentially unfair judgements of school quality. It also indicates that the operation of SPMs serves to marginalise some practices. This might include, for example, less 'traditional' qualifications like some creative subject areas or vocational qualifications which the current headline measures fail to fully capture. Since these subjects may be important to meet some students' individual needs, the sidelining of these subjects results in those students being made invisible, which I turn to next.

Impact on Special Needs/Alternative Provision

It is clear from many comments from participants that they feel that the system of SPMs disadvantages schools catering to diverse learning needs or offering alternative qualifications. This illustrates how SPM policies might adversely impact particularly vulnerable school populations.

> *'You can kind of go back to what happened after the Wolf reforms of 2014 when a whole raft of qualifications ceased to be eligible for performance tables. And you then look at the kind of attainment in special schools and AP [Alternative Provision] schools in particular since then and far less qualifications now available that schools like that can use.' Participant 4*

This extends the point made above with regard to the emphasis on traditional, academic qualifications. In addition to the design of SPMs, which prioritises academic qualifications, there are 4 other mechanisms perceived to squeeze non-academic qualifications from school offers. The first is the SPM related annual publication of approved key stage 4 qualifications for reporting in school and college performance tables. This indicates the size of the qualification, with each one having an

associated point score, as well as discount codes which ensure that a single course of study is credited only once. The second mechanism is the Register of Regulated Qualifications (by Ofqual). Third, there is the list of qualifications which meet the criteria for funding approval. The Department for Education will only fund students to undertake qualifications through a funded offer if the qualifications are on this list. As Participant 4 indicates above, the range of qualifications and their relative point scores etc has changed over time. The perception is that this has reduced the opportunities for schools to offer non-academic (e.g. vocational or creative) qualifications. Finally, there is the wider view that schools are strongly incentivised to build curricula around the courses which lead to those qualifications which meet the criteria and are found on these published lists and so contribute to a school's SPMs. Schools, in principle, have the autonomy to offer courses which do not contribute to their SPMs. However, doing so may risk being judged to be performing poorly, with all of the negative consequences related to this. Alternatively, schools which choose to build their curriculum offer around the qualifications which will contribute to their SPMs might then maximise their SPMs, in the hope that they might be judged to be performing well – with the associated rewards in terms of, for example, reputation, student numbers and funding.

Too much focus on measures

Comments arising from the discussion make specific reference to the effect of SPMs as leading to schools concentrating their actions on work which would bring about improvements, as defined by SPMs, to the detriment of other areas of school work. This is a broader comment than that discussed above in relation to the effect on the curriculum offer. This suggests that SPMs are shaping the wider foci of school priorities. They feel that there are many such areas which, while not featuring strongly within headline SPMs, might still be considered to be very important in order to meet the needs of the whole school community. For example, in this next quote, Participant 10 reflects on the effect of schools prioritising what will improve their SPMs over wider issues which otherwise would be worthy of attention.

> *'You get so fixated on certain measures, because you know that that's going to be an area of focus or that's perhaps not where you need to be or you're below other schools in that particular metric. So, you need to go after that one. And that's sometimes to the detriment of actually what's really going on on the ground' (Participant 10).*

Participant 11 echoes this sentiment, bringing into question the effect of a shift in focus from the needs of students to what is demanded to be seen as 'successful' within SPMs: *"Our prime priority is not our young people but it's as an institution, as a multi-academy, as a senior executive leader, is to demonstrate (quote unquote) 'success' or 'effectiveness' against this myopic data level".*

"That means that they distort what we do, so we begin to promote what is being measured instead of measuring and promoting what we value" (Participant 12). This comment hints at concerns around the loss of autonomy for schools and their leaders, their communities, their stakeholders to decide what is important, what they value and instead to be driven by an external concept of success which is defined by someone else via SPMs. I will now explore this idea in some more detail.

Loss of autonomy and agency

Some participants suggested that the existence of SPMs has diminished autonomy and agency for school leaders and teachers. This closely aligns with the ideas about the influence of neoliberal thinking discussed in section 4.1.1.2. It is felt by some participants that this loss of autonomy has affected schools on several levels. These include effects that operate at:

(i) school system organisation level

> *"'we're going to force you to be an academy' and the school performance measures were used manipulatively to do that".* Participant 12

This refers to the connection between SPMs, inspection and academisation. A school's Ofsted inspection outcome, based in part on various performance measures and judgments around a range of standards, directly determines if it is placed in a 'category of concern', an official status that can lead to forced academisation by the Department for Education. This point underscores the role of SPMs in shaping the organisational landscape for schools,

(ii) **school leadership/headteacher level**

> *"I think the large MATs have infantilised headteachers quite a lot and that's certainly my experience and particularly the MATs that have the very much command and control from the centre way of doing things, a lot of the decision making it has been taking away from headteachers".* Participant 13

The academisation process mentioned above typically leads to schools being forced by the DfE's Regional Director to join a specific sponsor Multi-Academy Trust (MAT), usually one that is recognised as having a strong track record. The MAT's track record will itself be based on improvements that are seen in SPMs and inspection outcomes of their schools.

(iii) **classroom teacher level** (talking about an English teacher they had spoken with)

> *"So his own agency and kind of professional judgment about what the subject of English should look like, frankly, was confusing the proxy with the thing. So, assessment is supposed to be a proxy of learning, not to define what should be taught".* Participant 9

Here, Participant 9, an experienced headteacher and MAT leader, discusses the impact of a focus on assessment and the related drive to improve SPMs through better examination results on what is taught in the classroom.

Measuring Inclusivity

Participants express concerns about the effectiveness of SPMs in capturing schools' complex social realities. This includes representing how schools are tailoring their work to meet individual needs.

> *"I mean if you've got a school that is more inclusive … there are more children with more complex special educational needs, for example. How are their needs met and how is that reflected?"* Participant 14

This raises questions about the distorting effect of SPMs on schools and the need for ways to represent a school's inclusivity and how it is meeting the needs of its student community.

Work pressure and High Stakes

Three participants, in their contributions to different focus groups, make reference in the discussion to Ruth Perry, the primary school headteacher who died from suicide in January 2023 while awaiting the publication of an Ofsted report that downgraded her school from 'Outstanding' to 'Inadequate'. This case has brought a great deal of attention to the question of the consequences of the high-stakes nature of the school accountability system.

While a great deal of focus is on the effect of work pressure arising from a high-stakes accountability system on the mental health of school leaders, the focus group discussion goes beyond this, considering the impact on the effectiveness of teachers and school leaders when working within such a system:

> *"I don't think teachers or school leaders can operate as effectively as they might do when they're under such crippling pressure".* Participant 10

This is extended by another participant in considering the potential effect of the pressure arising from a school accountability system on the students:

> *"when performance measures are used in that punitive way, it gets cascaded into the heads of young people and they feel inadequate, disenfranchised, incapable".* Participant 11

This makes the case that the pressures and associated potential harms of the school accountability system, including SPMs, extend beyond school leadership and permeate the school community, including its students.

7.1.4 Overarching Theme 4: Towards Holistic, Supportive and Contextualised Accountability

This overarching theme primarily concerns alternative approaches to SPMs and encapsulates a vision for accountability that prioritises continuous improvement, genuine support and a nuanced understanding of educational contexts, moving away from high-stakes, punitive measures towards a system that fosters both growth and equity. I will unpack several suggestions which might inform directions for alternative approaches to SPMs made by focus group participants below. I will return to these in section 8.5 when discussing alternative approaches.

7.1.4.1 SPMs are slow to change in order to be commensurable

A minor, but potentially significant point is made by two participants. This is around the perception of the slow pace of change when it comes to SPM policy. This in part speaks of the inertia and path dependency of data infrastructures (Kitchin & Lauriault, 2014). Once a system of measurement is established, particularly one designed for 'commensurability', it becomes incredibly difficult to alter, even when its underlying assumptions are challenged or its effectiveness questioned.

> *'You could look at the DFE documents for the school performance tables for every year. and then you can map the changes which have happened in the last five years and obviously, they're minute, which is rather, in terms of being a little bit pragmatic and realistic as to what might happen, the next five years. The change is so slow and so limited. For example, suddenly there's a "correction for outliers in progress 8". They're all kind of good things and well intentioned - but they're tiny in terms of the bigger picture'. Participant 7*

Participant 9 goes further by seeking to explain the slow pace of policy change in this area, arguing that it is the move to standardisation and the apparent need to be able to compare SPMs for schools year-on-year that is the source of the inertia seen here:

> *'…you could look at all the work around "governing by numbers" and the reforms that have taken place over the last 20, 30 years. It's about the move to make everything standardised in order to make it comparable. In order that comparison can be used to identify the winners and the losers. So, I think those two things go hand in hand - standardisation and comparison. It is assumed that you have to have commensurable measures.'*

This 'governing by numbers' approach may be seen to prioritise quantification and comparability over more nuanced understanding of school effectiveness, potentially leading to unfair assessments and hindering true innovation. This raises crucial ethical questions around the perpetuation of outdated, or even harmful systems, merely for the sake of historical comparison.

This point is more about participants' perception of barriers to change. I now move to consider the suggestions which arise from the focus group discussions which seek to address the issues which participants raise in terms of the effectiveness and effects of SPMs. This includes a reconfiguring of school accountability, an increase in the breadth of what is included in SPMs and a redesigning of SPMs and their communication to stakeholders.

7.1.4.2 Redefining Accountability: From Punitive to Supportive, From Judgment to Growth

This theme focuses on the suggestion of a fundamental shift in the philosophical basis of accountability, moving from a system primarily designed for external judgment and punitive measures towards one that actively supports school development and improvement. This involves redefining the role of oversight bodies and fostering internal learning and more authentic school improvement. This transformation is considered by participants to be essential for establishing accountability mechanisms that are truly fair, ethical and transparent.

Redefining the Inspectorate's Role

Participants imagine a redefinition of the role of Ofsted, shifting from an authoritative, punitive inspectorate to one that acts as a moderator, facilitator and 'safety net' for schools.

> *"If we start to see accountability as a safety net rather than the thing that drives the system we have a very different setup. This is related to the ideas of 'compliance' or 'minimum standards' and we could include safeguarding and whatever else. Then we could say that the purpose of accountability is to catch those schools where things have gone horribly wrong".* Participant 16

This exemplifies a desire for accountability structures that embody fairness and support, rather than simply exercising power and control.

Exploring the idea of a reformed, more supportive approach to accountability, Participant 6 considers that it might also lead to more efficient and less burdensome oversight.

> *"There's a moderation role for Ofsted which wouldn't actually cost an awful lot. This would better suit the time available from Ofsted because they're only going into schools for a day and a half".* Participant 6

These comments make explicit the links between the roles of SPMs and Ofsted. They speak of a desire to move to a new paradigm, where accountability for schools is reconfigured away from the potentially harmful, judgemental system based on a deficit model to one which is supportive and works alongside schools to overcome barriers and challenges. Such arrangements would require a redefinition of the relationship between Ofsted and the school system. This might include Ofsted

taking on a valuable role in verifying school self-evaluation within a culture with increased trust in the schools sector. This may include the use of school data in the form of, potentially reformed SPMs. This forms part of a re-imagining of accountability away from the existing model towards one more rooted in both professional trust and verification.

> *"we've got the data and then inspectors can come in and find out the detail and figure out what's happening. The inspection's role there is to sort of verify that they're using their own data more than having an independent sort of view on it".* Participant 16

This approach also suggests a move towards greater transparency in how schools themselves might present their performance, with external validation acting as an accountability mechanism that both confirms the integrity of the data and supports its interpretation, rather than simply imposing external judgments. As such, it implies a model based on a more ethical partnership between schools and their oversight bodies (e.g. Ofsted and DfE).

Prioritising a more formative data use

In addition to a redefined role for the schools' inspectorate, a move to an accountability system based around actively supporting school improvement would facilitate an emphasis on the primary use of data for internal school improvement and self-reflection, rather than primarily for external judgments.

> *"It's a different use of data - to inform what you might do next and your decision - making rather than just some sort of summative outcomes data".* Participant 3

That is not to say that the current accountability system prevents schools from using SPMs to inform improvement planning. The vast majority of schools will use available data to evaluate school processes and practices, to identify priorities and to assist in the formulation and monitoring of improvement plans. However, at least in part, this current use for many schools will be focused on the headline SPMs and on improvement strategies which are specifically designed to maximise the school's standing on these metrics. A shift to a more supportive system might allow schools the space to take more ownership of the data and how they are interpreted and used to inform strategic planning, built, perhaps, around other priorities and values. These might include national, regional, local, trust-wide or school level priority areas and values.

How schools support their local community, including other schools

The need for some information to be shared, as part of SPMs, around the extent to which schools support their local community is raised in 4 out of the 6 focus group discussions.

Participant 16 summarised this as schools acting as 'good citizens':

> *"we should be taking into account where the school supports teacher training where the school supports other schools, whether the school is a good citizen if you like and in the educational landscape".*

Participants envisage various forms of community engagement, including inter-school collaboration.

Schools working in partnership with community organisations/groups:

Some participants emphasise the role of schools as active partners within their local communities, rather than as institutions whose responsibilities are limited to narrowly defined educational outcomes. From this perspective, schools are understood as community hubs that engage with, support and collaborate with a range of local organisations and groups, including voluntary, cultural and social services. This view challenges the narrow focus of existing SPMs and advocates for measures that recognise and reflect schools' broader social contributions, such as community engagement, partnership working and their role in promoting social cohesion and local wellbeing. As Participant 14 notes, "*it should include factors around how well the school engages with wider community groups as well*".

Schools working to support other schools serving the same community:

Some participants draw attention to the importance of inter-school relationships and collaboration in managing shared responsibilities for pupils within a local education system. Practices such as Fair Access Protocols and Managed Moves require sustained cooperation, trust and relationship-building between schools to function effectively. However, these collaborative practices sit uneasily alongside an accountability system that primarily measures and rewards individual school performance. Participants therefore identify a perceived deficit in the current system's ability to capture and value systemic forms of 'performance' across a local educational ecosystem. By prioritising school-level outcomes within the existing data infrastructure, SPMs may inadvertently disincentivise the collaborative behaviours that are essential for collective improvement and for achieving equitable outcomes for all children within a community. As Participant 10 explains, "*So we will get 'FAPs' [Fair Access Protocols] or Managed Moves from other schools – and for them to work you've got to build those relationships with those other schools.*"

Shift to Collaboration

Participants express a strong desire to move away from individualistic competition towards between-school cooperation. I have previously discussed participants' concerns about the impact of competition on schools and their perceived tensions between competition and collaboration in section 6.1.1.4. I also examined the aims of SPMs as driving school improvement through the

operation of competition between schools operating in a quasi-market in section 4.1.1.4 and the collaborative turn as part of the policies expressed by the Blair government in section 4.1.3.2. Here participants move beyond expressing concerns and speak of what they consider a 'need' which would facilitate a move to collaboration, a system where schools are encouraged to support each other to improve. This also echoes the reference made previously to the London Challenge as a model which could be learnt from, with schools sharing data, resources and expertise to collectively overcome challenges.

> *"I think the key thing would be that we are saying that there needs to be a greater focus on understanding the importance of collaboration and reflecting that through accountability. So that would then have implications for what you measure and how it's measured".* Participant 14

It is worth saying here that the current Secretary of State for Education, Rt Hon Bridget Phillipson MP, has recently made clear her desire to move schools into a more collaborative system. In a speech to the Confederation of School Trusts (GOV.UK., 2024, November 7), she acknowledged that the existing system incentivises competition, something which she declared "ends now". She clearly stated that the key to creating the necessary conditions for a whole range of improvements, based around the themes of 'achieve' and 'thrive', was to replace competition with collaboration. However, what remains less clear from her speech is how the tensions between competition and collaboration will be overcome. The key mechanisms suggested by participants to bring about such a collaborative shift are:

- That the DfE will 'expect' all schools to support each other. This perhaps hints at priorities for how schools might be judged in the future, incorporating a requirement that schools collaborate into inspection frameworks, for example.
- A move to ensure that inspections are used to identify and disseminate excellent practice in schools as a priority. This signals a shift in emphasis for the role of Ofsted.
- Regional Improvement for Standards and Excellence (RISE) teams, coordinated by the most competent teachers and leaders, with those Multi-Academy Trusts identified as being strong playing a key part, will both facilitate support and contribute to the dissemination of excellent practice.

Despite this, the Secretary of State for Education also emphasised the continued centrality of examination outcomes, stating that they will "continue to be the anchor of our education system" (GOV.UK., 2024, November 7). This reading of the speech suggests that, while a shift in rhetoric towards collaboration and a broader conception of educational success is evident, little substantive change to SPMs may be envisaged in the short to medium term. This interpretation is derived from

the internal tensions within her remarks: on the one hand, a turn from competition in favour of collaboration and on the other, a reaffirmation of examination results as the primary organising mechanism of accountability. At the same time, her expressed ambition to move away from the "trap of chasing a narrow shade of standards" can be understood as signalling an expectation that schools pursue a broader set of educational goals, including supporting young people to "achieve" and "thrive". Taken together, these statements point towards a potential broadening of what is valued and recognised within the accountability system, which may align with participants' calls for greater attention to pupil wellbeing and holistic development within future SPMs, as I will discuss in section 6.1.4.3.10.

One mechanism designed to incentivise schools to collaborate might be to change the units of accountability from individual schools to a wider locus of accountability, an idea which I turn to next.

Change of locus of accountability

An alternative approach to SPMs, designed to address the issues around competition is discussed by participants. They consider approaches which would incentivise collaboration. One approach that they debate involves a shift in the locus of accountability from individual schools to groups of schools, particularly around a specific locality. This directly seeks to incentivise collaborative practice by introducing the concept of a shared accountability and a reorientation towards responsibility for the education of young people within the wider local community. Such a shift advocates for a more fair and collective form of accountability that supports more equitable outcomes across a wider geographical area. One participant describes an example from their career which employed such an approach.

> *"And what we set up, basically, was part of a flagship project from the DFE at that time. It was very, very collaborative. Our argument there was that what you need is to publish league tables and targets for localities, not for schools. That's because I'm not only going to be concerned about what happens in my school, I'm going to be held accountable for all the children in the town."* Participant 12

The approach described here has resonances again with that of the London Challenge and its optimistic and collaborative ethos (Kelly, 2011). While the London Challenge stopped short of forming formal shared accountability clusters in terms of their SPMs, it encouraged a sharing of responsibility for education beyond individual schools, providing opportunities to share strategy, planning, best practice and resources and for those involved to share in the apparent success (Stewart, 2011).

Safeguarding

Two participants suggest that safeguarding judgments, which, while not being specifically SPMs, form a central part of school inspection processes, should be separated from other judgements or measures. They advocate for a clear compliance model, with either 'met' or 'not met' as the potential outcomes. This reflects a view that certain areas of school operation are so fundamental to student wellbeing that they demand absolute adherence.

> *"If we think safeguarding is important in keeping children safe then check that schools are doing that. Don't put any additional burdens and workload on schools beyond that. That should be statutory and in legislation related to schools".* Participant 15

Differentiating core operations from academic outcomes by separating the assessment of fundamental operational aspects, like safeguarding, from academic performance was seen as important by some participants to ensure clarity and appropriate focus.

> *"Regardless of how you perform as a school - if you fall foul of safeguarding, then you're no good quintessentially at any level".* Participant 11

The reference to an evaluation of safeguarding here suggests that this comment is considering the wider school accountability system, with the monitoring of safeguarding procedures for schools a primary function of the school inspection process, rather than just SPMs. It does, however, suggest how a range of diverse SPMs could be managed, by categorising them, potentially by role or perceived level of importance.

> *"I think there are some of the non-negotiables, like safeguarding. They ensure that, as a parent if you send your child to school there are minimum standards which we expect to be met".* Participant 11

These views propose the existence of an ethical baseline for all schools. By separating safeguarding, it removes the potential for complex comparative metrics to obscure these non-negotiable minimum standards, with the aim of promoting greater clarity in accountability and ensuring greater equity for all children. It perhaps reflects a specific commitment to treating safeguarding as somehow separate from wider 'performance' evaluation, prioritising children's safety above other performance metrics. These views align with calls from all teaching unions in England for reform of the inspection system following the tragic death by suicide of headteacher Ruth Perry. As discussed in section 6.1.3.3.6, Ruth Perry died by suicide after her school's Ofsted Inspection grading was downgraded over safeguarding concerns (National Education Union, 2024, January 18). It is also in line with proposed improvements to Ofsted's inspection approach (Ofsted, 2025b).

7.1.4.3 Broadening What We Measure: Valuing Holistic School Quality

This theme addresses the need to broaden what is measured beyond narrow academic outcomes to encompass a more comprehensive understanding of school quality, including student and staff wellbeing, character development, life preparedness and the health of the school environment. This is seen to be crucial for ensuring fairness in the evaluation of school effectiveness and in developing ethical data practices that capture the full spectrum of educational value in schools.

Below I explore the most common ideas arising from discussions with participants to better represent the wider work of schools within SPMs.

Beyond Academic Metrics

Participant 14's insights capture much of the discussions around the current predominant focus of SPMs on academic outcomes and potential opportunities to address this:

> *"I guess to some extent there needs to be something that focuses on performance outcomes – I mean academic performance. I do think there is a need for that - and parents are one group that would be interested - and children as well. But I think it needs to be much broader than that. That could be vocational qualifications, but then I think some other form of outcomes. For example - things around inclusion and wellbeing will be critical things".*

This highlights a key tension: while academic outcomes are considered important information for parents, an exclusive focus on this aspect of schools raises some ethical concerns about neglecting other crucial aspects of student development and school provision. The current narrow definition of 'academic performance' does not simply reflect what are considered to be the important data related to schools but also implicitly constrains what is valued and measured. This has the consequence of potentially de-emphasising a more holistic view of education and perpetuating an overly competitive, one-dimensional view of school success. Expanding SPMs to include, for example, more vocational qualifications and metrics around inclusion and wellbeing, might foster a more fair and comprehensive assessment of a school's impact.

A common theme arising from discussions is one which advocates for the inclusion of measures that assess non-cognitive skills, character development and students' preparedness for future life.

> *"You would want students to be able to say that 'the school cares about my development as a person' or 'the school cares about me being confident'".* Participant 16

Such an approach would seek to broaden the concept of 'performance' beyond the dominance we see in the current SPMs of academic attainment, making space for a range of potential areas to 'measure'.

How well the school fits what a parent is looking for

One focus group explored an innovative approach to SPMs, shifting the emphasis from seeking an objective assessment of overall 'school quality' to how well a particular school matches a parent's specific needs and values. This is less about identifying 'the best' school and more about answering the question of which school 'fits best' in terms of, for example, ethos, approach, curriculum, wider offer, size or support. This approach might be seen to enhance SPMs by making them more relevant and actionable across diverse parental preferences.

> *"Schools are so diverse today. So, in the age of academies and MATs and schools having more autonomy, they are doing really different things and provide different offers, like academic specialisms, length of school days, discipline approaches. And I think it's a really positive thing if there is actual choice in the system that parents can choose a school that fits what they're looking for".* Participant 6

This suggestion possibly supports greater fairness by recognising the diverse needs of families and the varied strengths of schools, moving away from a single and potentially biased ranking system. Critically, Participant 6 then extends this further, raising the question of whether SPMs might be presented in such a way that allows a parent to see outcomes for students with a similar profile to their own child.

> *"How well does the school do for kids that come in with a particular level of performance? Because then parents can situate their child? Where is my child? What do they need? And how well is the school doing for my type of child?"* Participant 6

This proposal challenges the reductionist nature of current SPMs. It advocates for a system that seeks to empower individual users by supporting informed decision-making through the provision of relevant, personalised and contextualised information, potentially through the use of available interactive dashboard technology.

Evaluation of leadership

Relating to wider aspects of school accountability, one participant questions the efficacy of external metrics or inspection in assessing teaching quality. Instead, they argue that it may be more appropriate to evaluate leadership and governance, with important consequences for school autonomy.

> *"I don't think we will ever really have enough data to be able to say "this is good teaching" or "this is bad teaching" but I do think that we make a really pretty accurate assessment whether leadership and governance is working. If we can say "we trust these leaders, they're doing a good job and so we can leave them to it" - isn't that what we want?".* Participant 16

This perspective begins by challenging the legitimacy of attempts to accurately assess or quantify teaching quality. In its place, Participant 16 proposes that school accountability should be focused where data can be more reliably gathered and assessed i.e. at the leadership level. This implies a system designed to foster greater responsibility in school leaders and by identifying leaders who can be 'trusted', which, in theory, may allow for increased autonomy in school decision making and innovation to meet local needs.

Prioritising Student Development and Wellbeing

Participants express a desire for a move to a system which recognises and actively measures the crucial indicators of student wellbeing as being integral to school quality.

> *"We want students who feel safe and are thriving in ways which are immediate. It's not just some means to some other end 10 years hence. And the same with teachers, we want schools where teachers feel like they're developing -they feel like they're belonging".* Participant 16

Expanding SPMs to include metrics around wellbeing, would aim to foster a more fair and comprehensive assessment of a school's impact and operation. However, it is important to acknowledge the potential risks and ethical dilemmas associated with collecting and publicly using wellbeing data within a high-stakes accountability system.

> *"And suddenly you're asking them about their wellbeing. And then you're saying, 'right, we're going to put student wellbeing on the school league tables.' But as soon as we start measuring this, we've got a real problem. Imagine you're in a sort of difficult area. There's lots of child poverty. There's lots of kids with traumatic and distressed backgrounds. And just what a tough thing that could turn out to be".* Participant 16

This caution suggests the requirement for effective ethical frameworks in data collection and use, particularly concerning sensitive information.

The evaluation of student wellbeing and development is a common topic of discussion, arising in 5 out of the 6 focus groups in this round signifying a desire for SPMs to reflect a broader understanding of what schools are for and how they might be evaluated. One comment from Participant 16 suggests that, in addition to gathering information about student development, school evaluation might also include consideration of student voice.

> *"I mean, the obvious thing which is missing is that we don't really understand, in any of our measures, the students' experience of schools or their development as people". Participant 16*

Further detail of potential areas of focus are suggested by participants for a more holistic and fair measurement of student experience and development:

> *"Aspirations, character, attitudes, mental health, behaviours Yeah. Softer stuff - which people and parents value".* Participant 16

> *"Shouldn't we be prioritising things like tolerance and diversity, understanding of others, respect - these things are important".* Participant 7

These suggestions move beyond a narrow academic performance, emphasising the responsibility of schools to foster well-rounded individuals and contribute positively to society. One participant queries whether the impact of policy decisions in specific areas (e.g. cutbacks in mental health support services for students) are being adequately monitored.

> *"There's been cutbacks to services. But we don't really have any good data on the impacts of it, in terms of how many kids are there with higher risk or developing mental health issues and so forth".* Participant 4

This highlights an important perceived gap as well as a role for SPMs in monitoring the impact of policy decisions. This would allow policymakers to be held accountable for the consequences of their decisions.

This participant also makes reference to benefits which arose from the Every Child Matters (ECM) policy which was introduced by the government in 2003 (GOV.UK, 2003).

> *"I look back to the many years when there was Every Child Matters and we had the five outcomes (being healthy, stay safe, enjoy and achieve, make a positive contribution, achieve economic wellbeing). I kind of think that was the right approach - and that we would look across all of those different measures. Some of them we don't have any data on at the moment".* Participant 4

ECM may serve as an important historical precedent for a more holistic, ethically grounded approach to evaluating child wellbeing, highlighting a potentially missed opportunity in current data collection.

Some participants also follow up the ideas around student wellbeing and development suggesting practical approaches to gather such data, typically involving some form of student survey. These methods represent concrete steps towards obtaining data directly from the student perspective, contributing to a more comprehensive understanding of school life.

> *"There's a lot of schools who are using surveys that feel quite helpful … around school climate and covering lots of things, everything from motivation, efficacy, belonging, sense of orderliness, trust in teachers... the*

> *PASS survey [Pupil Attitudes to Self and School] which the Jubilee Centre in Birmingham were using around Character Education".* Participant 16

This demonstrates a desire to incorporate student voice and more qualitative approaches to bring new perspectives around student wellbeing and development into performance measures. This might also be considered to be a significant step towards greater democratisation of data collection, moving beyond top-down, quantitative metrics to embrace more nuanced, rich data which better reflect the 'whole school offer.' It promotes a system that listens to and values the experiences of its primary stakeholders – the students.

Staff Retention as a Quality Indicator

One participant suggested that information around staff retention can be considered a key indicator of a school's overall health and of effective leadership.

> *"Some schools are really hanging on to people - and some are just haemorrhaging staff. That needs to be considered in terms of what it says about the school".* Participant 16

The assumption being made here appears to be that staff retention should be a key school performance measure, since high turnover (e.g., reliance on supply staff) negatively impacts student learning. This is not without its risks. First, there are different potential interpretations of staff retention and turnover data.

> *"…you expect some degree of turnover in the school is a good thing - as people are moving on to better jobs - but equally too much is a bad thing".* Participant 4

While 'excessive' turnover might signal underlying issues related to fairness in workload or support, 'insufficient' might, for example, bring into question the opportunities afforded staff for professional development. Such ambiguity around who and what decides how much staff turnover is too much brings into question how such a measure might be consistently and fairly applied.

The publication of staff retention figures also risks schools being categorised as 'greenhouse schools' (growing leaders and exporting them) (Davies, 2009; McGovern, 2014) or, alternatively 'burnout schools' (haemorrhaging staff and experiencing high leadership turnover) (Friedman, 1991). While this may well highlight the variation in school environments, such measures may in reality be highly volatile. Labelling associated with a focus on staff retention may prove problematic for those schools experiencing challenges with recruitment. As with other measures, we would need to question the extent to which such features of a school are within or beyond a school's influence. Staff retention SPM might be considered a proxy for features of a school, including its approach to workload and work-life balance, school culture, pupil behaviour, leadership approach, opportunities for

professional development and career progression. However, it may also be shaped by factors beyond the school's influence, such as pay and the cost of living, housing, the local and regional recruitment market, transport networks, local amenities and the demographics and socio-economic context of the local community.

Publication of staff retention SPMs for those finding recruitment of staff challenging might also be damaging to the school's reputation, further exacerbating the challenge. This exemplifies the issue with SPMs in being a disincentive for transparency, discussed in 6.1.3.2.3.

Other discussions with participants explore ideas related to other aspects of the school workforce, for example staff attitudes, development and support and how these might provide valuable insights into the functioning of schools.

> "*What set of measures is going to shine a light on those if one of our priorities is keeping teachers happy and supporting them and recruiting them and training them".* Participant 16

This suggestion promotes the ethical responsibility of schools to prioritise the wellbeing and professional growth of their workforce.

> *"So are there light touch ways of understanding how teachers feel, what their problems are?"* Participant 16

Gauging attitudes are considered here to provide crucial, otherwise invisible, insights into school climate and working conditions.

Teacher empowerment

In response to the perception that SPMs stifle school leader and teacher autonomy and agency discussed previously in section 6.1.3.3.4, here a participant suggests capturing the degree of teacher autonomy afforded by schools.

> *"the question is how, in an accountability system, there can be some sort of recognition of teacher autonomy, or agency and professionalism".* Participant 14

This comment raises teacher agency as a sign of a healthy, professional environment. It also speaks to the use of SPMs to inform both school leaders to inform improvement strategies and policy makers of the impact of policy.

Schools as Part of a Wider Community and System

A frequent theme of discussions is the role of SPMs in representing the contribution that the schools make, and the degree to which they are accountable, to their wider community.

As part of an accountability system which is ensuring transparency around the use of public funds, one participant suggests including schools' budget management within a far broader range of SPMs. They also acknowledge that this would not be without its difficulties.

> *"You might also look at things like staff recruitment and some kind of budget management as well - although given the sensitivity of such information they are probably a very tricky beast to nail down".* Participant 4

While making financial data more transparent might prove to be a powerful tool for public accountability, providing information around effective financial stewardship, the 'sensitivity' acknowledged by the participant underscores the ethical considerations around data privacy and the potential for misinterpretation or misuse of complex financial information.

It is worth stating at this point that some publicly available information about schools' spending is available from the Financial Benchmarking and Insights Tool (FBIT) for schools in England. This is a relatively new service, launched in autumn 2024 by the Department for Education (DfE) as a replacement for previous tools. However, this tool is specifically designed to inform schools, trusts and local authorities. While it is available to all stakeholders, it is unclear how much use it would be for parents or the wider community to hold schools to account. This is in part due to some significant limitations around the data used there, which draws on historical data, including that from the Consistent Financial Reporting (CFR) return, Academies' Accounts Return (AAR) and the School Workforce Census. Some further work would be required to assess the reliability of the data and comparisons available from the FBIT. In addition, some of the tool's features are only accessible by logging into a secure account. This is only available to schools, trusts and local authorities.

Perhaps reflecting the current national focus on inclusion in schools (The Guardian, 2025; The Independent, 2025), participants suggest that inclusion should also form an element of a wider collection of school information and SPMs. This can be seen as a direct call for SPMs to embed more ethical principles at their very foundation.

> *"There needs to be a clear commitment to equality and to inclusion and inclusive practice".* Participant 14

Actively including measures of inclusivity might represent a deliberate effort to counteract historical biases in data collection that have often rendered marginalised groups invisible or reduced complex social dynamics down to simple metrics. Such SPMs could promote greater transparency around a school's effectiveness in fostering an inclusive environment and equitable practices.

The incorporation of 'inclusion' in SPMs would necessitate greater clarity about the conceptualisation and definition of the term. The term 'inclusion' is highly contested and relatively

poorly defined. No single, universally agreed-upon definition exists. This can lead to confusion and conflicting interpretations among different stakeholder groups. This lack of clarity can result in the term being used for various different, sometimes contradictory, purposes. Having said that, some valuable work exists that would provide a strong foundation to arrive at a clearer, more useful definition of inclusion. This will be examined in greater depth in the discussion chapter below.

Priorities for SPMs

Participants discuss the different priorities that are, or might in the future be, used to shape SPMs. This directly relates to the ethical implications of how data systems are designed to exert influence and to drive policy. Participants consider how SPMs might be made more useful for monitoring broader issues related to education. They also explore how revising the priorities for SPMs might provide a sharper focus on, and mobilise greater effort to address, key concerns. I will next look at some example approaches to this which participants consider.

Linking diverse data together to address key priorities at national, regional or local level

Participant 7 emphasises opportunities to use SPMs to address broader issues in communities:

> *"I think there are broader societal issues, but I think people often criticise the UK because the systems aren't joined up very well. There's obviously all kinds of interrelated services which support children in their communities beyond schools. But all the data is in different places, it's not linked … hopefully in the longer term more of these datasets and resources and so the different bits of public services can be linked up".*

This highlights the potential for SPMs to contribute to a more holistic, fair and, crucially, joined-up approach to public services, grounded in greater data linkage and transparency across sectors.

SPMs based around national priorities

Participants consider that SPMs are often opportunities to both identify and bring focus to specific national priorities:

> *"We often think about this from the point of view of a school – but instead I often think about it in terms of them [SPMs] being kind of about national priorities. If you think of something like the attainment of disadvantaged kids, that needs a national policy - that affects all schools".* Participant 4

This points to the potential of SPMs to drive accountability and mobilise action on national-level priorities, for example, in addressing systemic inequalities.

SPMs based around school priorities

Challenging the idea of SPMs being imposed externally, by central government and based around national priorities, participants consider an alternative approach. This is based around schools taking greater ownership of SPMs and framing them around the school's own priority areas.

One participant imagines a scenario where schools are provided with broad brushstrokes to incorporate as areas of focus and then add detail to decide how data related to that will be collected and discussed.

> *'"These are the ones we want to be focusing on. This is how we're going to measure it. This is how we're going to talk about it" '.* Participant 9

Another participant considers the possibility of even greater freedom for schools – allowing them to decide for themselves the areas of focus for evaluation and how they will be evaluated.

> *"These are our priorities and this is how we want to be judged."* Participant 14

The aim here is to empower schools through greater ownership of SPMs. Such a system would require some form of verification for what are effectively school targets. There is a risk that such a system would otherwise be open to abuse, with schools being free to choose SPMs, for example, around their particular strengths or areas where progress has already been made, rather than on the genuine key priority areas for the school, its students and its community.

This suggests a potential shift towards a more participatory and ethically grounded approach to school accountability, where schools have greater agency in defining what constitutes their success, rather than being solely dictated by external metrics.

What are schools for and how should that inform SPMs?

Further to the many comments outlined above, many of which relate to the types of behaviours which are encouraged through SPM policy, some of the discussions revolve around a more broad question: what are schools for and how might this inform the design and use of SPMs?

> *"…because first of all we try to get to an answer to the central question - which has to be: 'What is a school for?' 'What's the point of going to school?' ".* Participant 10

This is clearly an important question and one which might well be used to shape SPMs. However, the answers to it are neither neutral nor straightforward. They are shaped, for example, by the values and ideologies of those designing SPMs. These values and ideologies, in turn, become embedded within the data systems of SPMs. Since SPMs then contribute significantly to shaping the

concept of success, those values and ideologies influence practice at various levels within the school system.

For Participant 15, the search for a "*shared understanding as a society about what education is for*" is considered a "*completely fruitless thing to talk about*" since such a "*shared understanding is impossible*". The consequence of this, according to this participant, is the need to arrive at a set of standards with which schools would be required to comply and that, in the absence of a consensus on the purpose of schools, "*those standards have got to be set by the democratically elected government*".

This suggests a revised form of accountability, one which ensures that schools comply with a set of non-negotiable expectations, something discussed previously in section 6.1.4.2.6 in reference to safeguarding in schools. It implies a move away from attempting to represent school effectiveness within SPMs. Existing SPMs would be replaced with a set of judgments around minimum expectations, of 'met' or 'not met'. They also seek to clarify who would decide those minimum standards. In their opinion, the power to do this is conferred to those who are democratically elected to positions in government.

7.1.4.4 Contextualising Assessment and Tailoring Communication

Important themes from all focus group discussions focus on how SPMs should be communicated. This includes the imperative for accountability systems to acknowledge and account for the unique contexts of schools and to communicate performance data in a nuanced, accessible and audience-specific manner, recognising that a single, simple and universal approach is insufficient.

Contextualised data to allow fairer comparisons and to address bias

The impact of school context on SPMs is considered a high priority to address. Participants want to see the development of evaluation systems and methods built around and accounting for the unique student populations, community needs and specific circumstances of each school.

> *"I think you'd want them all to be contextual. So even things like inclusivity or staff wellbeing or pupil attitudes, you'd probably think that would vary systematically by the type of intake that the school is serving. So, I think some contextual information for those measures would also be important".* Participant 6

A concept which is returned to numerous times in different discussions is the question of fairness. SPMs which fail to take into account school context are considered to be fundamentally problematic and unfair. The urgent need for SPMs to incorporate the context in which a school operates is therefore seen to be critical to achieve a more fair system. Ignoring context risks perpetuating

significant biases when comparing schools that serve vastly different student populations and operate in extensively dissimilar contexts, as Participant 10 powerfully articulates:

> *"If you ignore the difference between schools, you fail to be able to compare them fairly. That's why we need to have contextualised performance measures – which could be done easily".*

This highlights how current decontextualised SPMs are perceived to fail the fairness test, as they are considered unable to equitably represent the efforts of schools serving diverse populations. From an ethical standpoint, the imposition of a single, unadjusted metric on such varied institutions is problematic. Participant 4 agrees with how easy it might be to begin to address the inherent unfairness in SPMs through contextualisation:

> *'…the measures that we have are reliable in that they "measure what they measure". And I think a lot of the complaints about Progress 8 are that it's unfair to schools in disadvantaged areas - and that's true. To begin to fix that we could easily bring back a contextualised version to sit alongside it. It's straightforward - and that wouldn't cost anything really.'*

This sentiment suggests that while the technical capacity for greater SPM fairness might exist, the political will to implement more nuanced and ethically sound data practices in this area may be lacking. The quote from Participant 4 above alludes to the previous use of contextualised SPMs. It also subtly critiques the transparency of the current system, as the mechanisms for accounting for disadvantage are perceived as being deliberately excluded. Another important point to draw from Participant 4's comment is a suggestion of a pragmatic approach to how policy changes involving the introduction of new SPMs might be implemented in practice. Rather than replacing SPMs, a new, contextualised version could be published alongside the existing SPMs. This is responding, in part to the point made above in section 6.1.4.1. The introduction of contextual SPMs was considered a priority recommendation from all focus groups.

Communication and interpretation of SPMs

Misinterpretation of measures

A recurring area of discussion across many focus groups reflects a perceived issue not so much with the design of the measures, but rather with how they are interpreted and the potentially detrimental consequences that arise from such interpretations.

> *"I think what I would say is I don't think there's anything inherently wrong with any of the measures. They measure what they measure. But they're very often misinterpreted. And really looking at a very narrow slice of the work of the school".* Participant 4

> *"There is nothing necessarily wrong with the measures. They do what they say. It's how they get used which is the problem".* Participant 7

This suggests that participants feel that SPMs are regularly being misinterpreted and as a result misused, raising serious ethical concerns about their real-world impact. This indicates a profound need for greater clarity in what performance measures purport to tell us about a school's 'performance' - including a clear definition of the meaning of the term 'performance' itself:

> *"I think we need clarity about what it is that we are discussing. Because if we talk about school performance and, fundamentally, it's about the students' outcomes - then it's not school performance".* Participant 11

while, crucially, acknowledging the inherent restrictions of what SPMs can tell us:

> *"…and recognising the limitations of performance measures".* Participant 16

It is worth noting here that this is not merely a question of how SPMs might be misinterpreted by users. As suggested by the findings of my quantitative analysis into the structure of SPMs (section 5), not only do SPMs fail to accurately represent school effectiveness, but they are also misleading in the way that the demographics of the school intake primarily shape them.

The way that data is communicated, including data visualisation

Participants raised a wide range of concerns about how school performance data is communicated, many of which relate directly to issues of misinterpretation discussed earlier (see section 7.1.2). Collectively, these discussions suggest that the communication of SPMs is a core component of accountability itself. If accountability is understood as a democratic good, premised on transparency and citizens' rights to access and understand information about publicly funded institutions, then the ways in which data are presented, framed and explained become central rather than peripheral concerns.

The use of rank orders: A recurring concern relates to the use of rank orders and league tables. Participants questioned the ethical and practical implications of representing school performance through explicitly competitive hierarchies. Such formats were seen as framing education as a zero-sum game, with winners and losers, rather than as a collective endeavour oriented towards the welfare of young people. As one participant notes, *"an organisational league table is flawed before it even starts because you're setting the terms of a game"* (Participant 11). This framing, they argue, shifts attention away from professional responsibility and collaboration and towards competition between institutions. From the perspective of accountability as a democratic good, rank ordering risks undermining transparency by concealing structural inequalities between schools and by presenting relative positions as objective truths, rather than as context-dependent representations.

Reduction of performance to a single number: Participants also expressed concern about the reduction of school performance to a single headline figure. This practice was seen as obscuring the complexity of what schools do and limiting meaningful understanding. As Participant 7 reflects, "*it's summarising what a school does into one number… Say it's 0.28. That is what your school does and is. But what does that even mean?*" Such simplification may create an illusion of clarity while offering little insight into educational processes or experiences. In democratic terms, this raises questions about whether citizens are being provided with information that is genuinely intelligible and useful for understanding school quality, or merely data that appears authoritative without being explanatory.

Confidence intervals (statistical significance): The communication of statistical uncertainty was another area of concern, particularly in relation to confidence intervals. Participants highlighted that statistical concepts are often presented without sufficient explanation, increasing the risk of misinterpretation. Participant 10 questions the practical value of such data, asking, *"If you've got a school with a Progress 8 of 0.13 and a massive confidence interval, how can I use that?"* This points to a broader issue within the accountability system: transparency is not achieved simply through disclosure, but through communication that enables users to interpret uncertainty appropriately. Without this, the publication of statistically complex information may undermine, rather than support, informed public accountability.

Colour in data visualisation – indicating some positive practice which may help appropriate interpretation: Despite these critiques, participants also identified examples of more effective data communication practices. In particular, the use of colour in data visualisation was seen as a constructive way to convey relative performance and statistical significance without over-reliance on precise numerical values. As Participant 7 explains, *"what you can see is the group, the green (above average) or the yellow (average)… they're doing some good things to communicate an effect size and statistical significance."* Such visual cues were viewed as enhancing accessibility and supporting more appropriate interpretation. From a democratic accountability perspective, this suggests that design choices can either obstruct or enable citizens' rights to intelligible information.

The language used around SPMs: Several participants also questioned the terminology used to describe SPMs themselves. The phrase "school performance measures" was seen by some as misleading, implying a level of precision or comprehensiveness that the data cannot support. Participant 11 argues that *"even the phrase itself is a bit of a misnomer"*, calling instead for greater clarity in how school performance information is described and framed. This highlights that transparency involves not only the availability of data, but also the language through which it is

presented. Inaccurate or overly broad terminology risks shaping public understanding in ways that exceed what the data can legitimately claim.

Enhancing data accessibility through narrative forms: Finally, some participants advocated for alternative modes of communication that move beyond predominantly quantitative representations. In particular, there was support for narrative-based reporting that synthesises data into accessible, contextualised accounts. Participant 16 suggests that it would be helpful to *"produce a narrative report which summarises the data… maybe just one page per school per year which says, overall, 'here are some strengths', 'here are some weaknesses', 'here are some things that stand out'."* This approach was seen as better aligned with the principles of democratic accountability, as it prioritises interpretability and informed understanding over technical complexity. Notably, this perspective resonates with recent policy developments, including proposals emerging from Ofsted's Big Listen consultation (Ofsted, 2024b) and the Department for Education's consultation on accountability reform (Department for Education—Citizen Space, 2025, 10 September), both of which signal a move towards digital school profiles or report cards designed to offer greater nuance, context and clarity for parents and the wider public.

Whether all data is made public

Linked, to some extent, with the ideas shared above about the different audiences for SPMs, participants ask whether all school performance data should be published publicly or if some data might be meant mainly for specific groups (schools, governors, trustees, policy makers, etc.) to use. These discussions explore a complex balance between transparency and usefulness.

> *"And I think that's where we are about stakeholders and their different uses for the data – and so what needs to be public and what doesn't".* Participant 13

> *"You could still have information made available to parents and you could also have performance measures which were not made public, that are solely to be used by policymakers and the schools themselves".* Participant 6

This approach recognises that not all data serve every purpose or audience equally. It advocates for a more sophisticated model in which different data streams serve distinct, clearly defined purposes.

Tailoring Information for Diverse Stakeholders

Some comments propose the customisation of the presentation and content of performance data to meet the specific needs and understanding of different audiences (e.g., parents, policymakers, inspectors and educators).

> *"Maybe with the metrics we have at the moment and the way that they're published they're trying to do too many different jobs for too many different*

> *people. So we've got the same sort of metrics are being used by schools for school improvement. They're being used by inspectors for inspection. They're being used by policy makers and they're also intending to be used by parents about choosing schools".* Participant 4

The diverse purposes, uses, interpretations and priorities of the different stakeholder groups that use SPMs are considered by participants as an issue requiring attention. Participants make reference to the repurposing of examination results data to inform schools, governors, parents, as well as policy makers and others (e.g. Ofsted). This highlights a critical challenge for greater transparency around SPMs: how can a single set of data be effectively and ethically communicated to diverse data users, each with their own distinct informational needs and levels of data literacy? The publication of one set of SPMs, which are intended to meet the needs of all stakeholder groups, results in perceived compromises that could be avoided if the manner of publication were tailored for each stakeholder group.

Addressing the Incompatibility of Multiple Aims

Participants are eager to recognise that a single accountability system cannot effectively serve multiple diverse and potentially conflicting purposes without risking the distortion of measures and their interpretation.

> *"It almost feels like there isn't going to be a solution that enables us to meet all of those purposes because one of the big problems is trying to have a single system that can meet all these different aims. Almost by definition, if you have all these multiple aims, no one system can do that".* Participant 9

This comment points to the adoption of different, targeted communication strategies to meet the different needs of each stakeholder group.

Schools using data to inform their work

Participants consider a key function of SPMs is to inform schools about their work and to aid with strategic planning. Their main focus was on SPMs providing data to support self-evaluation.

> *"It's how do you know that you're doing the best for the kids that you've got?"* Participant 10

> *"Teachers want to know whether their curriculum is effective".* Participant 16

In addition to school leaders and teachers using data, school governors are also expected to use data in the form of SPMs to "*to be able to hold the school to account and to advise it" (Participant 16)*.

This, perhaps, points to a tension in some of the arguments put forward in the focus group discussions. Participants are overwhelmingly critical of the existing SPMs, their use, their

effectiveness and their distorting effect on many aspects of school operation. However, here participants are suggesting that schools place value on SPMs in providing insights into their effectiveness. I will discuss the possible factors underpinning this tension in section 8.2.2.2.

What information do parents need?

One participant succinctly voices the links between information in the form of SPMs, parental choice and the concept of 'democratic good'

> "*Fundamentally, parents need choice and so parents need information. It's a democratic good*". Participant 16

This emphasises that access to information is essential for parents to participate knowledgeably, enabling them to make informed decisions within the school landscape. However, some points warrant further consideration. First, the comment does not specify what type of information parents need—what data is necessary for effective school choice and whether this remains consistent over time or varies among parents. Second, it assumes that school choice is a fundamental aspect of the school system. However, as discussed in section 4.1.1.4, school choice can reflect the neoliberal principles underlying the school accountability system in England. Indeed, the very idea or existence of school choice can be questioned, as examined in section 6.1.1.2.

What information do policymakers need about the wider school landscape?

Two participants in different discussions noted that, for policymakers, what matters is a 'system-level view' of performance. This view 'does not need to be published at the level of individual pupils or schools' (Participant 9), but through standardisation and comparison, it provides an overview of how effective the schools' system is overall.

> *"Since the government needs a system-level view of how our schools are doing, then an appropriate tool might be a blind kind of 'PISA' test of a sample of pupils in a sample of schools". Participant 9*

> *"just do a test for a random sample of students in a random sample of schools. It's going to be anonymous and we're doing it to understand the system". Participant 16*

These suggestions emphasise using data for systemic understanding and improvement without resorting to individual blame or competition. They propose both a practical and ethical approach to data collection that prioritises collective insight over potentially harmful granular comparisons.

A key lesson from these discussions is that using SPMs by different groups for various purposes is problematic because existing SPMs are not designed to meet all stakeholder needs. The solution is

to redesign SPMs and the broader accountability system with those needs in mind. Two participants support a pragmatic approach that starts with the purposes, designing SPMs accordingly.

> *"If our aim is school improvement, let's start from that aim and work backwards". Participant 9*

> *"I'd sort of try to find the key functions of schooling and what each group wants to find out - and work back to the measures". Participant 16*

This approach promotes a more transparent and ethically focused design process that ensures SPMs are suitable for their intended purpose.

Moderated Self-Evaluation and Triangulation

Promoting the use of moderated self-evaluation as a powerful tool, one participant suggests an approach which combines schools' insider knowledge with external review and the triangulation of evidence from multiple sources to achieve a comprehensive and reliable evaluation.

> *"I really like as a principle that leaders evaluate the school themselves and it's externally validated".* Participant 10

The principle behind this seems to align well with the idea of focusing school accountability on the evaluation of leadership (as discussed in section 6.1.4.3.3) as well as a redefining of the role of the school's inspectorate (section 6.1.4.2.1).

Many diverse measures and randomly choose which to focus on (to address gaming)

This discussion area focuses on a specific proposed approach to SPMs aimed at addressing the ongoing issue of schools 'gaming' the system, such as off-rolling, teaching to the test and narrowing the curriculum. These practices and their relation to SPMs were discussed in section 6.1.3.2.1. The approach involves a wide range of potential SPMs, which are only revealed shortly before publication. While recognising the extensive work of schools, this approach also seeks to prevent schools from unnecessarily focusing on specific, predictable SPMs.

> *"You could flood them with 101 measures and say to schools – we will judge you this year on a selection of these – now 'off you go' ".* Participant 7

> *"… you would have loads of potential measures and then you'd randomly pick a few each year to be the headline measures".* Participant 4

This innovative approach, first suggested by Professor Becky Allen (Allen, 2019, September 29) attempts to restore greater fairness to the system by disrupting the incentive structures that lead to 'gaming'. By introducing unpredictability, it aims to force schools to maintain high standards across a wide range of activities, rather than narrowly optimising known metrics.

7.2 Conclusion to Thematic analysis of focus group discussions on the subject of school performance measures

These findings show that participants view SPMs in England as serving various purposes—accountability, transparency, informing parents, improving schools, fostering competition and implementing policy—while also creating significant tensions and ethical issues. Participants questioned whether SPMs truly support parental choice or school improvement, pointing out that they often influence school behaviour in unintended ways, affecting curriculum, collaboration and adherence to policies.

Participants overwhelmingly viewed current SPMs as misleading, unfair and overly simplistic, reflecting student intake rather than school effectiveness. They emphasised that high-stakes use reinforces disadvantage, narrows educational priorities, undermines inclusivity and limits school autonomy, thereby creating perverse incentives and restricting schools' ability to address diverse community needs.

In response, participants proposed alternatives and enhancements emphasising holistic, ethical and context-sensitive accountability. These included measures that capture student and staff wellbeing, character development, leadership, inclusion, community contribution and workforce health; personalised, contextualised metrics; student and teacher voice; moderated self-evaluation; and linking wider public datasets. Participants also advocated for communication strategies tailored to different audiences, narrative reporting and the use of measures to prevent gaming.

Overall, the findings directly address the sub-questions for this phase:

- ***What do school stakeholders understand to be the purposes of secondary SPMs in England?*** Participants recognise multiple intended roles but highlight tensions between accountability, parental choice and school improvement.
- ***How do stakeholders perceive the effectiveness and wider effects of SPMs in practice?*** Current SPMs are widely seen as ineffective, misleading and generating perverse incentives, with considerable unintended consequences for schools, staff and students.
- ***What alternative approaches and improvements do stakeholders propose for the design, communication and use of SPMs?*** Participants envision a system that is contextualised, holistic, ethically grounded, collaborative rather than competitive and responsive to diverse stakeholder needs.

These findings underscore the call for a fundamental reconceptualisation of SPMs, moving from punitive, simplistic metrics toward nuanced, fair and improvement-focused measures that reflect the complex purposes of schooling and the diverse contexts in which schools operate.

8. Discussion

8.1 Introduction

In this chapter, I discuss the key findings of my research. My aim is to interpret the results presented in the Findings chapters, explore their significance and implications for understanding the effectiveness and effects of school performance measures (SPMs) and suggest directions for alternative approaches to them. I consider this in relation to the existing body of literature reviewed in Chapter 2 and the research aim and research questions as outlined in Chapter 1.

To do so, I draw on Critical Data Studies (CDS) (Iliadis & Russo, 2016) as a theoretical-analytical lens through which to interrogate the power dynamics, assumptions and biases embedded within SPMs. CDS will enable an examination of how SPMs both reflect and reproduce broader relations of power, particularly through their claims to objectivity, their role in legitimising policy interventions and their performative effects on schools and communities.

I also employ the FATE framework (Fairness, Accountability, Transparency and Ethics) (Inuwa-Dutse, 2023; Microsoft Research Lab, 2025) as a guide to consider how SPMs currently operate and how they might be reimagined or redesigned. While CDS helps reveal *why* current measures operate as they do and whose interests they serve, FATE provides a structure for exploring *how* they might be made fairer, more transparent and more accountable to the diverse purposes of education. Together, these frameworks allow for a critical yet constructive evaluation: one that moves beyond critique of the existing system to inform the development of more equitable and educationally meaningful approaches to school performance measurement.

In parallel, I situate this analysis within a critique of neoliberal influence in education policy, recognising that the design and operation of SPMs have evolved within a policy environment characterised by marketisation, managerialism and accountability. This perspective highlights how measures of 'performance' have been aligned with economic rationalities of efficiency and competition, shaping the behaviour of schools, policymakers and the public understanding of educational 'success'. In order to thoroughly interrogate the implications for the SPM policy of these influences, I have developed a Policy Logic theoretical-analytical framework (section 7.2) which is informed by work in the fields of policy logic, theory of change, programme theory, systems thinking, policy design and implementation theory, critical policy analysis and the sociology of education. This provides a framework for considering and, in turn, evaluating the underpinning policy logic for SPMs.

This research aims to examine how well secondary school performance measures in England serve their stated purposes and intended effects and what empirical analysis reveals about how they might be improved. This overarching aim, in turn, informs three research questions:

Research Question 1. What is the socio-political context of the production and publication of secondary school performance measures in England?

Research Question 2. To what extent is the production, publication and operationalisation of secondary school performance measures in England effective in meeting their stated aims, in this context?

Research Question 3. How can the findings of research and advances in technology be utilised to improve the communication and use of school data by stakeholders for school performance measurement purposes?

This discussion begins by analysing the findings from the three research phases which are presented in chapters 4, 5 and 6 through a policy logic lens. This allows me to examine the internal logic of school performance measures, exploring why they exist, how they are justified, what they prioritise and the conditions and assumptions they depend on. I then summarise the main research findings, emphasising how they address the research aims. I situate these findings within the wider theoretical and empirical context, noting areas of convergence and divergence with previous research. I also consider the implications of these findings for policy and theory, while acknowledging the study's limitations. Finally, I outline directions for future research.

8.2 SPM Underpinning Policy Logic

Before critically examining SPMs, I will first consider the policy logic that underpins them using the Policy Logic Framework detailed in section 3.3.3. This involves examining:

- Their **aims and rationales**, i.e. the purposes they are designed to serve.
- the **conceptual and ethical foundations** that justify their use.
- the **values and definitions of performance** embedded within them.
- the **conditions and mechanisms of operation** upon which their effectiveness depends.

To develop each dimension of this policy logic, I will draw on three main sources. The first is the thematic analysis of documentation related to SPMs. This research phases identified themes directly linked to the aims and purposes of SPMs. It also indirectly, through examining the socio-political context of the introduction, development and operation of SPMs, offers insights into their broader

rationales, conceptual foundations, embedded values, definitions and intended mechanisms. Secondly, I will utilise the findings from my thematic analysis of focus group discussions. One of the themes identified, as discussed in section 6.1.1, concerns the purposes, principles and priorities of SPMs. Lastly, I will draw upon an extensive body of existing research literature related, for example, to education policy, to further develop each of the four dimensions of the policy logic.

I will then use the four dimensions of the policy logic to collectively inform an SPM underpinning policy logic framework. The development of this framework establishes a basis for a systematic critique of how effectively SPMs, in practice, achieve their intended aims and/or produce unintended consequences.

8.2.1 Conceptual and Ethical Foundations of SPMs

8.2.1.1 Introduction to the Conceptual and Ethical Foundations of SPMs

School Performance Measures (SPMs) in England are the product of a long policy trajectory that has increasingly framed education through the lenses of measurement, competition and accountability (Leckie & Goldstein, 2017a). They are not neutral tools but expressions of particular ideological and managerial logics that shape how educational quality is understood, judged and acted upon. This section outlines the historical and ideological context that gave rise to SPMs and then considers their conceptual and ethical foundations. Together, these discussions illuminate how SPMs have redefined the purposes of schooling and the conditions of professional practice.

8.2.1.2 Historical and Ideological Context

The thematic analysis of policy documents and related literature reveals the persistent influence of neoliberalism on education policy in England from the late 1980s to the present. Examination of the themes represented in the SPM documents reveals how the influence extends across party lines and underpins the development, implementation and evolution of SPMs, with neoliberal ideas manifest through four recurring themes:

- **Standards**: an emphasis on addressing 'low standards' and raising educational performance;
- **Autonomy and accountability**: schools gaining managerial freedom while being held to measurable outcomes;
- **Economic focus**: privileging outcomes linked to national productivity and individual employability;
- **Quasi-markets**: the use of competition and parental choice to drive improvement.

Neoliberalism functions here as both an ideology and a policy framework, embedding the principles of efficiency, competition and performative accountability into the education system. Its administrative counterpart, New Public Management (NPM) (Wilkins et al., 2019), operationalises these principles through management approaches. NPM imports private-sector methods—target setting, performance indicators and inspection regimes—into state education. Schools are granted operational autonomy ('freedom to manage') but remain constrained by performance frameworks that bind funding and legitimacy to quantifiable results.

Alongside neoliberalism, neo-conservatism played a crucial role, particularly during the Thatcher governments. The New Right fused economic liberalism with social conservatism: market mechanisms were coupled with moral authority and traditional standards. This fusion was institutionalised through the 1988 Education Reform Act, which introduced the National Curriculum, standardised testing and Ofsted. These reforms reinforced centralised control over what counted as 'knowledge' and 'standards,' aligning economic marketisation with moral conservatism.

A further ideological layer was added under New Labour (1997–2010). Rather than reversing neoliberal reforms, New Labour reinterpreted them through a 'Third Way' lens, blending neoliberal efficiency with social democratic rhetoric. Policies such as Education Action Zones (Jones & Bird, 2000), Excellence in Cities (Bradley & Migali, 2012) and Sure Start Centres (Anning & Ball, 2008) framed accountability and data as tools for equity. However, these initiatives operated within the existing marketised framework, managing rather than challenging inequalities. This process of "roll-out neoliberalism" (Peck & Tickell, 2002, p.380) used targeted interventions to contain the social fallout of market reforms.

Similarly, the Pupil Premium policy introduced under the Conservative–Liberal Democrat coalition (2011) echoed social democratic aims of reducing disadvantage but functioned within neoliberal logic. It tied funding to individual pupils, reinforcing market mechanisms and quantifiable accountability. Success was measured by narrowing attainment gaps—again, through SPMs—thereby equating equity with efficiency rather than social justice.

Across these decades, a substantial body of scholarship has argued that SPMs have emerged at the intersection of neoliberalism, New Public Management (NPM), neoconservatism and elements of social democracy, forming what can be described as a hybrid policy regime (Apple, 2006; Ball, 2003b, 2016; Clarke & Newman, 1997; Lingard & Sellar, 2013). Within this literature, neoliberalism is commonly understood to provide the market-oriented logic of competition and choice (Ball, 2016; Lingard & Sellar, 2013), while NPM supplies the managerial technologies through which performance is measured, monitored and compared (Clarke & Newman, 1997; Gunter et al., 2016).

Neoconservatism, in turn, has been shown to add moral authority, reinforcing traditional values and legitimising surveillance and control (Apple, 2006).

Ball (2003b, 2016) conceptualises these kinds of pressures more broadly as a regime of performativity: a set of structural and cultural conditions in which schools and teachers are rendered visible, comparable and governable through data, targets and inspections. Building on this framework, the present research identifies the novel aspect of SPMs that they operate explicitly at the intersection of multiple ideological and managerial strands — neoliberalism, NPM, neoconservatism and social democratic influences — which together constitute the hybrid policy regime that shapes their design and influence. In other words, while Ball (2003b) provides the conceptual lens to understand the mechanisms and effects of performativity, this thesis extends that insight by empirically mapping the combination of policy logics and ideological influences embedded within SPMs, showing how they interact to produce a system that makes schools and teachers both accountable and governable through data.

8.2.1.3 Conceptual Foundations

The conceptual foundations of SPMs lie in a technocratic faith in measurement—the belief that educational quality can be objectively captured through standardised, comparable outcomes. This approach reduces complex learning processes to a narrow range of indicators, offering an illusion of precision and neutrality. It privileges what can be counted and reported, thereby shaping what is valued.

This quantitative rationality aligns closely with the human capital model of education, which frames schooling as an investment in the skills, knowledge and abilities of individuals that increase their productivity and economic value. From this perspective, education is not only a social or personal good but also an economic activity: its returns are measured in terms of improved earnings, employability and contribution to economic growth. The model emphasises the accumulation of cognitive and non-cognitive skills as a form of “capital” yielding both private and societal returns (Becker, 1964; Schultz, 1961). Students are thus framed as future workers and the purpose of education becomes tied to economic productivity and national competitiveness, aligning moral and social goals with economic imperatives.

SPMs also embody the neoliberal coupling of autonomy and accountability. Schools are presented as self-managing organisations free to innovate, yet they are simultaneously constrained by inspection regimes and data-driven oversight. Autonomy thus functions less as professional trust and more as a mechanism of self-governance—schools internalise accountability pressures and regulate themselves according to external metrics.

The introduction of league tables in the early 1990s epitomised this logic. By ranking schools according to test and examination results, the state established a public market of performance. Parents, positioned as rational consumers, were expected to drive improvement through choice, while schools competed for pupils and prestige. This quasi-market framing individualises success and failure, assuming that outcomes depend primarily on effort rather than structural inequality.

Conceptually, SPMs reconfigure education as a technical and managerial problem. Questions of purpose, meaning and social justice are displaced by a focus on targets, outcomes and efficiency. The result is a narrowing of educational aims and a redefinition of 'quality' in strictly performative terms.

8.2.1.4 Ethical Foundations

The ethical implications of SPMs flow directly from their conceptual premises. When performance data define educational value, ethical relationships are subordinated to managerial demands. Teachers and leaders are compelled to prioritise the measure over the learner, often engaging in what Ball (2000) describes as 'fabrications'—strategic representations of success designed to meet external expectations.

There is substantial evidence that performative accountability can generate chronic pressure, fatigue and cynicism rather than authentic professional growth. Qualitative research has shown that cultures of performative accountability channel teachers away from creative and professionally meaningful work toward compliance with targets and inspections, creating what Perryman (2022) describes as a *"discourse of disappointment"* that undermines teachers' wellbeing. Research drawing on Ball's (2003b) work on performativity highlights how systems driven by external metrics can lead to *cynical compliance* and inner conflict among educators, eroding professional trust and autonomy in favour of surface compliance. Moreover, large-scale international data indicate a positive association between accountability pressures and teacher stress, underscoring how performance-oriented systems can contribute to chronic strain rather than sustainable professional development (OECD, 2016).

Moreover, the datafication of accountability intensifies these ethical tensions. With the rise of digital dashboards, predictive analytics and algorithmic classifications, decisions about schools and pupils are increasingly mediated by opaque technical systems. This 'governance by data' (Ozga, 2009) embeds educational judgment within algorithmic logics that may reproduce bias and inequality under a veneer of neutrality. Ethical concerns about transparency, fairness and the depersonalisation of learning are therefore central.

SPMs also raise profound questions about justice and equity in education policy. Although official documentation sometimes gestures toward fairness and the mitigation of disadvantage (section 4.2.3), the practical effects of SPMs often diverge sharply from these stated intentions, with the rhetoric of objectivity and fairness co-existing alongside patterns of interpretation and use that reinforce existing inequalities. As shown through the phase 2 research findings (section 5.6) and the views expressed by participants in the focus groups of phase 3 (section 6.1.2.3), schools serving disadvantaged communities face greater accountability risks, reinforcing cycles of stigma and resource depletion rather than addressing disadvantage. The focus on closing attainment 'gaps' through competition diverts attention from the structural conditions that produce those gaps.

This critique resonates with a substantial body of scholarship showing that performance measures, even when framed as neutral or fairness-orientated, often reproduce inequities because they fail to account for contextual differences in school populations and resource distributions (Hursh, 2007; Lipman, 2011; Whitty & Anders, 2014). As Whitty and Anders (2014) argue, accountability systems based on standardised measures tend to penalise schools in more challenging contexts, embedding inequity under the guise of comparability. Similarly, Lipman (2011) shows that market-like competition in schooling can accentuate social stratification when performance data are publicly interpreted without adequate contextualisation. These analyses align closely with my empirical evidence that the focus on closing attainment gaps through competitive benchmarking can obscure structural determinants of educational inequality, reinforcing cycles of disadvantage rather than enabling substantive equity.

Overall, SPMs can be understood as embedding significant ethical tensions. Their emphasis on measurable performance may, in practice, crowd out professional judgement, care and contextual understanding, contributing to cultures of pressure, compliance and inequity rather than trust, transparency or educational justice.

8.2.1.5 Summary of the Conceptual and Ethical Foundations of SPMs

School Performance Measures in England thus represent a deeply embedded policy regime shaped by intertwined ideological, conceptual and ethical logics. Historically, they emerge from the convergence of neoliberalism, NPM, neoconservatism and social democracy, an alliance that combines market efficiency, moral authority and technocratic control. Conceptually, they are underpinned by assumptions that educational quality can be objectively captured through quantification and by the framing of education primarily in economic terms. Ethically, they produce a system that privileges performance over purpose, undermining professional integrity and reproducing inequality in the name of improvement.

A more balanced accountability framework would need to broaden the definition of educational success, value qualitative as well as quantitative evidence and reassert education as a public good rather than a competitive commodity. These ideas will be further explored in Section 7.6, which considers potential alternatives to the current SPM model.

Having examined the foundational ideas underpinning SPMs, I will now address the aims and rationales that SPMs are intended to serve before exploring their underlying values, assumptions, conditions and mechanisms. This will then provide a strong basis for assessing the effectiveness and effects of SPMs and considering directions for alternative approaches to them.

8.2.2 Aims and Rationales of SPMs

While existing academic literature has examined school performance measures in terms of accountability, performativity and governance (Ball, 2003b, 2016; Ozga, 2009; Power, 1997), relatively little work has sought to empirically unpack the explicitly stated aims of SPMs as articulated in policy documentation, or to examine how these aims are understood and interpreted by actors within the system. Existing studies tend to treat purposes such as parental information, accountability, or school improvement as taken-for-granted policy rationales or as objects of critique, rather than as empirically reconstructable claims. This study addresses that gap by systematically analysing government SPM documentation alongside focus group data, enabling a detailed examination of the primary aims attributed to SPMs and the tensions between their stated purposes and their perceived functions in practice. In doing so, it offers an empirically grounded account of SPMs as multi-purpose policy instruments, extending existing theoretical critiques by anchoring them in policy texts and stakeholder perspectives.

8.2.2.1 Informing Parental Choice and Accountability

A central justification for SPMs is their role in supporting parental choice by providing clear, comparable information about school performance. Within the quasi-market framework established by the 1988 Education Reform Act, informed parental choice is expected to stimulate competition, incentivising schools to improve. This rationale assumes equal access to information and capacity to act on it—assumptions that overlook social inequalities shaping who can exercise choice effectively.

SPMs also underpin inspection and accountability systems, particularly Ofsted's data-led risk assessments and judgments through the Inspection Data Summary Report (IDSR). Quantified indicators serve as both evidence and triggers for further scrutiny, shaping inspection frameworks and increasing the performative pressure on schools. The conflation of data-driven oversight with quality assurance embeds numerical performance at the heart of educational evaluation.

8.2.2.2 Driving Improvement and Intervention

SPMs are also framed as tools for continuous improvement within a 'self-improving school system.' Schools are expected to use performance data formatively to reflect, benchmark and plan. Yet in practice, the same data are used summatively for ranking, sanction and intervention, collapsing the distinction between developmental and punitive uses of data. Since the 1990s, performance metrics have determined which schools are deemed to be 'failing,' triggering interventions such as forced academisation or leadership replacement.

Recent initiatives such as the DfE's Regional Improvement for Standards and Excellence (RISE) teams suggest a shift towards support and capacity-building, yet these remain embedded within the same performance logic. Ethical concerns persist, as evidence shows that SPM outcomes are strongly shaped by school intakes and social context rather than true effectiveness—penalising schools serving disadvantaged communities.

8.2.2.3 Legitimising Policy and Promoting Transparency

SPMs can be understood through Ball's (2003b) theorisation of performativity, in which performance indicators operate as powerful regulatory mechanisms that shape organisational behaviour through comparison, competition and sanction. Building on this framing, my findings suggest that SPMs function not only to influence school practices, but also to legitimise government policy agendas under the rhetoric of "raising standards." The publication of performance data is commonly framed as a means of promoting transparency and democratic accountability, aligning with principles of openness and public scrutiny. However, participants' accounts (chapter 6) and SPM policy documentation thematic analysis (chapter 4) indicate that, in practice, these data more often reinforce surveillance, positioning citizens less as active democratic agents and more as passive consumers of performance information.

At a system level, the findings further show that SPMs contribute to forms of global comparability through data aggregation and international benchmarking, such as PISA. While Ball's work (2003b) foregrounds the effects of performativity on professional identities and practices, this study extends that analysis by highlighting how SPMs also operate symbolically, reinforcing the assumption that educational quality is measurable, comparable and internationally commensurable. Taken together, these dynamics point to a powerful but problematic regime in which performance data serve simultaneously as instruments of governance, mechanisms of policy legitimation and symbolic proof of progress.

8.2.2.4 Summary of Aims and Rationales of SPMs

Taken together, the stated aims of SPMs to inform choice, inspection, school improvement and intervention, promote transparency and democratic accountability and provide a basis for system learning and global comparison reveal an architecture of governance built around the use of school performance data. While framed as neutral tools for empowerment and improvement, SPMs function simultaneously as mechanisms of control, competition and policy legitimation. Their multiple purposes overlap and sometimes conflict, blurring the line between support and surveillance, information and intervention. In practice, SPMs sustain a performative culture in which improvement is demonstrated through data rather than enacted through learning.

The findings in this thesis resonate with a wide range of studies that have conceptualised performance data as central to contemporary modes of educational governance. Ball's (2003b, 2016) analysis of performativity highlights how accountability systems reshape professional practice and identity through constant measurement and comparison, while Clarke and Newman (1997) and Ozga (2009) emphasise the role of data in governing at a distance. Similarly, Lingard et al. (2017) show how performance metrics operate across scales, linking local school practices to national policy priorities and global benchmarking regimes. However, much of this literature has tended to examine these dynamics either at the level of policy discourse or through broad theoretical critique.

This study builds on and extends this work by empirically tracing how multiple, sometimes contradictory, aims are simultaneously articulated within official SPM documentation and interpreted by actors within the system. By foregrounding how informational, developmental, democratic and system-level rationales coexist within the same accountability architecture, the findings highlight the tensions and slippages between stated intentions and operational effects. In doing so, the research contributes new empirical insight into how SPMs function not only as technical instruments but also as political devices that define what counts as success, shape professional behaviour and sustain the governing logic of the educational market. These aims and rationales therefore provide a further basis on which to evaluate whether SPMs genuinely achieve their stated informational, developmental and democratic purposes, or whether they primarily serve to reproduce established modes of performative governance.

8.2.3 Values and Assumptions of SPMs

SPMs are fundamentally designed to quantify the success, effectiveness and quality, i.e. the 'performance', of schools. These data are not neutral. Instead, they are deeply embedded with various specific values and assumptions about the core purpose of education, the role of schools in society and the best ways to achieve systemic improvement. These values and assumptions are

rarely made explicit, yet they fundamentally influence how schools are incentivised to behave in practice. The design and public reporting of these measures implicitly define what matters most in education, what constitutes a 'good' school and who is responsible for student outcomes. In this way, SPMs do not merely reflect educational reality - they actively construct it.

This understanding aligns with critical policy scholarship that has shown how performance measures operate as normative instruments, embedding particular conceptions of value, responsibility and success within ostensibly technical systems of measurement (Ball, 2003b; Lingard, 2011; Ozga, 2009). However, much of this literature has focused either on high-level policy discourse or on professional subjectivities, rather than on systematically unpacking the values embedded within the design and operationalisation of specific SPMs.

In the following sections, I outline the key values and assumptions underpinning SPMs, focusing on (i) how 'success' and 'quality' are defined, (ii) the economic, social and political drivers that inform their design and (iii) the emphasis of SPMs on specific outcomes and groups.

8.2.3.1 Defining 'Success' and 'Quality'

At the heart of the use of SPMs lies the assumption that school effectiveness is measurable, objective and comparable across contexts. The embedded definition of school performance in secondary SPMs is primarily one of measurable academic attainment, particularly in core subjects (English and Maths). This privileges quantification and standardisation over more complex, relational and often unquantifiable dimensions of learning, including creativity, critical thinking, curiosity, emotional wellbeing, resilience and personal, moral, or civic development.

SPMs highlight an assumption that comparability between schools is achievable, desirable and fair, despite the diversity of school contexts.

This finding echoes long-standing critiques of performative accountability, which argue that standardised measures necessarily narrow educational value by privileging what is easily measured over what is educationally meaningful (Ball, 2016; Biesta, 2015). My analysis builds on this work by showing how these assumptions are not only discursively present but are materially embedded in the construction of SPM indicators themselves, thereby shaping what counts as 'quality' in practice.

8.2.3.2 Economic, Social and Political Drivers

As referenced previously (section 4.1.1.4), the market-driven improvement model, underpinned by ideas originating from a neoliberal worldview, further assumes that the publication of performance data will foster competition among schools, driving up standards through quasi-market mechanisms.

I will return to the mechanisms associated with the operation of the schools' quasi-market in section 7.2.4.

This assumption has been widely documented in analyses of education policy reforms in England and other market-oriented systems, where competition, choice and performance data are positioned as key levers for improvement (Le Grand, 2007; Whitty, 2002). The contribution of this study lies in empirically connecting these ideological drivers to both the stated aims of SPMs and the ways they are interpreted and experienced by school actors, highlighting tensions between policy intent and operational effects.

8.2.3.3 Focus on Specific Outcomes and Groups

Finally, SPMs embody particular value judgments about what should be measured and whose outcomes should matter most. The emphasis on English and mathematics within SPMs, for example, signals a belief in the superior value of core academic subjects, assumed to be aligned with economic utility, often at the expense of the arts, humanities and vocational learning. This is an idea which I will revisit when discussing the narrowing of the conceptualisation of 'performance' revealed by the quantitative analysis of SPMs in section 5.2.1.

SPMs appear to demonstrate some commitment to equity through their disaggregation of data for groups such as disadvantaged pupils or those with special educational needs. This reflects a belief that measuring disparities will lead to targeted interventions and fairer outcomes. I will later discuss whether it is reasonable to assume that this is the case, or whether, in practice, the measurement of 'gaps' serves to reinforce deficit framings, hold individuals and institutions accountable for such disparities in outcomes and, as a consequence, ignore the wider systemic inequalities.

This concern reflects a growing body of critical equity scholarship, which shows that gap-based accountability frameworks can individualise and localise responsibility for inequality while obscuring its structural origins (Gillborn, 2010a; Lipman, 2011). My findings extend this work by demonstrating how these equity-orientated measures are simultaneously mobilised as accountability tools, increasing risk exposure for schools serving disadvantaged communities.

8.2.3.4 Summary of Values and Assumptions of SPMs

SPMs are designed to measure and report school 'performance' as indicators of success, effectiveness and quality. They rely on certain key assumptions. These include that educational quality is measurable, comparable and improvable through transparency and competition. SPMs also embody specific values. They prioritise quantifiable achievement, accountability and standardisation, while also representing neoliberal, market-oriented ideas that connect education

to economic productivity. By concentrating on core academic outcomes (principally English and maths), SPMs determine what constitutes valuable learning and who is considered successful, often neglecting contextual factors and broader educational objectives.

While these dynamics have been theorised extensively in the literature on performativity, datafication and market-based reform (Ball, 2003b; Ozga, 2009; Lingard et al., 2017), this study contributes new empirical insight by integrating policy analysis, stakeholder perspectives and quantitative modelling to show how these values operate simultaneously across design, interpretation and effect.

To assess how SPMs operate in practice, I will examine whether they actually measure school 'effectiveness'. I will combine the reflections of school stakeholders in the focus groups with the findings of both the structural equation modelling analysis of the factors underlying SPMs and the cluster analysis of school information data. This will include an exploration of how the values embedded in SPMs shape school behaviours and priorities, whether their use reduces or reinforces inequities and whether SPMs broaden or narrow the concept of the purposes of schools.

8.2.4 Conditions and Mechanisms of the SPM System

This section outlines a framework for understanding the conditions and mechanisms necessary for school performance measures (SPMs) to function as intended. The framework is organised into two parts. The first draws on idealised neoclassical and neoliberal conceptions of how markets operate. The idealised school quasi-market is rooted in neoclassical and neoliberal economic theory, which assumes that market mechanisms can drive efficiency and improvement in public services, based on principles of rational actors, perfect information and competition and equilibrium.

However, policy design and implementation occur within far more complex and constrained realities. The second part of the framework, therefore, redefines these mechanisms in pragmatic, policy-operational terms, describing the real-world conditions necessary for SPMs to operate effectively within an education system, recognising incomplete rationality, imperfect information limited capacity.

8.2.4.1 Idealised Market Assumptions

SPMs are rooted in the idealised logic of quasi-markets, where rational actors, perfect information and competition are assumed to drive improvement.

- **Perfect information:** Idealised market-based accountability assumes the availability and use of faultlessly accurate, complete and transparent data. However, SPMs are often narrow, volatile and open to manipulation, providing incomplete or distorted signals.
- **Rational actors:** Parents are expected to make perfectly rational, informed school choices based on SPMs, while schools respond by improving performance. However, in practice, choices are limited by location, admissions and understanding and schools may engage in strategic rather than educationally meaningful behaviours.
- **Competition and equilibrium:** The model presumes competition will raise standards and efficiency. However, real systems feature fixed capacity, unequal access and complex admissions, meaning competition often benefits those already advantaged.

These idealised assumptions are widely discussed in the literature on quasi-markets and marketisation in education (Le Grand, 2007; Whitty, 2002; Chitty, 1989). The findings of this study build on this work by empirically demonstrating how these assumptions are both embedded in policy instruments and actively experienced by schools and stakeholders, revealing gaps between theoretical expectations and operational realities.

8.2.4.2 Pragmatic Conditions for a Functioning Schools Quasi-Market

This section presents a normative framework for understanding the conditions and mechanisms required for school performance measures (SPMs) to function effectively and ethically. It moves from the idealised assumptions of neoliberal and quasi-market models—where perfect information, rational actors and competitive equilibrium are presumed (Ball, 2003b; Le Grand, 2007)—toward a pragmatic account of the real-world conditions necessary for accountability to operate credibly. The framework distinguishes between operational conditions, which reflect the practical requirements for SPMs to support school improvement and stakeholder engagement and ethical-structural enabling conditions, which are necessary to ensure fairness, legitimacy, trust and system stewardship (Chitty, 2014; Ozga, 2009; Whitty, 2002). This approach recognises that the idealised conditions of a fully rational market are rarely present in education systems, but it identifies the minimal structural and ethical prerequisites that, in my view, are necessary for SPMs to achieve their intended purposes. By making these conditions explicit, the framework provides a benchmark for evaluating the extent to which SPMs in practice support both effective school improvement and equitable, accountable governance.

In reality, SPMs operate within various social, political, economic, structural, geographical and institutional constraints. A more pragmatic framework, derived from the principles of the idealised market assumptions above, therefore focuses on the practical conditions required for accountability

to function credibly. These conditions are divided into two groups. The first groups is made up of operational conditions. These are pragmatic interpretations of the principles of the idealised market assumptions summarised above. The second group of pragmatic conditions represent ethical-structural enabling conditions. These conditions, though often implied, are vital structural prerequisites that serve to ensure that the quasi-market mechanism maintains its moral legitimacy, social sustainability and overall systemic effectiveness. If these are neglected, the long-term viability of the system risks being compromised.

By integrating qualitative insights from school stakeholders with quantitative analyses of SPM data, this thesis extends existing scholarship (Ball, 2003b; Lingard et al., 2017; Ozga, 2009) by mapping the conditions under which quasi-market mechanisms may succeed or fail in practice, highlighting both operational and ethical-structural dimensions that have been underexplored in prior empirical work.

Operational Conditions

- **Information quality and accessibility:** SPMs must be valid, reliable and sufficiently broad. The metrics must also be resistant to gaming by schools and be presented with transparency and accessibility so a wide variety of users can understand them.
- **Parent and other stakeholder use:** SPMs must be understandable, relevant and usable to a range of users within real-world constraints. Beyond understanding the data, parents and other stakeholders must be able to then use the information in SPMs constructively.
- **School responsiveness:** School improvement depends on schools' having agency, capacity and resources with which to act. The incentives created by the SPMs need to be balanced - motivating improvement without causing distortion. Importantly, the system needs to ensure that opportunities for collaboration are preserved and not crowded out by competition, so that schools are able to share best practices.

These operational conditions are broadly consistent with existing critiques of accountability systems in education (Chitty, 2014; Le Grand, 2007). My research contributes further by showing how these conditions interact in practice, with some being met partially or unevenly across different schools and how these gaps influence the functioning of the quasi-market.

Ethical-Structural Enabling Conditions

- **Fairness and equity**: SPMs must have fairness with regard to intake e.g. the measures should not penalise schools that serve disadvantaged student populations. They also need to ensure that choice and accountability processes do not exacerbate issues around social equity, such

as increasing segregation between schools. Ultimately, the system should ensure equality of opportunity, so that all pupils benefit from the improvement driven by accountability.

- **Stability and trust:** Consistency, stability and responsible communication and framing are essential for SPMs to operate as intended since they build stakeholder trust and public confidence. Frequent policy shifts and media distortion risk sensationalism, undermining trust and creating uncertainty for schools and confusion for parents.
- **Accountability for the Unmeasured:** SPMs, by definition, privilege the measurable. For the system to be *ethically* legitimate, it should not be allowed to destroy the non-measurable elements of a holistic education (e.g., creativity, moral development, social cohesion).
- **Accountability for System Stewardship:** The neoliberal model assumes markets self-regulate towards an equilibrium state. However, in school quasi-markets, this rarely happens. Market limitations and failures (for example, the lack of school places in popular areas or the inability of unpopular schools to exit the marketplace) require active 'market shaping' or 'stewardship'. This function is carried out by the state, local authorities and/or Multi-Academy Trusts (MATs). Without this, any competition mechanism risk collapse, undermining the entire rationale of the quasi-market.

While these enabling conditions have been theorised by Ball (2003b) and Ozga (2009), this thesis empirically identifies how ethical-structural gaps, such as inequitable access or limited capacity for stewardship, can distort accountability signals, reinforcing advantages for some schools and undermining the intended systemic effects.

8.2.4.3 Summary of Conditions and Mechanisms of the SPM System

SPMs are designed to drive market mechanisms that encourage improvement through informed parental choice and school competition. This relies on credible information, meaningful engagement, responsive schools, equity and trust. When these are absent, accountability signals become distorted, undermining both fairness and the system's capacity for genuine improvement.

This framework enables me to evaluate whether SPMs in practice satisfy these pragmatic conditions and mechanisms.

In contrast to previous scholarship (Chitty, 2014; Le Grand, 2007; Lingard et al., 2017), these findings provide new evidence about the real-world interplay between operational and ethical-structural conditions, demonstrating that even well-designed SPMs may fail to achieve their intended market-driven or improvement-oriented outcomes when systemic constraints, inequities or misalignments with school capacities exist.

8.2.5 Conclusion to SPM Underpinning Policy Logic

By reflecting on the conceptual and ethical foundations, aims, rationales, values, assumptions, conditions and mechanisms of SPMs, I have now formulated an SPM underpinning policy logic framework. This framework outlines how SPMs in England are firmly rooted in neoliberal ideology and New Public Management practices that frame education as a competitive market and an economic investment. It sets out how they promote autonomy with accountability, data-driven performativity and the quantification of success, risking the narrowing of educational purpose and the obscuring of structural inequalities. SPMs are intended to make school performance visible, comparable and actionable. They are designed to inform parental choice, guide inspection, drive school improvement, trigger intervention and promote transparency. They rest on the assumption that educational quality can be objectively measured and improved through competition, standardisation and accountability. Effective SPMs require meaningful and interpretable data, equitable design and trust in accountability.

This SPM underpinning policy logic framework, therefore, details the underlying principles and various purposes for SPMs, as well as the varied prerequisites required for them to operate effectively and as intended. I will next use this framework as a basis upon which to critically evaluate SPMs against their stated aims.

8.3 SPM Effectiveness

In this section, I address Research Question 2 by evaluating how well the production, publication and operationalisation of secondary school performance measures in England achieve their stated aims. To do this, I start with the SPM underpinning policy logic framework developed earlier, which was informed by my thematic analysis of government SPM documentation. I draw on the findings of the quantitative analysis phase and insights from the focus groups to assess SPMs in this context, using concepts from Critical Data Studies as a lens to examine how SPMs function.

Various issues with SPM effectiveness have been identified through the research phases included in this thesis. These issues were identified via a) examining the implications of the strong influence of neoliberal ideology on SPMs' operation, as revealed through thematic analysis of SPM documentation, b) insights gained by analysing the structure of SPMs using quantitative methods and c) the broad range of issues discussed in stakeholder focus groups. I have organised these into four categories: 1. SPMs lack breadth and do not reflect whole-school effectiveness, 2. questions around SPM fairness and bias, 3. The communication and interpretation of SPMs and 4. the wider implications for the school system. It is worth clarifying that categorising the SPM issues in this way aims to provide a structure to systematically explore all the questions raised about SPMs across the three research phases. However, it is important to recognise that these categories are interconnected and intertwined and do not fall into neat, distinct classifications. Appreciating the complex, entangled nature of the SPM policy landscape will also be crucial when I turn to suggest directions for alternative approaches to SPMs.

8.3.1 SPMs Lack Breadth and Do Not Reflect Whole School Effectiveness

One of the summary assumptions of the effective operation of SPMs stated above is that SPMs constitute ‘meaningful’ data. It is worthwhile unpacking the term ‘meaningful’ before evaluating SPMs based on this idea. Drawing on research in the fields of public administration (Behn, 2003; Hatry, 2006), policy evaluation research (Innes, 1990; Nardo et al., 2008) and psychometrics (Kane, 2006; Messick, 1989), we can consider the term ‘meaningful’ to incorporate three critical concepts. First, for SPMs to be considered meaningful, they would need to include data of sufficient scope and balance (otherwise termed ‘content validity’ or ‘relevance’) for the user. This might, for example, raise the question of whether SPMs cover an appropriate range of desired educational outcomes for a variety of users and are not solely focused on one narrow area. Second, for SPMs to be meaningful, they must be actionable (otherwise called ‘criterion validity’ or ‘utility’). The data included in SPMs would allow them to be used effectively. For example, they would provide parents with information

to help them choose a school for their child. Finally, the data would need to accurately reflect the aspects of schools they purport to measure (i.e., 'construct validity'). I will turn to each of these considerations in the sections below.

8.3.1.1 Scope and balance

An important theme which emerged from focus group discussions about the effectiveness of SPMs was that they are not focusing on the right things. An outcome of the evaluation of SPMs for participants was that, while they will always, by definition, have a limited scope and balance, SPMs are particularly narrow, focusing primarily on academic outcomes rather than broader aspects, such as personal development or individual emotional wellbeing. Due in part to this narrow extent, SPMs are presented in an oversimplified manner. SPMs are considered to fail to align with what all stakeholders value. This observation aligns with prior research showing that performance measures often privilege easily quantifiable outcomes at the expense of broader educational goals, including personal, social and emotional development (Biesta, 2015; Hargreaves, 2003), but extends it by showing how this misalignment is perceived by stakeholders in England.

A key finding of the quantitative analysis of the structure of SPMs was that the 6 headline SPMs represent just 4 underlying factors. These are the constructs of 'Attainment', 'Curriculum', 'Progress' and 'Destinations'. Of these, three headline SPMs (Attainment 8, Basics and EBacc Average Points Score) strongly correspond to 'Attainment'. Correlation analysis also shows a strong association between these three headline SPMs, indicating a substantial narrowing of the concept of 'performance' within the 6 headline SPMs.

It would be worthwhile at this stage to consider the reasons for this apparent redundancy in the factors represented within SPMs and unpack it in context. In other disciplines, such as Psychology, factor analysis is used with redundancy built in by design (Boyle, 1991). A personality questionnaire, for example, might include multiple items intended to measure the same personality aspect. In this case, the redundancy is intentional. It allows, through factor analysis, the detection of a common underlying construct via shared variance as well as the filtering out of noise. The analysis in research phase 2 identifies, for the first time in the English SPM context, that three attainment-focused measures load onto a single factor, revealing empirical redundancy. This finding is novel because it demonstrates that not all headline SPMs provide distinct information, highlighting a narrowing of the conceptualisation of 'performance' that has not been documented in previous research. While a design of SPMs containing such redundancy might be justified in the same terms as the example from Psychology above, i.e. that it is intentional to provide a more robust model and to filter noise, this raises some further questions. Why is only the attainment factor designed in this way and not

the other SPM factors? Progress 8 has often been treated, since its introduction in 2016, as the prominent, key performance metric, emphasised in league tables and policy discourse (Ofsted, 2023). The fact that only 'Attainment' includes such redundancy, rather than 'Progress' or the other SPM factors, means that the other factors are underrepresented or not captured cleanly. This suggests that the rationale for this lies elsewhere. Alternative explanations for including three very highly correlated 'Attainment' SPMs, while statistically acceptable, include the following conceptual, institutional and political rationales:

- **Policy layering** (Daugbjerg & Swinbank, 2016; Choi & Seon, 2021; Nevile et al., 2019). As educational accountability systems evolve, new measures are added, for example, to 'improve fairness', 'raise standards' or 'widen accountability', without removing old ones. This can lead to an accumulation of similar indicators.
- **Symbolic legitimacy** (Elsbach, 1992). The use of multiple measures can create the appearance of rigour. Redundancy, therefore, might be used to enhance political legitimacy and lend greater credibility to the accountability system, while adding little or no new information to SPMs.
- **Multiple audiences** (Baumet al., 2019). A pragmatic response to the competing demands of different stakeholders could be that each group gets a repackaged version of the construct in a form it prefers.
- **Conceptual narrowness** (Dulude & Milley, 2020). A more direct explanation might be that the policy is designed to reflect a narrow, simplistic conceptualisation of 'performance' as equated with attainment.
- **Institutional inertia** (Rosenbaum, 2022). Once SPMs are established, they gain institutional inertia. School systems, new accountability frameworks and data-collection processes are built around them. This can lead to resistance to more radical change, even when new measures are introduced, resulting in a layering of SPMs over time.
- **Strategic communication** (Chaffee, 1991). Redundancy allows flexibility in messaging. For example, policymakers can switch between different indicators without changing the underlying logic. This can help maintain the stability of the narrative.
- **Performative effects** (Ball, 2003b). Redundancy provides a mechanism which reinforces the dominant paradigm. By presenting multiple attainment-based metrics, SPMs continually reproduce the message that "attainment defines performance". In turn, schools internalise this logic, aligning teaching, curriculum and resource allocation toward those same dominant metrics.

These explanations resonate with other analyses of performative accountability (Ball, 2003b; Power, 1997), which highlight how repeated and overlapping metrics reinforce a narrow understanding of success, but my findings provide empirical evidence showing the specific concentration on attainment in England's secondary SPMs and its perceived consequences among stakeholders.

Taken together, these explanations suggest that the redundancy observed among the attainment-based SPMs might not be simply a statistical artefact but rather a reflection of the historical, institutional, conceptual, ideological and political structures underpinning England's school accountability system. Rather than indicating a deliberately balanced design to enhance measurement reliability, this redundancy functions to reproduce and legitimise a narrow conception of school performance, centred on attainment. In this sense, the overlap among attainment measures helps to stabilise a dominant policy narrative, reinforcing rather than broadening the narrow dimensions through which schools are evaluated.

8.3.1.2 Actionability and the Contextual Blind Spot

A key purpose of SPMs, as discussed previously, is to inform parents and guardians when choosing a school for their child. The assumption is that SPMs can provide information to parents to identify which local schools are the 'best', i.e. the most effective. The potential impact of school reputation on school admission numbers and, therefore, school funding is central to the mechanism through which competition between schools is expected to drive school improvement through accountability pressures.

This aligns with research showing that accountability metrics are intended to influence parental choice and school improvement (Chudowsky et al., 2019; Hout & Elliott, 2011), but my findings suggest that in practice parental decisions are constrained by socioeconomic factors and local geography, highlighting the limitations of assuming rational, fully informed choice.

The implication in terms of actionability is that SPMs may still serve their intended purpose of informing parents in their choice of schools. However, those choices may be being based on the socio-economic attributes of the students at the school rather than on which school has the best provision. This has far-reaching consequences for equity of opportunity and addressing inequalities. Furthermore, while SPMs provide information to parents, in practice the degree of associated choice afforded to parents will be limited and will be dependent upon a range of factors. School admissions can be complex arrangements, coordinated through local authorities and consisting of a number of interconnected factors, including school admissions policies and oversubscription criteria, geography and transport availability, limited school capacities (published admission numbers), academic selective criteria or religious affiliations.

SPMs also have a stated purpose of informing school improvement. Schools are expected to use SPMs as part of their self-evaluation processes and related strategic school improvement planning. However, when SPMs are treated as if they accurately reflect school quality, there are important consequences. They continue to shape policy and practice as though they provide a valid representation of school effectiveness. The effects of this cascade through every level of the education system. At the policy level, misleading data distort priorities as policymakers base reforms and resource allocation on misleading information and interventions target the wrong schools and/or issues. SPMs serve to reinforce inequalities and stratification by amplifying existing mechanisms related to parental choice, advantage and housing markets. This illusion of precision and objectivity risks resulting in the entrenchment, through institutionalisation, of flawed metrics and the erosion of trust and confidence in the school accountability system. This finding builds on earlier critiques of accountability systems, including the work of Ball (2003b) and Power (1997), which emphasise how performance metrics can reproduce inequality, but it provides new empirical evidence that quantifies the extent to which school intake determines SPM outcomes in England.

Within schools, accountability pressures, in a system where schools are held accountable for factors beyond their control, drive schools to prioritise their performance within SPMs over genuine school improvement. This can lead to the gaming behaviours discussed in the findings of the focus group discussions in section 6.1.3.2.1. In particular, SPMs are dominated by effects of school intake, incentivising schools to manipulate the make-up of their student body, through unethical admissions processes, exclusions or 'off-rolling' of students likely to have a detrimental effect on a school's SPMs. For students, this translates into a reduced educational experience. Ultimately, when flawed SPMs retain their authority, the system becomes performative rather than transformative, sustaining the illusion of accountability while perpetuating the very inequalities it was designed to address.

These findings are consistent with research on the performativity agenda in education, which documents how high-stakes accountability measures can incentivise gaming and strategic behaviour that prioritises metrics over learning outcomes (Ball, 2003b; Lubienski & Lubienski, 2006), but my study adds empirical evidence demonstrating the magnitude of the effect of student intake on SPM outcomes in the English secondary system.

From a Critical Data Studies (CDS) perspective, SPMs can be seen not simply as tools for assessing school quality, but as powerful systems that shape how education is understood, managed and controlled. While they appear objective and neutral, this apparent fairness hides the political choices behind how SPM data are defined, collected and used. By presenting themselves as technical and

data-driven, SPMs give legitimacy to government reforms and interventions, assisting policymakers in justifying actions that might otherwise be contested. Once in place, these measures become part of the machinery of governance, used to identify 'failing' schools, justify structural reforms and direct resources and oversight.

At the same time, SPMs help to create and sustain an education quasi-market, encouraging schools to compete for position in league tables and shaping how parents, teachers and students think about success. While this has the appearance of promoting autonomy and accountability, in practice it increases centralised control, as standardised data systems make schools ever more visible to inspectors, trusts and government departments. Over time, SPMs have become deeply embedded in how schools operate and how performance is discussed. This makes them hard to change, even when their limitations are clearly presented and widely recognised.

This echoes critiques in the literature on data-driven governance and audit cultures, which emphasise the ways in which supposedly neutral metrics reinforce centralised control and institutionalise accountability narratives (Power, 1997; Strathern, 2000). My research, however, demonstrates these dynamics specifically within the English secondary school context, linking them to concrete mechanisms such as league table positioning and intake manipulation.

The stories that are told through school performance data, whether about improvement, excellence, decline or disadvantage gaps, have political power. They reduce complex educational realities into numbers and rankings that appear to be transparent and trustworthy, providing governments with a means to demonstrate their influence and ambition for even higher school standards. Yet these stories often mask inequalities, rewarding schools that already have advantages while penalising those working in more challenging contexts. In this way, performance measures shape not only what counts as 'success,' but also who gets to define it. Ultimately, the continued existence of SPMs does not indicate that they are effective. Instead, it shows their value as governance and legitimacy tools. They help sustain an illusion of fairness and objectivity, even when the actual system remains both unequal and contested.

These insights align with studies of the political economy of education accountability, which argue that performance measures often reproduce existing inequalities while appearing neutral (Au, 2007; Ball, 2003b), but my findings quantify the specific extent to which SPMs privilege schools with advantaged intakes and disadvantage schools in challenging contexts.

8.3.1.3 Accuracy

The final condition for meaningfulness is accuracy — the extent to which SPMs measure what they purport to measure. Focus group participants, as discussed in section 7.1.2.1, suggested that a significant issue with SPMs is that they are both oversimplified and reductive. This aligns closely with the narrowing of the conceptualisation of 'performance' referred to above (section 7.1.3).

Findings from the quantitative analysis of SPM structure indicate that over 77% of the variance in attainment can be explained by the demographic profile of school intakes, rather than by between-school differences that might indicate differential school effectiveness. In terms of the accuracy of SPMs, this is a significant finding. It suggests that only a relatively small proportion of the variance in SPMs — up to 23% — can be explained by school effectiveness, implying that SPMs are poor proxies for school quality. This supports prior research highlighting the limitations of either value-added or attainment-based metrics, which are heavily influenced by socio-economic context and therefore provide a distorted picture of school performance (Eyles & Machin, 2015; Reardon, 2011).

This finding resonates strongly with Thrupp's (1999) critique of school effectiveness research, which accepts the existence of modest school effects - typically reported in the region of 5–15% in post-adjustment studies - while emphasising the socially embedded and contextually constrained nature of such effects. Thrupp argues that school composition, or 'school mix', should not be treated merely as a control variable, but as a constitutive feature of schooling that fundamentally shapes school processes and outcomes. The present study extends this critique by demonstrating empirically, using factor analysis and structural equation modelling, that pupil characteristics are not simply external influences on school performance measures, but are central to the latent structure of those measures themselves. In this sense, the dominance of context identified by Thrupp sociologically is shown here to be embedded statistically within the measurement of school performance. The somewhat larger residual variance observed in this study - up to 23% - is therefore best understood not as evidence of a larger causal school effect, but as a reflection of the broader definition of school-level variance inherent in latent variable modelling, which captures all remaining between-school differences not explained by pupil composition, rather than isolating narrowly defined institutional effects.

From a measurement perspective, this raises concerns about construct validity: SPMs claim to assess school 'performance' but, in practice, operationalise it primarily through attainment, which is heavily context-dependent. If SPMs primarily reflect contextual advantage rather than institutional performance, their continued use as indicators of effectiveness misrepresents the nature of

educational quality. The result is an accountability framework that measures social composition more accurately than school impact.

The implications of this inaccuracy are significant. When flawed SPMs are assumed to be accurate and objective, the resulting data are used to justify interventions, reforms and public narratives that may be misplaced or misleading. It can lead to interventions that target the wrong schools, reinforcing socio-economic inequalities and eroding public trust in accountability systems. Institutional and school-level consequences include goal displacement, gaming behaviours and resource diversion toward improving metrics rather than learning, while accountability becomes punitive and distorted. For school leaders, the pressure to improve metrics beyond their influence undermines morale, trust and ethics, leading to disillusionment, burnout and the normalisation of gaming behaviours. Students experience reduced educational quality, greater inequity and internalised notions of failure tied to their school's status. More broadly, these dynamics reproduce social stratification, constrain innovation and erode the purposes of education, transforming a system intended to promote improvement and fairness into one that entrenches inequality under the guise of meritocracy. Issues of accuracy are therefore directly linked to questions of fairness and bias (section 7.3.2) and to how SPMs are interpreted and communicated (section 7.3.

These findings build on earlier work on the limitations of accountability systems and meritocratic metrics (Au, 2007; Biesta, 2015; Power, 1997), but provide novel, quantitative evidence and reinforce the argument that SPMs are systematically biased toward socio-economic advantage.

8.3.2 Questioning SPM Fairness and Bias

Next, I will examine school performance measures related to the principles of fairness and bias, assessing them not only by what they value and measure but also by whose success they recognise and reward and whose challenges they obscure. I will consider a range of perspectives on fairness and bias, including design fairness, deeper structural or systemic issues, the mechanics and transparency of measurement, representation and inclusion.

This analysis builds on a substantial literature on educational accountability and fairness, which highlights that performance measures often reflect and reinforce existing social and contextual inequalities (Leckie & Goldstein, 2018b; Whitty, 2002). While prior research has primarily focused on statistical adjustments or equity in reported outcomes, my approach extends this by examining fairness and bias through multiple lenses—including whose success is recognised, whose challenges are obscured and how design, measurement mechanics and transparency interact to shape systemic

effects. This provides a more holistic and nuanced understanding of how SPMs operate in practice beyond more conventional technical or statistical evaluations.

To examine the extent to which these issues of fairness and bias are embedded within measurable school outcomes, multivariable regression models were estimated for each of the six headline school performance measures. These models included both contextual variables (KS2_APS, FSM6, EAL, SEN) and school character variables (selective admission, religious character and single-sex status), allowing for the estimation of unique associations while holding all other factors constant.

Across all outcomes, prior attainment (KS2_APS) emerged as the dominant predictor of attainment-based measures, particularly Attainment 8, EBacc APS and Basics, confirming that headline performance is strongly structured by intake characteristics. Socio-economic disadvantage (FSM6) showed consistent negative associations across all outcomes, with the strongest effects observed for Destinations and Progress 8, reinforcing the presence of a persistent social gradient across the accountability system.

Notably, EAL status showed a consistent positive association across all six outcomes after controlling for prior attainment and deprivation. This effect was strongest for Progress 8 and EBacc-related measures, suggesting that linguistic composition of schools is associated with distinct performance patterns not captured by deprivation alone. SEN status showed small but consistently negative associations, although effect sizes were comparatively modest.

Among school character variables, selective admission policies showed no consistent positive association with attainment once intake characteristics were controlled for, and were negatively associated with Progress 8 and Destinations. Single-sex schools showed small but consistent positive associations across outcomes, while religious character showed minimal and inconsistent effects.

Taken together, these findings indicate that headline SPMs are not neutral indicators of school effectiveness, but are systematically patterned by pupil intake and composition.

8.3.2.1 Equitable design

For SPMs to be effective, they require equitable design, as explained in section 7.2.5. This means creating systems that measure schools fairly, recognising different school contexts, characters and challenges (Leckie & Goldstein, 2018b). Scholars argue that assessment indicators embedded within accountability systems need to be interpreted in light of contextual diversity to uphold fairness and equity (Shepherd & Adams, 2025). The design would aim to ensure that SPMs reflect what schools can genuinely influence and might also value diverse forms of success, while avoiding the reinforcement of social inequalities. There is a body of evidence, including that arising from this

project, which suggests that SPMs fail to accurately and fairly reflect school effectiveness. As such, they perpetuate social inequalities by providing an unfair and biased performance measurement system. The central issue here is the strong influence of school intake demographics on headline SPMs, as explained in section 5.4.2.

In order to control for these demographic influences on SPMs, a mechanism for contextualising them would need to be deployed. SPMs, in their current form, do not include contextualised measures. A statistical measure known as Contextual Value Added (CVA) for secondary schools in England was published for the years covering 2006 to 2010. When the Conservative-Liberal Democrat government took power to replace New Labour in 2010, they abolished CVA from the headline SPMs. The reasons given for this were: a) it was difficult for stakeholders to understand and b) different pupil groups were expected to make different progress (Department for Education, 2010).

Evidently, a measure that aims to incorporate multiple contextual factors, including, for example, gender, ethnicity, SEND status, age, mobility, deprivation, EAL status, etc., requires a much more complex calculation than a raw attainment measure or a simpler value-added measure. This might justify identifying strategies to improve the communication of CVA and support its proper interpretation, rather than removing it altogether. The reason it was scrapped seems more related to what was presented as a moral stance: that it is "morally wrong to have an attainment measure which entrenches low aspirations for children because of their background" (Department for Education, 2010, p. 68). The removal of CVA as an SPM, as discussed in the findings section for phase 3 (section 6.1.2.3), described by the Secretary of State for Education at the time, Michael Gove, as "rejecting the soft bigotry of low expectations" (Department for Education, 2010, November 22), exemplifies the ideological framing embedded within England's accountability system. By obscuring the influence of student background on academic outcomes, this policy reinforces a worldview prioritising individual responsibility, self-reliance and meritocratic achievement. Structural inequalities are presented as challenges to be overcome through improved teaching and leadership, rather than recognised as systemic factors requiring broader intervention (Ball, 2003b; Whitty, 2002). Research on value-added models suggests that removing contextual adjustments risks misrepresenting school effectiveness, particularly for schools serving disadvantaged populations and can exacerbate inequities by privileging schools with more advantaged intakes (Education Policy Institute, 2025; Leckie & Goldstein, 2018b). The simplification of performance measures also aligns with quasi-market principles, making school data more accessible for parents, while simultaneously legitimising policy narratives that emphasise competition and individual achievement over systemic fairness.

It is worth exploring this position in more detail. This deliberate obscuring of the influence of student context on their academic outcomes reflects an ideology. This includes a belief in individual responsibility, self-reliance and the opportunity open to all to forge their own destiny. It reflects a deep belief in meritocracy. Importantly, it frames structural inequalities as challenges to overcome through better teaching and leadership, rather than as meaningful determinants which require wider systemic change. Simpler SPMs were considered preferable because they better met parents' needs when choosing schools (Department for Education, 2010), thereby supporting the principles of the school's quasi-market. It adopts a 'no excuses' discourse, placing the moral and professional responsibility on schools to raise every child's attainment, regardless of background. It also demonstrates distrust in what were perceived as complex, technical, statistical adjustments, which were vulnerable to manipulation. Together, this reflects a neoliberal moral economy in which equality of opportunity is presumed and inequalities of outcome are attributed to individual or institutional failure rather than systemic inequity. The influence of neoliberal ideology is seen to have brought about a shift in moral norms and a redefining of 'fairness' (Brown, 2015; Harvey, 2007). Rather than seeking to make 'fair' judgments about school effectiveness, by controlling the influence of wider systemic inequalities issues which are beyond the influence of schools, 'fairness' involves holding all schools equally accountable, regardless of context. Such simplistic approaches to SPMs introduce significant issues of comparability between different school types, for example, selective or non-selective, rural or urban or coastal, large or small.

For students, this intensifies inequalities, as those in disadvantaged schools face narrowed curricula, teaching to the test and the internalisation of deficit narratives, as shown in the findings of the phase 3 focus groups in sections 6.1.3.2 and 6.1.3.3. Parents are offered simpler but ultimately misleading indicators of school quality (section 6.1.2.1), reinforcing middle-class advantage and deepening social segregation through perceived choice (Gewirtz et al., 1997). Schools serving deprived communities are punished in league tables, facing reputational and resource decline, while being held morally responsible for wider systemic inequalities. At the system level, SPMs validate competition, autonomy and performativity rather than recognising contextual constraints. This produces a landscape that rewards schools already advantaged by their intakes and frames structural inequity as a result of individual or institutional failure.

I will return to the important question of the concept of contextual adjustment fairness in relation to alternative approaches to SPMs in section 7.5.2.

8.3.2.2 Structural and Systemic Bias

As discussed above, SPMs may reproduce existing social, economic and educational inequalities. Place-based inequalities, such as regional disparities in funding, challenges around the recruitment and retention of qualified staff or the socio-economic profile of catchment areas, further shape outcomes in ways that are outside schools' immediate control. The impact of SPMs on school reputation can serve to exacerbate some of these inequalities. Furthermore, cultural and linguistic assumptions embedded within the curriculum and assessment frameworks can also privilege certain groups, reinforcing biases rooted in class, language or ethnicity. As a result, SPMs risk encoding the structural stratification of the education system into ostensibly neutral measures of 'performance.'

This interpretation aligns with extensive research showing that school accountability systems often reflect and reproduce existing social inequalities. Scholars have demonstrated how performance metrics interact with place-based disadvantage, cultural capital and institutional reputation to advantage already privileged schools and communities (Ball, 2003b; Bourdieu, 1986; Reay, 2017). By embedding these structural conditions within ostensibly objective data, SPMs extend earlier critiques of performativity and marketisation, showing how inequalities become normalised and legitimised through measurement itself rather than being addressed as systemic problems (Gillborn, 2010a; Goldstein, 2017b).

These structural patterns are consistent with the regression findings reported in section 6.5, which show that socio-economic composition and prior attainment remain dominant predictors of all headline performance measures, even after controlling for school-level characteristics.

8.3.2.3 Statistical and Procedural Fairness

While transparency and methodological integrity are often presented as important, or even necessary, foundations of fairness in school performance measures (SPMs), they are not, on their own, sufficient to guarantee fair outcomes. Research across education policy, public administration and algorithmic governance highlights that fairness is a socially constructed and value-laden concept, shaped by stakeholders' values, purposes and interpretations rather than determined solely by technical rigor or openness (Espeland & Stevens, 1998; Braun, 2005). In accountability systems, even statistically sound and transparent models can be perceived as unfair if they conflict with stakeholders' values about what schools should be responsible for, how contextual disadvantage should be addressed or how performance information should be used (Kane & Staiger, 2002; Newton, 2016). Moreover, work on transparency in complex evaluative and algorithmic systems cautions that increased disclosure does not necessarily enhance perceived fairness or legitimacy, particularly when methods are difficult to interpret or when power asymmetries shape how

information is received and acted upon (Ananny & Crawford, 2018). From this perspective, fairness in SPMs cannot be reduced to procedural transparency or statistical integrity alone, but must also be understood in relation to the normative assumptions embedded in measures and their alignment with stakeholders' values and educational purposes.

Modelling choices, such as the form of value-added estimation, weighting of variables, or treatment of outliers, might create or conceal bias, particularly where assumptions about linearity or comparability are left untested. Practices such as data suppression for small cohorts or shrinkage adjustments can distort results for specific types of schools, most notably small or specialist providers (DfE, 2016; Goldstein & Rasbash, 1994; REL Mid-Atlantic, 2022). Equally, limited public access to the underlying data, code, or decision rules undermines opportunities for independent scrutiny or challenge (GOV.UK, 2025, October 16). The fairness of an SPM therefore rests not only on its design but on its methodological robustness, technical transparency (Diakopoulos & Koliska, 2017) and its stability across time and cohorts.

Recent policy initiatives in the UK suggest tentative steps towards addressing some of the opacity surrounding data and algorithmic processes used in public sector decision-making, including SPMs. The UK Government's Data Ethics Framework (UK Government, 2025), built on the principles of Fairness, Accountability, Transparency and Ethics (FATE), encourages decision-makers to consider ethical implications throughout the data lifecycle, from collection to communication. Complementing this, the Algorithmic Transparency Recording Standard (ATRS) (Government Digital Service, 2023). provides a structure for documenting how algorithms are designed, implemented and governed, with the aim of making automated or data-driven decisions more open to scrutiny. While these frameworks have yet to be systematically applied to SPMs, their existence signals an emerging recognition that procedural fairness in educational accountability requires not only robust statistical modelling but also transparency about the assumptions, choices and limitations underlying those models.

8.3.2.4 Inclusion, Accessibility and Voice

SPMs can only be fair if they inclusively and accurately represent the diversity of schools and learners. In practice, the experiences of pupils in alternative provision, special schools or small cohorts are often excluded or statistically marginalised, limiting the representational fairness of the system (O'Higgins et al., 2015; Roberson, 2025). Measures may also fail to accommodate the needs of students with disabilities, those for whom English is an additional language or those from culturally diverse backgrounds, introducing subtle biases through differential accessibility (School approaches to the education of EAL students, 2012). Beyond technical inclusion, fairness also relates to whose

voices shape the very definitions of 'performance' and 'success', as expressed by participants in the focus groups of phase 3 (section 6.1.2.4). When using the research phase 3 findings to explore alternative approaches to SPMs, I will revisit how the legitimacy of SPMs might rely on whether educators, parents and communities are genuinely involved in shaping and evaluating them, ensuring that measures reflect a broader and more just conception of educational value.

8.3.3 Communication and Interpretation of SPMs

The effectiveness of SPMs is limited by a number of conceptual and practical challenges in terms of both communication and interpretation. The assumption that data are inherently interpretable and that they are communicated in a way that supports appropriate interpretation by different stakeholders, is often misplaced, as discussed in reference to the complexity of SPMs in section 6.1.2.1. In practice, SPMs, including current metrics such as Progress 8 and Attainment 8, are widely perceived as complex, confusing and poorly understood, even among educational professionals. Focus group discussions and quantitative analysis using SEM align with existing literature (Education Policy Institute, 2025; Leckie & Goldstein, 2019b; OECD, 2008) in highlighting that SPMs often fail to convey a clear or accurate picture of school 'performance,' partly because they obscure the extent to which outcomes reflect school intake characteristics rather than the quality of provision.

Moreover, SPMs suggest the promise of meaningful projections of school outcomes for students in 5 or even 7 years' time, from the time of choosing a school to leaving it, thus reducing their relevance for decision-making in relation to school choice. Different audiences - parents, policymakers, researchers and school leaders - have divergent purposes, levels of data literacy and interpretive needs, yet, as shown in the thematic analysis of documentation in Phase 1 (section 4.2.1), SPMs are expected to serve them all. This 'one-size-fits-all' approach means that performance data are stretched too thin, attempting to communicate too much to too many and in the process, often misrepresenting reality.

In some government publications on SPMs, there is emphasis on the importance of recognising their limitations. Users are advised to exercise caution in their interpretation and to use them only in conjunction with other information when drawing conclusions about schools. This formed part of the findings in section 4.2.3. However, there is a perceived disjoint, expressed by participants in phase 3 (section 6.1.2.1), between these warnings and the way that SPMs are used, including by government departments and the media. This raises questions about the consistency of the message and whether more could be done to support the appropriate interpretation of SPMs, addressing the diverse needs, priorities and statistical skills of stakeholder groups.

The terminology used in SPMs, including the very notions of 'performance' and 'school', adds further ambiguity, while the volume of available indicators, changing grading systems and the technical design of measures like the EBacc, Progress 8 and Attainment 8 contribute to confusion. This complexity in SPMs was expressed by participants in the phase 3 focus groups (section 6.1.2.1) and aligns with other research. For example, the Education Policy Institute (2025) highlights how measures such as Progress 8, Attainment 8 and the EBacc create interpretive challenges for stakeholders due to their technical design, while Prior et al. (2021) emphasise that even well-intentioned value-added models can be misunderstood, particularly when combined with changing grading systems and the volume of indicators. Together, these suggest that SPM complexity and terminology can lead to the obscuring of school performance. Meanwhile, the use of the term 'measure' in SPMs implies a sense of precision and objectivity, suggesting that school performance can be quantified and compared with scientific accuracy. This assumption may obscure the interpretive and value-laden nature of what is being measured. As a result, rather than clarifying school effectiveness, current approaches to communicating SPMs frequently distort interpretation—undermining their intended purpose of supporting informed understanding and accountability.

8.3.4 Wider Implications for the School System

Stakeholders' beliefs about trust, transparency and fairness in accountability systems significantly shape the behavioural and systemic effects of school performance measures (SPMs), influencing how educators respond to metrics, how parents interpret results and how the wider education system adapts to accountability pressures (Braun, 2005; Espeland & Sauder, 2007; Newton, 2016; Ofsted, 2017; Whitty, Power & Halpin, 1998). This is echoed by the views expressed by participants in the phase 3 focus groups (sections 6.1.1.2, 6.1.2.1, 6.1.2.4, 6.1.3.2 and 6.1.3.3). The current system rests on the belief that performance data and associated measures are legitimate tools for holding schools to account. However, as focus group discussions revealed, SPMs often operate within a zero-sum framework, exemplified by Progress 8 and league tables, where schools are positioned in competition rather than collaboration, creating a system of 'winners and losers'. This framing distorts educational purposes and creates systemic pressures that prioritise comparative outcomes over collective improvement. The close alignment between SPMs and Ofsted gradings amplifies this effect: limited, context-insensitive quantitative metrics heavily influence inspection judgements, reinforcing a perception that accountability is misleading and unbalanced. This dynamic also exposes tensions between Ofsted's evaluative authority and DfE driven performance measures, producing an accountability imbalance that undermines professional trust. Schools consequently tend to focus on satisfying the requirements of the inspection framework rather than fostering internal, self-directed improvement, reflecting an assumption that external pressure is necessary for progress. Evidence

from comparative research suggests that such intensified accountability pressure often leads to unintended consequences, including narrowing of the curriculum, constrained professional autonomy and increased stress for educators as systems prioritise external compliance over collaborative improvement (Altrichter & Kemethofer, 2015; Brill et al., 2018). These outcomes can be seen as emergent properties of a tightly coupled, high-stakes accountability environment - one that feeds back on itself, narrowing the scope for learning and innovation.

These dynamics also highlight how SPMs are not objective, neutral tools but socially and politically situated instruments of data power that shape perceptions, behaviour and decision-making within schools, consistent with critical data studies showing that data infrastructures and metrics co-produce social realities rather than merely represent them (boyd & Crawford, 2012; Kitchin, 2014). From this perspective, misleading or poorly interpreted SPMs do not simply reflect school performance. Instead they actively construct narratives of success and failure, framing what counts as legitimate knowledge and whose interests are served by particular definitions of performance (Beer, 2016).

Ultimately, the misleading and competitive nature of SPMs leads to an erosion of trust, destabilising the legitimacy of the accountability system and constraining the school system's capacity for genuine, sustainable improvement.

8.3.5 Conclusion to SPM effectiveness

Drawing on the policy analysis, quantitative data and stakeholder focus groups, this discussion of the effectiveness of SPMs suggests that the production, publication and operationalisation of SPMs in England fall significantly short of their stated goals to reflect whole-school effectiveness or ensure fairness accurately. A core problem is the narrow focus; SPMs are overwhelmingly focused on academic attainment, overlooking crucial aspects such as personal development and wellbeing. Furthermore, quantitative findings reveal that three of the six headline SPMs overlap heavily, strongly reflecting a single underlying factor: attainment, suggesting a limited concept of 'performance'. This overlap among headline measures reinforces a narrow, attainment-centred concept of performance, potentially misleading stakeholders into treating correlated indicators as independent signals of school quality. This is a nuance rarely emphasised in existing discussions of SPM design (Education Policy Institute, 2021; Gorard & See, 2013; Leckie & Goldstein, 2019b).

A profound issue lies in fairness and bias. The current SPM structure is demonstrably dominated by school intake demographics, with the majority of attainment variance explained by student background. While previous studies have documented the strong association between attainment-

based SPMs and pupil background, the Phase 2 quantitative analysis extends this work by demonstrating the extent to which intake characteristics dominate the overall structure of headline measures, casting doubt on whether the system meaningfully captures school performance at all. In this sense, SPMs function as inadequate and unreliable proxies for genuine school quality. The absence of systematic contextualisation within current SPMs can be interpreted as reflecting an ideological commitment to a neoliberal-influenced 'no excuses' accountability culture, reinforced by neo-conservative emphases on standardisation and central control, in which 'fairness' is framed as holding all schools equally accountable regardless of context, thereby risking the reproduction of existing social inequalities. This interpretation is consistent with critical policy research which argues that school accountability systems grounded in 'no excuses' logics deliberately downplay social and contextual inequality, framing uniform standards and sanctions as fairness while legitimising unequal outcomes (Apple, 2006; Ball, 2003b; Gillborn, 2010b).

Beyond structural flaws, interpretive challenges severely limit SPM effectiveness. SPMs are widely perceived by diverse audiences as complex, confusing and poorly understood, as reported by participants in the Phase 3 focus groups (section 6.1.2.1). This finding aligns with previous research showing that school performance measures such as Progress 8 and Attainment 8 are poorly understood by parents, teachers and even school leaders, who frequently struggle to interpret their meaning, limitations and appropriate uses (Hutchinson, 2018; Mansell et al., 2009). However, this analysis extends that literature by showing that interpretive confusion is not merely a communication failure but also interacts with the underlying intake-dominated structure of SPMs, amplifying the risk of misattributing performance and reinforcing misleading narratives of school quality. The underlying assumption that SPMs clearly communicate school quality fails because these measures often obscure the extent to which observed outcomes are driven by student characteristics rather than the quality of educational provision. Research on England's secondary school accountability system shows that SPMs, when unadjusted for pupil background, can over- or underestimate school performance because differences in socioeconomic and demographic intake account for a large share of the variation in outcomes (Leckie & Goldstein, 2019a). A broader review identifies statistical weaknesses in how such measures handle pupil background and cautions that unadjusted performance metrics can mislead interpretations of school contributions to pupil progress (Prior et al., 2021). Work examining adjustments for multiple pupil characteristics further demonstrates substantial impacts on school rankings and classifications (Leckie & Goldstein, 2019a; Leckie & Goldstein, 2020).

Finally, the wider system implications of contemporary school performance measures highlight a corrosive dynamic. The high-stakes, zero-sum nature of SPMs, particularly when they are explicitly

linked to accountability mechanisms such as Ofsted gradings, actively fosters competition at the expense of collaboration. This competitive environment incentivises strategic behaviours, including manipulating student cohorts, focusing resources on marginal test-takers and prioritising metric satisfaction over substantive educational improvement, that can distort professional practice and organisational priorities. These dynamics not only undermine trust in peer relationships and local networks but also erode the legitimacy of the entire accountability framework. Empirical evidence supports these concerns: for example, von Stumm et al. (2021) demonstrate that school quality ratings, such as those produced by inspection regimes like Ofsted's, are weak predictors of students' achievement once contextual and background factors are taken into account, calling into question the validity of linking such ratings to high-stakes performance measures. In doing so, this research confirms broader critiques of SPM systems, i.e. that they misrepresent school contributions to student outcomes and builds on them by showing that the very measures used to enforce accountability may be poor proxies for the qualities they purport to capture, thereby exacerbating the corrosive effects on educational practice and trust within the system.

In section 7.5 I will turn to developing ideas for directions for alternative approaches to SPMs, aiming to address the various issues arising from an analysis of SPM effectiveness. The underlying principles which raise crucial questions about the effectiveness of SPMs in meeting their stated aims are:

1. Headline SPMs are excessively narrow in scope. They concentrate on a limited and overlapping set of outcome measures, primarily academic attainment, progress, curriculum breadth and post-16 destinations, while marginalising many of the broader educational, developmental and relational functions of schools. This thesis extends existing critiques of metric narrowness by demonstrating how this restricted scope shapes both school behaviour and public understandings of what constitutes educational 'performance'.
2. SPMs insufficiently account for the contextual challenges faced by schools serving different communities. As shown in earlier analyses, a substantial proportion of variation in headline measures is explained by student intake characteristics rather than by differences in school provision. Building on this evidence, the thesis argues that the failure to meaningfully integrate context into SPM design not only produces misleading signals of school quality, but also contributes to systematic inequities in accountability judgements.
3. SPMs are required to serve a broad range of target users, including parents, teachers, school leaders, governors, inspectors and policymakers. This thesis makes an original contribution by foregrounding the role of data literacy, interpretive capacity and user purpose in shaping how SPMs are understood and acted upon. It shows that the

communication of SPMs rarely reflects these differences, creating predictable risks of misinterpretation, misuse and oversimplification.

4. The purposes of SPMs are embedded within an evolving policy logic that rests on particular conceptual and ethical foundations, aims and rationales, values and assumptions and conditions and expected processes of operation. By unpacking SPMs as a complex assemblage of interacting ideas, incentives and institutional practices, this thesis exposes tensions between trust, fairness, accountability, transparency and control that are typically treated in isolation. This system-level perspective reveals why incremental technical adjustments to SPMs are unlikely to resolve their more fundamental legitimacy problems.

These four considerations related to SPM efficacy, alongside ideas which emerged from the focus group discussions, will provide a basis for the suggested alternative approaches discussed in section 7.5.

8.4 SPM effects

The complex, entwined nature of SPM effects, as recognised in the three research phases constituting this thesis, makes systematic unpacking challenging. To structure this analysis, I draw on a rich literature showing that school accountability systems produce effects across system, organisational and individual levels (Hanushek & Raymond, 2004; Loeb & Figlio, 2011).

These levels are:

- **Macro level** - captures systemic, structural and political effects embedded in the policy architecture of accountability. Analyses identify the incentives and unintended consequences generated by performance-based accountability systems (Hanushek & Raymond, 2004; Loeb & Figlio, 2011).
- **Meso level** - highlights organisational and cultural transformations within schools and local systems. Accountability pressures have been linked to changes in school culture, staff morale and organisational climate (Erichsen & Reynolds, 2019; Smith & Holloway, 2020).
- **Micro level** - shows every day behavioural and professional adaptations to performative pressures. Research demonstrates that teachers and school staff adapt their instructional practices, experience heightened stress and modify professional behaviour in response to high-stakes accountability demands (Ball, 2003b; Perryman, 2009; OECD, 2019).

Having said that, these levels are deeply connected and serve here primarily as an analytical tool rather than a claim that SPMs have effects that can be attributed solely to specific levels within the school system's structure. For instance, SPMs can motivate teachers to 'teach to the test' (micro level), which not only narrows classroom practice but also influences institutional priorities (meso level) as schools reorganise curricula around accountability metrics. This, in turn, amplifies systemic pressures (macro level) that equate educational quality with test performance, creating a self-perpetuating cycle in which performance data drives, rather than reflects, educational practice.

This multi-level framing also allows the analysis to connect with insights from CDS and FATE research. CDS research has traditionally focused on either micro-level everyday interactions with data (e.g., teachers' interpretations and use of SPMs), meso-level organisational practices (e.g., school adaptation to accountability regimes), or macro-level systemic and political effects (e.g., policy logics and structural incentives), but rarely integrates all three (Russo & Iliadis, 2016; Schäfer et al., 2024). By situating SPM effects across macro, meso and micro levels, this thesis offers an original synthesis that spans these typically siloed perspectives. Similarly, while FATE research often attends to fairness and ethical concerns at the micro level, awareness of meso and macro-level contexts is essential to

understanding how individual adaptations are shaped by organisational pressures and systemic incentives (Jobin et al., 2019). The macro/meso/micro framework therefore enables a more holistic account of the ethical, political and practical implications of SPMs.

8.4.1 Micro level (typically Professional/Behavioural)

8.4.1.1 Teaching to the test

Due to pressures and incentives arising from the publication of SPMs, teachers adapt their classroom practice to align with measurable indicators of success, as shared by participants in the Phase 3 focus groups (section 6.1.3.3). The logical behaviour for a teacher is to focus professional effort and curriculum design towards achieving outcomes that demonstrate compliance with policy expectations rather than fostering deep, holistic learning in students (Ball, 2003b; Perryman, 2009). Such rational behaviours are incentivised by potential rewards or sanctions in the shape of reputation, inspection grades and, ultimately, funding, through improved student recruitment. The analysis of SPM policy documents in research Phase 1 suggests that SPM policy's implicit, and somewhat simplistic, assumption is that a focus on measurable outcomes drives improved teaching and learning. However, too often the logical consequence is a strategic shift away from complex, non-measurable educational aims toward only what can be counted. Participants in the Phase 3 focus groups speak clearly about their perception of this effect and their views are consistent with previous research on high-stakes accountability (Au, 2007; Ball, 2003b; Perryman, 2009). This creates a conflict between SPMs' purposes (the improvement of learning) and their effects in reality (tactical behaviours to maximise scores in SPMs). I made reference to this and the concerns expressed in the focus groups about the prevalence of such practice in section 6.1.3.2.1.

'Teaching to the test' illustrates how SPMs, as data infrastructures and assemblages (Kitchin, 2014), are neither neutral nor purely descriptive. Instead, they are socially and politically constructed entities that define what counts as valid learning. SPMs become the ultimate arbiters of 'good' or 'successful' teaching, privileging quantifiable outcomes over relational, creative or holistic aspects of pedagogy, which are consequently rendered invisible, raising concerns around justice and equity. This process involves data power, the capacity for data to wield influence, create inequalities and shape society. From this perspective, SPMs exercise professional control, subtly shaping what is valued and what knowledge is legitimised. The act of publishing SPMs is performative. It does not simply report on educational practice, as suggested by the Phase 1 policy document analysis (section 4.2.1.1), but actively reshapes it, creating feedback loops that, from the perspective of focus group participants (section 6.1.3.3), influence teacher behaviour and curriculum design over time. Within this governance assemblage, teachers negotiate their role as educators with the demands of

producing data for accountability systems, balancing pedagogical aims with the pressures of reporting, comparison and evaluation, as suggested by participants in the Phase 3 focus groups (section 6.1.3.3). Meanwhile, the complex, relational, or creative aspects of pedagogy are consequently rendered invisible within the governance assemblage of SPMs, as they are not recognised by policymakers, inspectors, or parents and are often sidelined by teachers navigating performative pressures. From this point of view, SPMs do not merely measure performance; they co-produce educational realities, privileging what can be counted and systematically marginalising what cannot.

'Teaching to the test' practices also undermine fairness and explainability. The practice privileges students whose learning aligns with narrow forms of assessment (Au, 2007), neglecting wider skills like creativity, critical thinking and personal development. While the system achieves some "pseudo-transparency" **(**Vujnovic & Kruckeberg, 2016, p.1) by making test scores and examination grades visible, ultimately it erodes transparency about what is being valued and assessed and it obscures the explainability of outcomes. SPMs show 'performance,' but not the quality of teaching or depth of learning that sits behind it. At the same time, accountability for the systemic failure to support these broader educational aims is lost, as the focus is almost entirely on examination grades rather than the learning process itself. High-stakes accountability systems encourage this focus by privileging measurable outcomes and SPMs explicitly use exam results as the key indicator of school performance (Au, 2007; Ball, 2003b; Leckie & Goldstein, 2019b; Kitchin, 2014; Perryman, 2009).

8.4.1.2 Impact on wellbeing

The high-stakes school accountability system and published performance data in the form of SPMs are policy instruments intended to create pressure for improvement, as shown by the SPM policy document analysis of Phase 1 (section 4.2.1.3). The logical, albeit somewhat unintended, consequence is the creation of a stress-inducing environment driven by a fear of sanction or poor ranking. This was clearly expressed by participants discussing work pressure and the high-stakes nature of the school accountability system in the focus groups (section 6.1.3.3). School leaders and teachers are therefore responding rationally to the pressures described by focus group participants when they, for example, work beyond their contractual hours (Perryman & Calvert, 2020), internalise the SPMs' goals as their own primary professional goals (Parcerisa et al., 2022) and constantly, often subconsciously, monitor and adjust their own teaching practice and professional decisions to ensure they align with the expected data outcomes (Au, 2007; Jerrim & Sims, 2022). The psychological toll and its consequences for professionals' wellbeing, as discussed by participants of the focus groups, contradict the policy's broader goal of sustaining a high-quality education system. This illustrates a

failure in the policy's conditions of operation where intense pressure arising from the operation of SPMs generates self-defeating negative externalities, referred to by Valli and Buese (2007, p. 519) as "unanticipated and often negative, consequences for teachers' relationships with students, pedagogy and sense of professional well-being".

This is an example of data governance extending into the emotional and psychological spheres of work (Beer, 2016). The demand for continuous, measurable performance requires emotional labour (Hochschild, 1983) and self-surveillance (Siegel et al., 2022). Teachers internalise the metrics, not just for external inspection, but for their own peace of mind, as discussed by participants in the Phase 3 focus groups (section 6.1.3.3). From a critical data studies perspective, this internalisation of SPMs by teachers illustrates the performative and political power of data, emphasising that SPMs are socially and institutionally constructed artefacts that shape professional behaviour. This creates a culture in which the educator's body and mind are positioned as extensions of the data infrastructure, constantly subject to scrutiny and optimisation pressure, with the resulting stress becoming a side effect of the data-driven pursuit of institutional success. This aligns closely with Beer's (2016) concept of metric power, which highlights how the measurement, circulation and interpretation of data not only shape organisational practices but also exert power over individuals' behaviours and self-assessment, revealing how governance through metrics extends beyond formal rules into the psychological space of those governed.

These impacts on wellbeing highlight a significant lack of accountability for harms. While schools are held responsible for their data outputs, the system itself often escapes accountability for the negative effects it has on its workforce, which focus group participants (section 6.1.3.3) and other research identify as a) work pressure and stress (Jerrim & Sims, 2022), b) loss of autonomy and professional agency (Valli & Buese, 2007) and c) impacts on recruitment and retention (Wei et al., 2009). This raises questions about fairness by shifting the burden of systemic risk—such as poor SPM policy design or insufficient funding—onto individual professionals. The resulting harms often remain invisible because SPMs focus on student achievement, not staff emotional wellbeing, enabling policymakers to overlook the systemic damage and the threat it poses to the long-term sustainability of the education system.

8.4.1.3 Narrowing of pedagogy

The SPM policy logic, based on the idea that quality can be measured and compared, leads to the prioritisation of core academic subjects that are suitable for standardisation (Au, 2007) as observed by focus group participants (section 6.1.3.3). Any pedagogical approach (e.g., creative arts, critical thinking) that deviates from the measured criteria poses a logical risk to the school's SPM standing.

SPMs thus inherently discourage diverse or complex teaching methods that are hard to quantify, thereby limiting the concept of education and professional practice to fit the metric. This was discussed by focus group participants (section 6.1.3.3) and is supported by the work of Au (2007, p. 258), in which he argues that the "primary effect of high-stakes testing is that curricular content is narrowed to tested subjects, subject area knowledge is fragmented into test-related pieces and teachers increase the use of teacher-centered pedagogies."

Furthermore, SPMs reconfigure pedagogy itself, as indicated by Au (2007). Focus group participants discussed how pedagogy is transformed from a complex, context-specific professional craft into a collection of standardised practices focused on what can be easily measured (section 6.1.3.3). This is often exhibited through rote learning, skill drilling or content coverage, as explained by National Research Council (2011, p.40), "*Overly narrow instruction might include such practices as drilling students on practice questions, focusing on the limited subset of skills, knowledge and question formats that are most likely to be tested*". This situation exemplifies quantification bias, where the easily measurable replaces what is professionally valued. SPMs serve as a prescriptive technology, prescribing how teachers should teach and leading to the algorithmic erosion of professional autonomy and agency discussed by focus group participants (section 6.1.3.3), since the 'data-correct' approach substitutes the teacher's professional judgment of diverse student needs (Verger et al., 2016).

This narrowing of pedagogy also fundamentally questions fairness for pupils who thrive with alternative learning styles (e.g., project-based, experiential, or creative approaches), another idea arising in Phase 3 focus group discussions. It promotes an unfair standard of success that only suits learning modes most compatible with standardisation, as exemplified by a quote from the Dr Adam Boddison, former Chief Executive of nasen (UK's primary professional body for SEND) in GL Assessment (2019, p.8) "*It is through the arts, heritage and culture that some learners with SEND can access education, but instead these learners are spending increasing amounts of time on the subjects that carry the most value within the accountability frameworks*". At the same time, this narrowing of pedagogy diminishes the transparency of what constitutes 'success.' While the data is transparently reported, the definition of educational quality behind the metric remains opaque and biased (section 6.1.3.3), failing to recognise that the system chooses to prioritise one narrow form of learning success over others, thus obscuring a genuine, open view of educational quality (Biesta, 2015).

8.4.2 Meso level (typically Institutional/Organisational)

8.4.2.1 Perverse incentives and gaming

From a policy logic perspective, gaming, in the form of 'off-rolling' (the removal of students from the school roll for reasons that suit the school's, rather than the student's, interests), curriculum narrowing or unethical admissions practices, as considered by focus group participants (section 6.1.3.2.1), is a rational and logical institutional response to an SPM policy and a school accountability system designed around high-stakes incentives. When a school's legitimacy and survival are tied to SPMs, institutions will strategically comply by focusing on actions that improve the SPM i.e. the visible output, rather than necessarily improving underlying educational quality (Cullen & Reback, 2006), ideas explored by focus group participants (section 6.1.3.2). From a critical data studies perspective, this represents a case of metric performativity, where a policy instrument, in this case SPMs, reshapes behaviours in ways that undermine its own objectives. Once success is defined through a narrow range of quantified indicators, strategic compliance and data optimisation can be prioritised over substantive educational improvement (Beer, 2016; Espeland & Sauder, 2007).

When examining the role of SPMs in incentivising gaming behaviours, we can consider the performative effects of metrics. Schools do not just measure their performance; they actively 'produce' good data to survive and thrive within the audit culture, an idea explored at length in the Phase 3 focus groups (section 6.1.3.2). Gaming is a political performance designed to satisfy the data infrastructure's demand for high scores. It is a practice which is widely recognised by stakeholders across the English schools' system, for example by Ofsted (2019a) and House of Commons Children, Schools and Families Committee (2008) and by researchers examining the impact of accountability on schools (Au, 2007). Gaming behaviours demonstrate that the data is not a neutral reflection of reality, but a socially constructed artefact that schools are able to engineer. This performativity reinforces the power of the data system, as high scores appear to validate the policy, even if the underlying educational reality has actually been strategically warped.

Perverse incentives and gaming directly undermine fairness. This has been widely recognised for a considerable length of time, as demonstrated in a speech by the Secretary of State for Education (2012) "…*league tables can be corrupted. Too much reliance on one measure as a target — however well-designed that single target may be — will mean gaming can occur*". The accountability system risks rewarding manipulation and strategic data management rather than genuine, sustained educational improvement and penalising schools that choose not to 'game' the system. While the data appears transparent, the underlying, unacknowledged manipulation creates an opacity of true 'performance'. The system is accountable for the score, but not for the deceptive means used to

achieve it, making it difficult for stakeholders (parents, policymakers) to rely on the data as a true measure of quality, ultimately undermining the credibility of SPMs as performance metrics.

8.4.2.2 Disincentives for transparency

Given the high-stakes, competitive nature of the SPM system, as revealed in Phase 1 (section 4.1.1.4) and Phase 3 (section 6.1.1.4) analyses, one of its key conditions of operation, the logical institutional behaviour for schools is risk management. Openly reporting poor data or contextualising failure risks inviting sanctions and reputational harm, which would be irrational in a competitive quasi-market, as focus group participants noted (section 6.1.3.2). Therefore, the logic of the SPM policy creates a fundamental tension between the goal of transparency and the institutional aim of survival (Roberts, 2009). As a result, SPMs risk inadvertently encouraging data opacity, thereby protecting school reputation while undermining the policy's aim of system learning and informed accountability.

By controlling the availability and interpretation of data through this data opacity, schools can seek to manage external perceptions and thereby their position in the competitive system. This opacity is a mechanism of power (Beer, 2016). It obscures underlying inequities, complexities, or failures (Ball, 2003b), shielding the institution from accountability. Despite its ultimate goal of visibility, the data system inadvertently enables schools to selectively perform transparency while obscuring the 'messy' data that might threaten their perceived quality or market standing (Lubienski, 2007).

This behaviour directly weakens transparency and explainability, two fundamental principles of data ethics. Without open reporting or honest contextualisation, the information accessible to the public and policymakers remains incomplete, making it challenging to accurately explain the drivers of either success or failure. It also undermines accountability, as the system becomes less able to identify genuine problems when schools are incentivised to conceal them (Power, 1997). This lack of clear, context-rich information hinders the fair and targeted distribution of support and resources needed to address systemic shortcomings, as noted by Ball (2000) in his consideration of the relationships between data, accountability, transparency and governance in education.

8.4.2.3 Loss of autonomy and hindering local innovation

By establishing centralised SPMs as the definitive measure of 'success,' the SPM policy limits local discretion and innovation capacity, as noted by focus group participants in Phase 3 (sections 6.1.3.2 and 6.1.3.3) as impacting schools on multiple levels, from classroom practice to school system level and restricting schools' ability to adequately serve the needs of their community. The rational school leader understands that pursuing local, innovative practices that are not reflected in, or that pose a risk to, the mandated SPMs carries significant risk (Biesta, 2010). The policy logic effectively

subordinates the professional judgment of the needs of the local community to the requirements of the national data system, resulting in a loss of agency in favour of maintaining central control and adhering to standardised metrics (Ozga, 2009).

Focus group participants shared their perception that local practice, knowledge and diverse community needs are thus subordinated to national metrics (section 6.1.3.2). In their comments, we can see how the data infrastructure functions as the tool of this centralisation, converting complex local realities into standardised, comparable and centrally accessible data points (Ozga, 2009). This highlights the role of SPMs in exercising control through data power (Beer, 2016), as seen in one focus group participant's reference to headteachers being "infantilised" (section 6.1.3.3) within a schools' system shaped by SPMs and accountability (Ball & Olmedo, 2013). This process eliminates local differences, enforces a uniform definition of value and ultimately enables the government to govern through data, effectively extracting autonomy while claiming to provide an objective measure of school quality (Ball, 2003b).

This dynamic exposes the tension which Ball and Olmedo (2013) refer to as existing between accountability and agency, which fundamentally weakens both fairness and explainability (Hopmann, 2008). It is unfair to hold local school leaders accountable for SPMs when central policy simultaneously limits their professional autonomy and ability to address local needs effectively. The lack of local discretion diminishes the system's explainability. When performance is poor, it becomes impossible to determine whether it results from local failure or the restrictive nature of the centralised mandate. As Ozga (2009) notes, centralised data systems "*reduce the capacity to explain the causes of failure or success*" (p. 151). The accountability system, therefore, becomes a tool of control rather than a shared responsibility for improvement (Strathern, 2000).

8.4.2.4 High-stakes pressure & performativity culture

A performative culture is a consequence of the SPM policy's design, through its emphasis on accountability and standardisation. The high-stakes nature of SPMs forces schools to prioritise the visible performance necessary to satisfy external scrutiny. This pressure drives a risk-averse compliance where schools focus on producing good data rather than engaging in the complex, reflective work required for genuine improvement. The logical imperative, therefore, is to manage the external gaze, making performativity both a rational and pervasive feature of the policy's operation.

This can be characterised as an example of 'audit culture', the application of accounting and financial management techniques to the governance of organisations and people, along with associated social and cultural effects (Shore & Wright, 2015, 2024). Schools are trained to be increasingly self-monitoring, continuously anticipating and managing the demands of the external data system. Ball (2003b, p.216) argues that "performativity is a technology, a culture and a mode of regulation which employs judgements, comparisons and displays as a means of control". This culture transforms internal processes into perpetual preparation for external scrutiny, through the publication of SPMs and inspection reports, where data become instruments of internal discipline, as noted by Perryman (2009). The performance is not just for the outside world; it is internalised, leading to a constant state of anxiety and a diversion of resources to improve data production. Decisions are justified in relation to their effect on SPMs or inspection gradings rather than on their educational merits. As Power (1997, p. 57) observes, "actors respond rationally to the incentives created, producing behaviours aligned with audit requirements rather than substantive improvement". This ultimately consumes energy and time that could have otherwise been devoted to real educational innovation.

The consequences of this high-stakes performativity extend beyond procedural compliance, impacting both staff and students. Research has shown that intense accountability pressures are associated with increased teacher stress, burnout and attrition (Ingersoll, 2001; Lavy, 2015). Teachers report feeling constrained in their professional judgement, often narrowing the curriculum and adopting teaching-to-the-test practices to safeguard measurable outcomes rather than fostering deep learning (Au, 2007; Flockton et al., 2017). These studies collectively highlight that the design of SPMs, when coupled with high-stakes consequences, does not merely reshape institutional behaviour but can produce tangible psychological and educational harms.

Through the operation of SPMs, the state holds schools accountable, but the design of the high-stakes system itself is often shielded from any reciprocal accountability for its systemic effects. Ball (2003b) notes that top-down performance measures are imposed with the aim of "holding schools and teachers accountable while the policies themselves are rarely subject to critique or accountability" (p. 218). While data are transparently reported, the pervasive, risk-averse and crucially performative nature of the culture means the resulting data are a poor and biased reflections of reality, compromising their explainability as valid indicators of true educational effectiveness (Perryman, 2009).

8.4.2.5 Competition rather than collaboration

By using SPMs as league tables to promote parental choice and competition (Leckie & Goldstein, 2009), a direct consequence of the policy's embedding within a neoliberal and market-driven

ideology, SPM policy structures schools as rivals. The logical institutional response in a competitive environment is to hoard resources and expertise in an attempt to gain a market advantage. There is a call for greater collaboration between schools, in various forms, to encourage school improvement. This can be found within, for example, government policy documents (section 4.1.3.2), more recent remarks by the current Secretary of State for Education, Rt Hon Bridget Phillipson MP (Dutaut, 2025) and comments from many of the stakeholders in the focus groups (section 6.1.4.2.4). In this context, the policy's explicit aim of improvement is thus undermined by its implicit competitive logic, which renders cooperation illogical and structurally prevents system-wide collective accountability or learning.

The SPM data infrastructure actively promotes comparison and ranking. The design of league tables, data dashboards, benchmarks and normalised scores aims not just to track performance but to foster a competitive environment, as noted by Espeland and Sauder (2007, p.3), rankings "*do not simply measure performance but actively reshape organisational behaviour and relationships through competition*". In this context, SPMs act as a tools of comparison, turning neighbouring schools into fierce rivals. This process is a crucial aspect of data-power (Beer, 2016), which shapes relationships between organisations (Power, 1997), using quantified metrics in the form of SPMs to uphold a competitive dynamic that fragments the system and hinders the development of collective, systemic efforts to tackle common, often locally rooted challenges.

Indeed, the competitive structure unfairly penalises collaboration. Schools that do collaborate may see their performance relative to rivals suffer, creating a clear fairness gap. Focus group participants noted how SPMs disincentivise collaboration between schools (section 6.1.1.4), despite what they perceived to be recognised benefits as exemplified in the gains made in the London Challenge (Ogden, 2008). These comments echo the remarks made by Hargreaves (2003, p.106), that market-oriented accountability systems "*reward individual institutional success while penalising collective effort and shared responsibility*". Brown and Lauder (1996, p.12) argue that this lack of collaboration also reduces systemic transparency and discourages "*the open sharing of professional knowledge, undermining collective learning*". Knowledge, insights and lessons learned remain trapped within individual institutions (Ozga, 2009), making it harder to understand and address system-wide issues. This also weakens overall accountability because systemic failures become more difficult to attribute or resolve when institutions are encouraged to operate in isolated, self-serving silos.

8.4.3 Macro level (typically Systemic/Political)

8.4.3.1 Stratification (SPMs as measures of intake; resource effects; recruitment/retention)

A key finding from the Phase 2 quantitative analysis (section 5.4.2) was that headline SPMs, especially the attainment metrics that are dominant, are heavily influenced by intake demographics. This aligns with patterns long established in the literature, as discussed in section 2.4.1 and was noted by participants in the Phase 3 focus groups (section 6.1.2.3). Consequently, SPMs, as tools for comparing schools based on the neoliberal assumption of a 'level playing field' (Reay, 2012), an idea identified in the Phase 1 SPM document thematic analysis in relation to the formation of school league tables (section 4.1.1.4), reinforce existing structural inequalities. As Ball (2003a, p.23) puts it, "*advantage is systematically converted into further advantage*". Schools with high socio-economic intake are viewed as more desirable by parents, as noted by focus group participants (section 6.1.2.1), while Burgess et al. (2015, p.1279) find that higher socio-economic status parents are "*more successful in translating preferences into actual school admissions*". Parents who are less limited by factors like transport costs or inflated house prices tend to be more successful in school admission applications than their disadvantaged peers, thereby entrenching advantage, as discussed by Allen and Vignoles (2007). SPMs, intended to ensure accountability, naturally result in resources and reputation flowing towards advantaged schools, creating the "*winners and losers*" described by focus group participants (section 6.1.2.3), thus systematically worsening, rather than alleviating, social stratification (Lupton, 2005; Thrupp, 2001).

This highlights how SPM data assemblages interrelate, for example, ranking, funding and inspection, to reinforce patterns of advantage (Ball, 2016; Ozga, 2009). This exemplifies the CDS axiom that "data are not neutral" (D'Ignazio & Klein, 2020, p.18); it is a technology that reflects social hierarchies by capturing advantage through metrics. Kitchin (2014) emphasises that data "*are framed by the systems that produce them and the interests they serve*" (p. 4). SPMs, correlated with socio-economic status, thus become data proxies for wealth rather than educational effectiveness, as argued by Thrupp and Lupton (2006). The data infrastructure that transforms social advantage into institutional 'success' makes the resulting stratification seem objective and meritocratic (Porter, 1995). The use of SPM metrics legitimises a highly unequal system as a fair outcome of differences in school quality (Espeland & Sauder, 2007), rather than a reflection of existing privilege and systemic inequity.

This systemic stratification is unfair and damages accountability. It is fundamentally unjust to evaluate schools mainly on their performance data (Gorard, 2010) when such data are largely influenced by the social makeup of their student bodies. As Thrupp and Lupton (2006) argue,

accountability systems that ignore context "*penalise schools for factors largely beyond their control*" (p. 309), undermining both fairness and the ethical basis of accountability itself (Biesta, 2004).

8.4.3.2 Inclusivity (SEND and AP)

Since the system defines success narrowly through standardised attainment, the rational institutional behaviour is to manage the student body to maximise those scores (Ball, 2003b). Students with complex learning or assessment needs (SEND or in Alternative Provision) act as a disincentive to high performance because their diverse outcomes do not align with the standardised metrics, as discussed by focus group participants (6.1.3.1). The policy logic, therefore, structurally promotes practices such as strategic exclusion that marginalise inclusion in favour of data optimisation, a practice that Gillborn and Youdell (2000) term "*educational triage*". The most vulnerable students are therefore systematically de-prioritised in the pursuit of high scores (Gillborn, 2010b). Since SPMs fail to capture relevant outcomes for these groups, the SPM policy overlooks real differences in school mission and context (Biesta, 2009). I will explore the implications of this for the suggested policy redesign later in sections 7.5.1 and 7.5.2.

This is a powerful illustration of how invisibility arises from omissions in what is measured (Espeland & Stevens, 1998). Students and more unusual educational routes, such as vocational training, that do not fit the standardised metric are rendered invisible. Indeed, this was observed by Alison Wolf (2011) in her review of vocational education for 14- to 19-year-olds: performance measures have historically "failed to recognise the value of vocational and applied learning routes" (p. 7). At the same time, the data system, focused on normalisation and comparison, creates the category of the 'outlier' or the 'exceptional' (Ozga, 2009), allowing schools to strategically manage these students outside the core performance metrics. Ultimately, the measures systematically exclude or poorly measure the progress of these groups, making the final metrics inherently unfair to the schools that serve them (Biesta, 2009; Thrupp & Lupton, 2006).

This marginalisation represents a profound failure of both fairness and transparency within SPM policy. Accountability systems that evaluate schools without recognising context are "fundamentally unjust to schools serving disadvantaged or atypical intakes" (Thrupp & Lupton, 2006, p. 309) and systematically penalise those facing the greatest complexity (Lupton, 2005). At the same time, performance data strip away contextual information in order to enable comparison, reducing transparency by rendering key challenges invisible (Kitchin, 2014; Ozga, 2009). As Biesta (2004) argues, such systems are ethically compromised when they hold institutions responsible for outcomes without acknowledging the conditions under which those outcomes are produced.

8.4.3.3 Policy centralisation (loss of local discretion)

The logical function of the centralisation of SPMs is to standardise and scale accountability, enabling governance at a distance through comparable data (Ozga, 2009; Power, 1997). However, their role in implementing government policy endangers local agency by favouring systemic coherence and controlled compliance across the entire schools sector, choosing regulation and comparability over local context and autonomy (Ball, 2021), thus "*prioritising system coherence over local responsiveness*" (Hargreaves & Fullan, 2012, p. 93). Phase 3 focus group participants discussed their concerns around this stifling of local innovation (section 6.1.3.2) by the operation of SPMs.

SPMs, as centralised data systems, are not just technical tools of measurement; they are powerful instruments of governance and surveillance that "*shape and regulate conduct at a distance*" (Rose, 1999, p.49). Their infrastructure dictates who can hold whom accountable and how they might do this (Power, 1997). SPMs, as mechanisms of control, contribute to the consolidation of centralised power within national bodies (Ozga, 2009) e.g., the Department for Education or Ofsted. Meanwhile, focus group participants raise questions about the degree of democratic accountability of Ofsted (section 6.1.2.2), while they underscore the central role of Ofsted in shaping school responses to SPMs.

Centralisation leads to a concentration of accountability without reciprocal transparency, with processes which "*concentrate accountability upwards*" (Power, 1997, p.43). The central government holds schools accountable via data, but the centralised policy decisions themselves often lack transparency in their formation, restricting meaningful challenge or a robust mechanism for accountability regarding their systemic impacts (Ball, 2003a). From an ethical perspective, governance structures should be accountable to those they affect (D'Ignazio & Klein, 2020). This imbalance is fundamentally unfair since local actors bear the risk. They are held accountable via SPMs, while the policy architects who define the metrics and the system's structure are often insulated from accountability for the negative consequences of their decisions (Thrupp & Lupton, 2006).

8.4.3.4 Flawed assumptions (system continues 'as if' valid)

A fundamental problem within the SPM policy logic is the political system's requirement for legitimacy and institutional credibility. Despite evidence questioning the validity of the SPMs' measurements or causal assumptions, the political drive remains to treat them 'as if' educational quality can be measured accurately (Espeland & Sauder, 2007) and that SPMs genuinely reflect school performance (Gorard, 2010). The rational choice for policymakers is to preserve the framework's consistency and stability rather than risk political upheaval by acknowledging that the

accountability system's foundation is flawed (Ball, 2021). This is a dangerous policy pathology. It undermines the policy's fundamental rationality and calls for a wide-ranging and critical examination of the assumptions justifying SPM-driven interventions.

CDS terms this 'epistemic closure' (Williams, 2025): datafication creates a feedback loop in which metrics become the only visible evidence, while alternative knowledge is sidelined. As Espeland and Sauder (2007) argue, **"***numerical indicators come to shape - and be treated as - the very reality they claim merely to represent***"** (p. 3). The data system begins to act as a self-referential authority, where SPMs are used to confirm the assumptions that created them (Porter, 1995). Their ongoing operation on flawed premises serves to reproduce misinterpretations as facts. Policy debates are closed down by the assertion that 'the data proves it', even if the data itself is only a poor proxy for reality. By ignoring the flaws, limits and uncertainties exposed through critique of the data, the metric (e.g., SPM) is afforded a higher epistemic status than other forms of knowledge or judgement (Fricker, 2007), such as professional, contextual or qualitative.

This fundamental issue calls into question the explainability of SPMs. Policies may be justified by data that cannot meaningfully account for context or causality since they are based on a flawed metric (Gorard, 2010). If the metric is a poor or biased measure of quality, then any policy derived from it is also unjust and ineffective. It also fails to meet the ethical standard of transparency, which requires that statistical adjustments, limitations and modelling assumptions be open and understandable to all stakeholders (Kitchin 2014; Porter, 1995). The entire system lacks accountability because the ultimate justification - the metric's validity - is not open to genuine external scrutiny or audit (Biesta, 2004). The opaque, flawed assumptions at the macro level create a governance system that lacks the essential ethical ability to honestly explain its actions and consequences. It could be argued that, ethically, it is the duty of policymakers to demonstrate the validity of SPMs and to mitigate harms when their underlying assumptions break down.

8.4.4 Conclusion to SPM effects

The analysis shows that SPMs, despite their intentions, create a complex web of interconnected and often problematic effects across the entire schools' ecosystem. At first glance, what was introduced as a policy to improve school quality and transparency can now be seen as a powerful data infrastructure that controls behaviour (Beer, 2016), distorts priorities (Ball, 2003b) and reinforces inequality (Gorard, 2010).

From micro-level pressures that push teachers to 'teach to the test' (Au, 2007) and compromise their professional wellbeing (Ball, 2003b), to meso-level institutional gaming , risk-averse opacity (Power,

1997) and stifled local innovation (Biesta, 2010), the system consistently values measurable metrics over authentic educational quality. At macro level, this results in systemic stratification, marginalisation of vulnerable students (Thrupp & Lupton, 2006) and the concentration of centralised political power (Ball, 2012).

Ultimately, reliance on SPMs sustains a flawed, self-confirming policy logic where the data infrastructure functions less as a neutral reporting tool and more as a powerful technology that determines what is considered valid learning, thus undermining fairness, transparency and the very autonomy and collaboration essential for genuine educational improvement. Across all levels, the logic of quantification overrides professional judgment, justice and context, creating an illusion of objectivity. SPMs not only measure but actively construct an educational reality—maintaining a self-referential system that prioritises data over learning, compliance over understanding and control over genuine improvement.

Next, I will turn to consider suggested directions for alternative approaches to SPMs, which would seek to address the various issues arising from an analysis of the effects of SPMs. I can summarise the two main influences behind these undesirable effects as being:

1. SPMs operate within a high-stakes accountability system rooted in the neoliberal policy logic of competition, choice and performative accountability, which emerged in the late 1980s. While the literature widely acknowledges the unintended consequences of such systems, my research extends this work by empirically tracing how these effects are produced and sustained. Through thematic analysis of policy documentation, I demonstrate how the stated purposes of SPMs are tightly aligned with this policy logic, while stakeholder focus groups reveal how the resulting accountability pressures shape professional, institutional and system-level behaviours in practice. This triangulation shows that unintended consequences are not incidental side-effects but structurally embedded outcomes of the policy logic within which SPMs function.
2. Headline SPMs are narrow in scope and, as critics have long suggested, are strongly influenced by student intake characteristics. My quantitative analysis advances this critique by systematically quantifying the extent to which intake demographics explain variation in headline SPMs, demonstrating that they predominantly reflect contextual factors rather than school effectiveness. Despite longstanding cautions regarding the validity and reliability of these measures, the accountability system continues to treat them as precise indicators of school quality. By integrating this quantitative evidence with an analysis of SPMs through the lenses of a Policy Logic Framework and the lenses of Critical Data Studies (CDS) and Fairness,

Accountability, Transparency and Ethics (FATE), my research provides a coherent evaluation of SPMs that moves beyond critique to a structured assessment of their practical and ethical limitations.

I will use these, in addition to the four considerations related to SPM efficacy (section 7.3.5), to provide a basis for the suggested alternative approaches discussed in the next section.

8.5 Alternative Approaches

The evidence examined earlier indicates that the limitations of current school performance measures (SPMs) in terms of their effectiveness and unintended consequences stem not only from technical flaws in their design, which fail to account for school context and offer a very narrow view of student outcomes, making them both misleading and unfair, but also from the underlying policy logic that shapes their development and use. SPMs operate within a neoliberal-influenced, high-stakes accountability system rooted in competition, market logics and performance comparisons. This has resulted in certain system behaviours, such as gaming, curriculum narrowing, risk aversion, reduced opportunities for collaboration and the marginalisation of schools serving disadvantaged communities. Additionally, the diverse range of user groups for SPMs—each with different needs, priorities and capabilities—raises important questions about how to best support appropriate communication and meaningful interpretation of these measures. All these factors interact in complex ways, ultimately undermining SPMs' primary goal of enhancing educational quality.

Alternative approaches must therefore go beyond the necessary technical 'fixes' to also address the conceptual and structural foundations of the system. The following four subsections explore possible directions for SPM reform:

1. expanding the scope of SPMs to reflect a more holistic conception of school quality
2. embedding contextualisation into the design of SPMs to ensure fairer and more meaningful comparison
3. improving communication and interpretation for diverse audiences
4. reconfiguring accountability to support learning and collaboration rather than competition, compliance and sanction

While extensive research has critiqued the limitations and unintended consequences of school performance measures (Ball, 2003b; Espeland & Sauder, 2007; Gorard, 2010), proposed alternatives have tended to focus on technical changes to existing metrics, such as broader indicators (Biesta, 2010), contextualised comparisons (Thrupp & Lupton, 2006) or narrative reporting (Ehren & Shackleton, 2016), rather than on the underlying purposes, assumptions and uses of SPMs. An original contribution of this thesis lies in developing an integrated set of alternative approaches that move far beyond technical fixes to address the scope, design, interpretation and governance of performance measurement as a coherent policy system, grounded in robust empirical evidence and informed by critical policy analysis, Critical Data Studies and FATE perspectives that reconceptualises SPMs simultaneously as measures, representations and instruments of governance.

8.5.1 Moving Beyond Academic Metrics to Provide a More Holistic Picture of School Quality

SPMs reflect a logic model that assumes quantifiable performance pressure on academic outcomes will lead to systemic improvement (Biesta, 2009), a dominant theme identified in the SPM documentation thematic analysis of Phase 1 (section 4.2.1.2). This reasoning constrains the meaning of 'performance' to what can be measured through standardised tests, thereby limiting a broader understanding of the institutional purpose of schooling, summarised by Ball (203) as "*only that which can be counted counts*" (p. 215). The result is a system that primarily understands itself through metrics (Power, 1997) of attainment and progress, marginalising other potential dimensions of educational value. Consequently, a fundamental problem with the current SPM system is its epistemic narrowness—it recognises only a very limited subset of what schools actually do (Bowker & Star, 1999).

This is partly a consequence of the datafication of schooling exemplified by SPMs (Williamson, 2017). A rich and diverse school life is reduced to mere data points, while what remains unmeasured becomes invisible and devalued (Espeland & Stevens, 1998). SPMs play a role in actively shaping educational realities, not just in recording and reporting them (Beer, 2016). Although academic attainment and progress may be seen as key indicators, they alone cannot encompass the broader aims of schooling — such as supporting personal development of capable, ethical and fulfilled individuals, fostering civic responsibility, nurturing creativity and sustaining inclusive, cohesive communities (Biesta, 2010).

A redesign of SPMs grounded in FATE principles would aim to counter these tendencies by expanding the scope of school work to include ethical, human and community dimensions. This might involve prioritising student development and wellbeing within a more comprehensive range of indicators, as recognised in the focus group discussions on this topic (section 6.1.4.3.4). A system which included wellbeing data as a vital aspect of school quality would require careful management, utilising ethical frameworks to guide the collection and use of personal, sensitive data (UNESCO, 2017b). The inclusion of student wellbeing data would likely require adding qualitative data alongside quantitative academic assessment data as part of SPMs. While headline SPMs currently provide only a very narrow slice of information about the work of schools, as discussed by focus group participants (section 7.1.2.4), primarily around academic attainment and progress, the potential of the inclusion of student wellbeing as part of a far broader, more granular picture of schools is one which has both emerged in the past as well as having some more recent examples. This approach draws on some

ideas and information that are already available to stakeholders. A broader picture of a school might be formed by combining the following:

- SPMs, including wider aspects of SPMs beyond the headline measures
- Ofsted inspection reports, which include information about student behaviour, attitudes and personal development
- Ofsted ParentView[3], an online survey for parents and carers to share views on their child's school, primarily designed to inform inspection
- Information about schools from school publications (e.g. prospectuses) or available through school websites
- School data available from a range of government, media and independent sources, including a range of dashboards targeted at different user groups (school leaders, parents) with different priorities and functionalities
- Directly through a school visit, including school open events etc.

However, locating, compiling, comparing and understanding these data can be challenging for many users.

Secondly, the School Report Card (SRC) (DCSF, 2008), which was developed at the end of the New Labour government period but never reached the implementation stage, was intended to include indicators of student wellbeing. The project's ambition was to incorporate a wide range of such indicators, using quantitative measures such as attendance, exclusions, or take-up of school lunches, alongside qualitative indicators that might be derived from perception surveys of parents and pupils.

Finally and more recently, there have been a number of developments to widen the scope of school information available to stakeholders. This perhaps suggests an appetite, reflected in the opinions of participants in the focus groups, for a move to provide stakeholders with information about schools which goes beyond purely academic outcomes. These developments include:

- Revised Ofsted inspection grading and 'report card' (Ofsted, 2025) (not to be confused with the SRC above), which increases from 4 to between 6 and 8 evaluation areas, including personal development and wellbeing, attendance and behaviour and inclusion. Unlike the School Report Card proposed in 2008, which aimed to incorporate quantitative wellbeing indicators and qualitative perceptions from pupils and parents, the new Ofsted report cards represent a

[3] Most recent publication of management Information from Ofsted on the use of Parentview (Ofsted, 2025, October 31) shows that the mean response rate (proportion of pupils in school who have submitted an entry on Parentview) is only 13%

relatively modest step toward a more comprehensive view of schooling that nonetheless stops short of the full wellbeing-centred, stakeholder-driven model envisioned in the original SRC proposals

- The Department for Education has announced plans to introduce Digital School Profiles (Department for Education, 2025d). While details are still to be released on this initiative, the aim is to widen the scope of information available to parents about schools by recognising the breadth of what a school does across a range of criteria. A separate service is to be developed for use by schools with the aim of providing data-insights to inform school improvement work and supporting greater collaboration between schools
- There are numerous examples of groups of schools (LocalEd2025, 2025) and other organisations (Rethinking Assessment, 2024) have recently developed models for 'school report cards', 'school profiles' or 'balanced scorecards' which would include diverse information about schools, in many cases using innovative approaches which utilise the opportunities afforded by new technologies

Some of the qualitative data which would inform such broader SPMs would include the use parent and student voice. Schools have used surveys using psychometric design to explore student attitudes for some time (GL Assessment, 2025). However, these are typically used by schools to identify non-academic barriers to learning. As such, they would need some significant revisions to be used in a different context as part of the information about schools or the school system. The data generated from student and parent surveys might also be used alongside other forms of data, including quantitative data in the form of, for example, participation rates for extra-curricular activities (Robinson, 2024) or engagement (e.g., enrolments, withdrawals and completions) with support service programmes (Prebble, 2005). The UK Department for Education also commissions research (GOV.UK, 2025, 29 September) to understand the perspectives of parents and students. This is typically used to inform the design of effective policy in relation to schools and so, again, would not translate directly to use within a reformed structure of SPMs for parents, for example. Having said that, these established models can provide a basis for further work to incorporate robust data around non-academic aspects of the work of schools in the information shared about them with various stakeholder groups.

An idea complementary to broadening the information available to stakeholders is to move away from providing prejudged 'performance data' and instead provide a range of information that allows the user to interpret it in their own way. In this sense, parents, for example, could take a personal view of all of the information provided about a school rather than being told which schools perform

the ‘best’ or ‘worst’, thus embracing more diverse opinions of school ‘quality’. Parents could use such information to find out how well a school fits their values, priorities, interests and preferences. This approach would signal a move away from the simplistic rank ordering of schools based on a small number of, mainly academic attainment, metrics, to allowing stakeholders some freedom to judge for themselves what the data tells them about schools. A parent might then choose a school, not because it is objectively the ‘best school’ but because it is subjectively the ‘best school for their child’. Rather than claiming that SPMs are data which objectively capture school ‘quality’, SPMs, in a reformed state including greater breadth, might instead be considered to provide users with the non-judgemental ‘qualities’, i.e. attributes of schools.

Schools do not operate in isolation. They are interconnected through various means, including with other schools through informal connections, formal federations (Howland, 2015) or Multi-Academy Trusts (MATs) (Chapman, 2015) for shared governance and resources. They also engage with Local Authorities (LAs), Higher/Further Education institutions (e.g., for teacher training) (Glazzard & Tate, 2025), local businesses (Bonnen, 1984) and their immediate parent/local communities (Giovannini, 2023). To better understand the wider contribution of a school in supporting its local community, including other schools, we might therefore move beyond just asking “what does a school do for its students?” to “what does a school contribute to the educational system and its locality?”

To address this, within SPMs, we might include aspects of the broader work of schools as the ‘good citizens’ mentioned in section 6.1.4.2.3, contributing positively to other schools, organisations, businesses, groups and communities. We might envisage new indicators of their collaboration with other schools, partnerships with local organisations and service to local communities, which would acknowledge schools’ civic role.

The active involvement of schools with their local community and broader society is fundamentally connected to inclusion. Schools that actively support and engage with their local communities tend to promote greater inclusion, as such partnerships enhance social cohesion, foster mutual understanding and create learning environments that value diversity and participation beyond the school gates (Ainscow, 2020; Dyson & Todd, 2010). Measuring school inclusivity involves multiple components, including inputs (e.g., equitable admissions, physical access, inclusive hiring and staffing), processes (e.g., inclusive policies, classroom practices, student voice and participation, culturally responsive pedagogy) and outcomes (e.g., sense of belonging, retention, attendance, fair access to opportunities, reduction in exclusion and disciplinary disparities) (Education Policy Institute, 2022; Mezzanotte, & Calvel, 2023). For instance, the widely used Index for Inclusion conceptualises inclusion through three interconnected domains of culture, policy and practice and provides indicators such as “all students are equally valued” and “barriers to learning and

participation are reduced" (Booth & Ainscow, 2002). Simultaneously, EPI's proposed quantitative measures at the school-group level include odds ratios of admissions of disadvantaged pupils, rates of persistent absence and suspensions and achievement gaps—all proxies for the inclusiveness of a school's intake, retention and support systems (Mills, 2019). What remains a challenge and is often highlighted in the literature, is that many aspects of inclusion, such as lived student voice, relational dynamics, subtle exclusion or off-rolling. Since these are difficult to quantify, mixed methods or survey instruments capturing perceptions (e.g., staff and pupil responses on inclusive ethos) are often used alongside quantitative data (Mezzanotte & Calvel, 2023). Perception scales, such as the Sentiments, Attitudes and Concerns about Inclusive Education [SACIE] scale, can be utilised to assess teacher self-efficacy, school leadership and the extent of authentic stakeholder participation, including students and parents (Loreman et al., 2014).

Here, we observe the complex interplay of various facets of school work. Leadership approaches shape school policies and culture, which in turn shape, for example, admissions practices and teacher empowerment, which in turn impact student and staff demographics and teacher job satisfaction and retention (Loreman et al., 2014; UNESCO, 2017a). High staff turnover is both costly and disrupts learning. Measuring staff retention, especially in key subject areas and high-deprivation contexts, might serve as one proxy for the effectiveness of school leadership, the school culture and more general working conditions. Indeed, the centrality of the role of school leadership and governance (Colman, 2021) within these entangled aspects of schools, as illustrated in the example above, prompted the suggestion from focus group participants that leadership and governance might be the key aspects of schools which might require evaluation (section 6.1.4.3.3). This might include

a. capturing leadership behavioural practices (such as vision-setting, strategic resourcing, staff development and monitoring teaching and learning) which can be captured through validated survey instruments e.g., the School Implementation Leadership Scale (Lyon et al., 2022; Wills & van der Berg, 2020) or the Multi-factor Leadership Questionnaire (MLQ) (Bass & Riggio, 2006)
b. leadership distribution and structure, operationalised via tools like the Distributed Leadership Inventory, which assesses how leadership responsibilities are spread across staff (Leithwood, 2012; Spillane, 2022)
c. outcomes-based metrics, such as the existing SPMs or changes in student achievement which might be modelled using value-added approaches to estimate the leader's contribution (Clifford et al., 2012; Wills & van der Berg, 2020)

Crucially, the literature highlights the need for instruments with strong psychometric properties (e.g., reliability, validity) and cautions that leadership measurement must account for its contextual

complexity, the indirect nature of leadership effects and the limitations of attribution (Wills & van der Berg, 2020).

Teacher stability, support and empowerment are certainly important for sustained school improvement. Measures of professional learning, workload balance and staff satisfaction might well act as useful proxies for organisational health. Student and staff demographics, the results of perception surveys and measures of inclusive practices would no doubt provide valuable information. For schools, these evaluations might support strategic school improvement planning, while for policymakers, they might provide insights into wider system health as well as where resources might be best targeted. However, their publication for use in a broader set of SPMs, intended to support school choice, might raise several questions.

First, there is a risk of misinterpretation. Measures of staff wellbeing or professional culture are inherently complex and context-dependent, yet public reporting would inevitably require some degree of simplification. Parents and other stakeholders may interpret low satisfaction or workload scores as evidence of poor school quality, when in fact they may reflect transparent self-assessment, recent organisational change or other challenging external conditions. Moreover, schools with more reflective or self-critical cultures could be penalised for honesty, while others may appear more favourable due to less candid reporting. These are the “disincentives for transparency” discussed in section 7.4.2.2.

Second, public dissemination may generate perverse incentives, as discussed previously regarding the existing SPMs in section 7.4.2.1. High-stakes publication tends to encourage data management behaviours such as selective reporting, coaching of responses or prioritisation of appearance over substance. School leaders may become more concerned with ‘performing’ organisational health than meaningfully improving it. In more extreme cases, such pressures could deter open discussion of workload or morale issues, undermining the very improvement processes these indicators might be intended to support.

Third, there are ethical and privacy considerations. In smaller schools, survey data could risk identifying individuals or small staff subgroups, compromising confidentiality. Public reporting may also discourage honest participation in perception surveys if staff fear reputational consequences for the school or themselves.

Fourth, publication of organisational health data could contribute to market distortion and inequality. Parents with greater social and cultural capital are more likely to use such information to avoid schools perceived as less favourable. Meanwhile, disadvantaged parents might lack the means to avoid such schools due to the impact of house prices or transport costs, further concentrating

disadvantage. This dynamic risks creating self-reinforcing cycles of reputational harm: once labelled as having poor staff morale or inclusion practices, schools may find it harder to attract skilled staff or engaged families. This is the process of accelerated stratification discussed in relation to the existing SPMs in section 7.4.3.1). Additionally, competition for positive reputational outcomes may weaken inter-school collaboration, which is often seen as essential for sustained improvement.

Finally, there are conceptual and measurement challenges. Indicators of staff wellbeing, professional learning and inclusivity are subjective, variable and heavily influenced by contextual and systemic factors such as funding, policy pressures and local demographics. As such, their validity and comparability across schools are limited and their attribution to school-level leadership may be uncertain.

In sum, while organisational health indicators have considerable potential for internal evaluation and professional dialogue, their use as public instruments for parental choice risks reproducing many of the distortions already observed in existing SPMs. A more constructive approach may involve using such data within trusted professional networks, peer review systems or collaborative improvement partnerships, where the emphasis is on learning rather than competition, or by policymakers, to inform funding and support decisions. I will revisit this idea when discussing the communication of SPMs in section 7.5.3 below.

Ultimately, the introduction of a broader set of school-related data as part of reformed SPMs would represent a step change in their operation. First of all, while I have continued to use the term SPMs throughout for consistency, many aspects of schools would be either difficult or impossible to capture in a quantified metric. Their inclusion in the school information provided to stakeholders would require the use of both qualitative and quantitative data, moving away from the concept of a defined, small set of School Performance Measures towards collections of school information that present information to users rather than pre-baked judgements. Secondly, what is included within potential future collections and what is omitted, would depend on who gets to decide on their content, what they consider schools are for and what aspects of schooling are most valued to them. The use of focus groups, comprising a range of stakeholders, in this phase might provide a foundation upon which to build, one in which school data users might be actively involved in future decision-making regarding SPM design, communication and interpretation.

8.5.2 Contextualised Data to Allow Fairer Comparisons and to Address Bias

Even using broader school information, any attempt to evaluate school performance must grapple with the challenge of contextual variation. Schools operate in environments shaped by socio-economic, cultural and demographic factors that strongly influence their outcomes. Treating

performance data as if it were independent of context distorts both fairness and interpretability. For the current headline SPMs, the failure to account for contextual factors means SPMs reflect intake far more than school effectiveness (section 7.3.1).

As referenced in section 6.1.2.3, contextual value-added (CVA) was a prominent headline SPM between 2006 and 2010, when it was discontinued. The reasons cited for its removal were that it was difficult for the public to understand and, importantly, that adjusting for differences in pupils' socioeconomic and demographic backgrounds meant CVA reinforced low educational aspirations among disadvantaged student groups (Department for Education, 2010). This reflects a neoliberal and meritocratic ideological stance that emphasises individual responsibility and the moral agency of schools and students. From this perspective, educational successes or failures are viewed as the result of individual effort and school quality, rather than structural inequalities.

However, a contextualised approach to SPMs would recognise that student and, therefore, school outcome differences often reflect disparities in opportunity caused by systemic inequity, rather than just variations in institutional quality. Methods such as multilevel modelling (MLM) and, as presented in section 5.4.2, structural equation modelling (SEM) can help disentangle the influence of contextual factors (e.g., socio-economic status and prior attainment) from the effects due to school practices. The inclusion of contextualised SPMs, therefore, would go a considerable way to addressing one of the fundamental flaws of the current system, its 'contextual blind spot', providing measures more closely aligned with school effectiveness. Contextualisation, therefore, is not an excuse-making exercise; it is an essential step towards a fairer system. By recognising the social determinants of educational outcomes, it aligns performance measurement with both empirical accuracy and moral integrity.

On a more exploratory level, as discussed in section 5.5, cluster analysis can be used to identify patterns within groups of schools facing similar challenges. This approach could provide new and valuable insights into the existing work on the concept of 'similar schools' (Department for Education, 2020; FFT Education Datalab, 2025; SchoolDash, 2025), making performance comparisons more meaningful. It also helps reduce perverse incentives related to ranking and allows policymakers to spot systemic patterns, such as regions or school types facing especially severe challenges.

Moreover, the recent efforts to compile collections of information about schools discussed in the previous section (7.5.1), suggest opportunities for linking diverse data sources such as pupil background data, workforce statistics, school funding, student outcomes and neighbourhood characteristics to create a much richer evidence base in the form of 'school profiles'. Such models of school information might also include some degree of flexibility to include information related to

national, regional, local or institutional priorities, providing a focus for users on identified issues and how they are being addressed. While such 'school profiles' may need to be communicated with care depending on what data are included in them—an idea I will address in the next section —they could offer a significantly improved foundation for informing parental choice, school improvement, the diagnosis of problems around structural inequalities and guide both targeted support and wider policy development.

8.5.3 Reconceptualise SPMs as Communicative Tools to Meet Different User Needs and Support Interpretation and Understanding

SPMs have multiple audiences. These include parents, teachers, governors, inspectors, policymakers and the media. Each of these will have distinct information needs and varying capacities for data interpretation. The current mode of SPMs, which privileges simplified tables and single-number rankings, encourages over-interpretation and misinterpretation. It values reputational school hierarchies over informed decision-making.

Reform of SPMs requires not only technical redesign, for example, using some of the ideas suggested above, but also a reconceptualisation of SPMs as communicative tools. To meet their stated aims, the data presented in SPMs should aid sense-making for parents, schools, inspectors and policy makers. This might involve considering the language used around SPMs. For example, reviewing and clarifying the term 'school performance measures' itself might support better interpretation. This might provide greater clarity around:

- What is the 'school' that SPMs refer to? Is it the school leadership, the school's students, the teaching and support staff or the entire institution?
- How is 'performance' conceptualised? What is valued and what is ignored? And, linked to the question above, whose performance do SPMs represent and how might this information be used?
- How accurate, precise and reliable are SPMs as 'measures' and might the use of other terms, for example, 'indicator' rather than measure, guide users to exercise greater care when using them, leading to more appropriate and consistent interpretations?

Sense-making using SPMs could also be supported by:

- Layered presentation of information, where users can move from simple summaries to more detailed contextual explanations as needed.

- Interpretive guidance highlighting the limitations, margins of error and appropriate uses of each measure.
- Tailored reporting for different audiences—for example, professional dashboards for school leaders focusing on improvement planning, alongside more accessible summaries for parents.
- Narrative contextualisation, allowing quantitative results to be accompanied by interpretive commentary or case illustrations that humanise data.

All of the above would be enhanced by adopting consistent approaches to the communication and interpretation of SPMs. For instance, it is important that any narrative commentaries accompanying SPMs adhere to the guidance provided by the DfE referred to in section 4.2.4. This could include using more than one SPM when comparing schools or carefully considering margins of error when using measures, rather than relying on simple rank orders based on raw scores.

The layering of information availability and customised reporting for different user groups within alternative approaches to SPMs could extend beyond simply how school information is presented. Instead, various user groups might be provided with access to different types and levels of information tailored to their specific needs. This could build on some of the current practices. For example, Ofsted inspectors receive the Inspection Data Summary Report (IDSR) for schools before inspection (Ofsted, 2025). These reports are not made public but are accessible to schools via a UK government website login. The IDSR contains detailed data on schools, including in-depth breakdowns of student academic outcomes, as well as staffing, student mobility, attendance and behaviour. Recently, the DfE announced (The Guardian, 2025) that every school in England will be issued with an AI-generated target for minimum pupil attendance. These targets will not be published or shared with Ofsted but will instead be used to connect schools with higher-performing institutions working in similar contexts, facilitating the exchange of ideas. This perhaps indicates a more supportive, positive and collaborative approach to using school data—concepts I will explore further in the next section (7.5.4).

Improved communication and interpretation could therefore strengthen data literacy across the whole system, reduce the misuse of statistics and promote more informed, reflective engagement with SPMs. This shift would also support greater accountability by improving data transparency, comprehensibility and openness to critical dialogue.

8.5.4 Reconfiguring Accountability to a More Supportive System, Built on Improved Trust

The most profound alternative direction involves going beyond the design and communication of SPMs towards a wider reimagining of the purpose and structure of accountability itself. Current arrangements rely heavily on top-down inspection, external judgment and punitive intervention, with SPMs serving as triggers for actions, including inspections and interventions. This model has achieved limited success in improving outcomes and has generated substantial collateral damage—narrowing curricula, demoralising staff and exacerbating inequality.

The entire school accountability system in England fundamentally rests on a foundational assumption of distrust: that schools and teachers will not improve without external pressure and surveillance (discussed in section 6.1.2.2.3). Systems of accountability are therefore considered essential to bring about the improvements needed. This presumption shapes a system oriented around control rather than collaboration (Didau, 2015), producing compliance behaviours, risk aversion and performativity rather than intrinsic, meaningful school improvement. An accountability system built on distrust, corrodes professional morale, weakens internal motivation and fosters adversarial relationships not just between competing, rival schools, but also with inspectors and policymakers. Drawing on Onora O'Neill's account of trust (O'Neill, 2002), the way forward is not to demand blind trust but to cultivate trustworthiness (O'Neill, 2018)—through transparency about methods, honesty about limitations and integrity in purpose. Trust cannot be legislated, but it can be earned when institutions demonstrate competence, reliability and ethical intent. Reorienting accountability towards demonstrable trustworthiness would mean evaluating not only outcomes but the quality of processes and relationships: whether schools act ethically, communicate openly and are committed to learning and improvement. A trust-based accountability model could thus replace external compulsion with reciprocal responsibility—where schools, policymakers and inspectors each earn and sustain one another's trust through trustworthy practice.

A new accountability paradigm might not be based only on trust, but also on support and professional learning rather than fear and compliance. Some suggested directions which emerged from focus group discussions were:

- Moderated self-evaluation and triangulation: Schools could undertake systematic self-evaluation, using SPMs as one strand of evidence, moderated through peer review or local collaborative networks. This would position performance data as diagnostic rather than summative - a means for identifying areas for inquiry and support.

- Redefining the inspectorate's role: Inspection could move from judgment to facilitation, supporting schools' self-evaluation processes and capacity-building.
- Collaborative accountability: Clusters or partnerships of schools could share responsibility for collective outcomes within an area, reflecting the interdependence of schools within educational ecosystems. This aligns with complexity theory's recognition that system-level improvement arises through relationships and feedback rather than central command.
- Diversified measures and selective emphasis: Incorporating a wide range of metrics and rotating the focus areas (rather than using a fixed set of headline indicators) could disrupt gaming and reward authentic breadth of provision.

Such a model would reposition accountability as a relational and formative process—one that builds trust, nurtures professionalism and fosters continuous improvement. It would also align the educational system more closely with ethics of care, fairness and collaboration, principles considered to underpin the broader purpose of education.

8.6 Future directions for SPMs

Future development of school performance measures might therefore move beyond incremental reform and address the systemic alignment between measurement, accountability and educational purpose. The re-envisioned approaches discussed above—holistic, contextualised, communicatively responsible and trust-based—point towards a system that treats schools as partners in learning rather than subjects of surveillance.

This reconfiguration requires co-design involving educators, researchers, policymakers and communities, recognising that valid and legitimate performance frameworks are social constructions, not neutral instruments. Building such frameworks will involve sustained dialogue about values, evidence and the kind of education system society wishes to sustain.

In this sense, the future of SPMs is not merely a technical matter but a moral and political choice—about what we value, how we define success and how we hold each other to account in the shared enterprise of education.

8.7 Conclusion to Discussion

The persistence of School Performance Measures (SPMs) in English education, despite clear analytical flaws exposed by the Policy Logic Framework (PLF), Critical Data Studies (CDS) and FATE framework, reveals that they endure not through technical oversight but because they perform

powerful political, institutional and governance functions. From a policy logic perspective, SPMs provide apparent legitimacy for reforms such as academisation and rebrokering (the process by which a school moves from one academy trust to another, typically to improve performance, governance or support), anchor accountability systems administratively and sustain narratives of crisis and improvement that justify ongoing intervention. Policymakers value them not for their accuracy but for their symbolic and administrative utility: they offer simple, communicable metrics that signal oversight, reinforce the ideological framework of competition and choice and create the appearance of evidence-led governance. Once embedded within inspection, funding and governance infrastructures, SPMs become administratively indispensable and resistant to replacement or radical reform. After 35 years of SPMs, even envisaging a schools system operating without SPMs in their current, or similar, form becomes a challenge for many stakeholders. SPMs represent an enactment of neoliberal ideology, as clearly evidenced by the thematic analysis of government SPM documentation. However, these ideas, which are rarely (if ever) named as neoliberal-influenced, are so deeply embedded in school practices and policies that rather than being seen as policies arising from a particular contestable political and economic ideology, they are seen as essential parts of the system. As such, they are considered to be something more akin to a 'natural law', an unquestionable truth (Monbiot & Hutchison, 2024).

Viewed through a CDS lens, SPMs are not neutral representations of school quality but sociotechnical artefacts that shape knowledge, behaviour and power. They render schools and teachers legible to the state, structure institutional responses and performatively generate the outcomes they purport to measure. Data production and interpretation are never neutral: they simplify complex educational realities, obscure methodological assumptions and depoliticise debate by shifting attention from structural inequalities to school-level performance. Recognising this reveals that improving SPMs demands not only technical refinement but also a fundamental reconsideration of how data are produced, interpreted and embedded within policy systems.

From an ethical and normative standpoint, the continued reliance on flawed SPMs raises serious concerns regarding Fairness, Accountability, Transparency and Ethics (FATE). Measures that inadequately adjust for context misrepresent school quality, reinforce systemic inequities and invert the logic of accountability—disciplining schools rather than scrutinising the system itself. The opacity of statistical models and technical assumptions further erodes trust, undermining both professional autonomy and democratic legitimacy. Moving forward, the focus should shift from governing through data to fostering ethical and authentic accountability. This means developing data systems that illuminate rather than obscure structural inequalities, ensuring contextual transparency and

embracing multidimensional accountability models that balance oversight with professional judgment and democratic participation.

The future of school performance measurement should therefore be an integrative and ethical project—transforming data from a mechanism of top-down control into an instrument of systemic improvement and social justice. By adopting a critical data infrastructure approach, policymakers can design metrics that expose the structural conditions shaping educational outcomes and hold themselves accountable for the fairness of the system as a whole. Attending simultaneously to the political, epistemic and ethical dimensions of SPMs would allow the development of measures that are not only technically sound but socially legitimate, educationally meaningful and genuinely supportive of schools. In this evolution, data would cease to serve as a tool of surveillance and instead become a foundation for trust, transparency and equitable educational governance.

9. Conclusion

9.1 Overview of the Research

This thesis set out to examine how well secondary school performance measures (SPMs) in England serve their stated purposes and what empirical analysis reveals about how they might be improved. It was motivated by a persistent paradox within English education policy: despite decades of critique concerning their validity, fairness and unintended consequences, SPMs remain central to accountability, governance and public discourse. Rather than treating this persistence as a technical failure to be corrected through methodological refinement alone, this research approached SPMs as complex socio-political and sociotechnical constructs whose functions extend well beyond measurement.

To address this problem, the thesis adopted a three-phase mixed-methods design. The first phase investigated the socio-political context underpinning the production and publication of SPMs, analysing government documentation to identify the ideas, assumptions and policy logics embedded within their stated aims and principles. The second phase employed quantitative approaches to interrogate the breadth and meaning of 'performance' as operationalised through headline measures, examining the extent to which these measures capture school quality and align with their intended purposes. The third phase explored, through a series of focus group stages, how SPMs are interpreted, communicated and used by different stakeholder groups and how advances in research and technology might support more meaningful and ethical engagement with school performance data.

These phases were brought together through an explicit synthesis, guided by an overarching research aim: **to examine how well secondary school performance measures in England serve their stated purposes and intended effects and what empirical analysis reveals about how they might be improved**. This overarching aim, in turn, informed three research questions which address context, effectiveness and improvement:

Research Question 1. What is the socio-political context of the production and publication of secondary school performance measures in England?

Research Question 2. To what extent is the production, publication and operationalisation of secondary school performance measures in England effective in meeting their stated aims, in this context?

Research Question 3. How can the findings of research and advances in technology be utilised to improve the communication and use of school data by stakeholders for school performance measurement purposes?

The analysis was framed by an integrated theoretical approach drawing on a novel Policy Logic Framework (PLF), Critical Data Studies (CDS) and Fairness, Accountability, Transparency and Ethics (FATE). This combination enabled the thesis to move beyond evaluative judgments about whether SPMs 'work' and instead interrogate how they function politically, epistemically and ethically within the English education system and subsequently propose directions for alternative approaches to SPMs.

In doing so, the thesis addressed its stated aims and objectives: evaluating the effectiveness of SPMs against their explicit and implicit purposes; critiquing the headline measures and their operationalisation; examining the role of performance data within wider system dynamics; appraising the consequences of SPM publication and use; developing critical analytical frameworks capable of capturing the complexity of the SPM data assemblage; synthesising the findings to suggest meaningful alternative approaches to SPMs as part of the school accountability system.

The remainder of this conclusion synthesises the key empirical and theoretical contributions of the research, sets out recommendations for policymakers and identifies priorities for future research.

9.2 Empirical Contribution

The empirical contribution of this thesis lies not in identifying previously unknown weaknesses in SPMs, but in providing a robust, multi-layered account of how these weaknesses are produced, sustained and normalised across policy, data and practice.

First, the qualitative analysis of government documentation offers a detailed empirical account of the socio-political context in which SPMs have evolved. This analysis demonstrates that SPMs are underpinned by a coherent, if often implicit, set of assumptions, associated with neoliberal ideology, aligned with market-oriented governance: competition between schools, parental choice as a driver of improvement and performance visibility as a mechanism of discipline. While these ideas are rarely named explicitly, their persistence across policy cycles suggests that SPMs function as stabilising devices within a broader ideological framework. Empirically, this work shows how stated aims such as fairness, transparency and improvement coexist with and are frequently subordinated to, administrative imperatives and political narratives.

Second, the quantitative analyses contribute empirical evidence regarding the narrowness and context-dependence of the performance constructs embedded within headline SPMs. This study makes a novel empirical contribution by constructing a latent factor model of school performance at the school level, quantifying the relative contribution of each headline measure and empirically demonstrating redundancy across the SPMs: a contribution not previously realised in the literature. By examining relationships between multiple measures and contextual variables, the thesis demonstrates that a substantial proportion of variation in school outcomes is associated with factors beyond schools' direct control. While this finding aligns with existing effectiveness research, the contribution here lies in showing how multiple headline measures collectively reinforce a narrow conception of performance, rather than offering complementary or corrective perspectives. The combination of factor analysis, SEM and cluster analysis in this thesis represents an original contribution, providing a multi-layered, system-level understanding of how SPMs encode and perpetuate contextual bias. Empirically, the thesis illustrates how the proliferation of measures has not resolved longstanding problems of validity or comparability, but instead risks compounding confusion and driving rather than addressing social stratification, while maintaining the appearance of precision.

Third, the stakeholder-focused analysis provides empirical insight into how SPMs are interpreted and used in practice. The findings reveal significant divergence between the intended communicative function of SPMs and their lived effects. For many stakeholders, SPMs operate less as informative tools and more as instruments of risk management, reputational control and compliance. These effects are unevenly distributed, with schools serving more disadvantaged communities experiencing heightened scrutiny and reduced scope for professional judgement. Empirically, this work highlights how data use is shaped by power relations and institutional incentives, rather than by shared understandings of educational quality. I used the various themes which emerged from Phase 3 focus group discussions, informed by the Phase 1 thematic analysis of SPM documentation and the Phase 2 quantitative analysis of SPM data, to arrive at a series of suggested directions for alternative approaches to SPMs, including to broaden the scope of SPMs, contextualise data, consider how the communication of SPMs might be improved and, ultimately, reconfigure school accountability to be more supportive and built on trust. Taken together, the quantitative findings constitute an original empirical contribution by linking latent performance constructs, intake effects and school clusters, demonstrating not only the narrow operationalisation of performance but also how these measures actively shape schooling practices and reinforce structural inequalities.

Taken together, these empirical findings demonstrate that SPMs do not simply fail to meet their stated aims; they actively reshape the conditions under which schools operate, often in ways that

undermine those aims. This thesis therefore contributes empirical evidence that reframes the debate from one of technical inadequacy to one of systemic misalignment between measurement, governance and educational values.

9.3 Theoretical Contribution

The theoretical contribution of this thesis lies in its integrative conceptualisation of SPMs as sociotechnical governance instruments, rather than neutral representations of educational performance.

A central theoretical contribution is the development, application and extension of the Policy Logic Framework to the domain of school performance measurement. The analysis demonstrates how SPMs function as policy technologies that render schools understandable to the state, anchor accountability systems and legitimise intervention. By showing how methodological flaws coexist with political utility, the thesis advances theoretical understanding of why ineffective or contested policy instruments persist over time. This contribution moves beyond critiques that attribute policy failure to poor design or implementation, instead highlighting the importance of symbolic, administrative and ideological functions. Although developed in relation to school performance measures in England, the Policy Logic Framework may have broader applicability as an analytic tool for surfacing and interrogating the often-implicit assumptions underpinning policy design, particularly those associated with neoliberal modes of governance that emphasise performance, measurement and individualised responsibility for system improvement.

The thesis also contributes to Critical Data Studies by empirically grounding CDS concepts within the context of education accountability. It demonstrates how SPMs act as performative data artefacts: they do not merely describe school performance but actively participate in its production. Through processes of simplification, categorisation and ranking, SPMs shape institutional behaviour, professional identities and public perceptions of schooling. The research extends CDS scholarship by illustrating how data infrastructures in education depoliticise debate, shifting attention away from structural inequalities towards school-level responsibility.

A further theoretical contribution arises from the application of the FATE framework to SPMs. By foregrounding fairness, accountability, transparency and ethics, the thesis articulates a normative critique that complements political and epistemic analyses. It shows how inadequate contextualisation, opaque modelling and high-stakes use of data raise ethical concerns that cannot be resolved through statistical adjustment alone. The integration of FATE into education policy

analysis provides a framework for evaluating not only whether measures are accurate, but whether they are justifiable and legitimate within a democratic system.

Finally, the thesis contributes conceptually by reframing accountability itself. Rather than treating accountability as a unidirectional process focused on individual schools, the research advances a system-oriented understanding that emphasises shared responsibility, structural conditions, collective outcomes and trust. This reconceptualisation challenges dominant accountability paradigms and provides a theoretical foundation for alternative models of performance measurement and governance.

9.4 Recommendations for Policy Makers

The findings of this thesis suggest that meaningful reform of school performance measurement requires a shift not only in technical practice, but in the underlying policy logic that governs how performance data are produced, communicated and used. Incremental adjustments to existing measures are unlikely to resolve the fundamental problems of fairness, trust and legitimacy so long as SPMs continue to function primarily as instruments of control within a high-stakes accountability regime.

First, policymakers should reconsider the purposes to which SPMs are put. If performance data are genuinely intended to support improvement rather than merely signal compliance, their role within high-stakes accountability systems must be substantially reduced. This includes limiting the weight given to headline measures in inspection, intervention and public ranking, and repositioning SPMs as diagnostic tools that inform professional judgement, inquiry and support. Such a shift would require moving away from punitive uses of data towards models of system stewardship that focus on identifying need, allocating resources and enabling learning.

Second, reform requires a broadening and rebalancing of what counts as school performance. The dominance of attainment-based metrics reflects an epistemically narrow conception of schooling that marginalises important dimensions of educational value. Policymakers should support the development of multidimensional school profiles that incorporate wellbeing, inclusion, leadership, organisational health, civic contribution and collaboration, drawing on mixed quantitative and qualitative evidence. Crucially, not all such information should be publicly ranked or standardised; some indicators are more appropriately used for internal evaluation, professional dialogue and system monitoring rather than market-facing comparison.

Third, contextualisation should be embedded as a core design principle of performance measurement, rather than treated as an optional or politically inconvenient adjustment. The

evidence presented in this thesis demonstrates that current SPMs largely reflect student intake and structural conditions rather than school effectiveness. Policymakers should therefore reintroduce and modernise contextualised approaches—using robust methods such as multilevel modelling, peer-group comparisons and cluster analysis—to distinguish institutional contribution from social disadvantage. Doing so would improve empirical validity, address bias and support more morally defensible interpretations of school outcomes, while reducing the harms associated with decontextualised league tables. Importantly, the cluster analysis findings provide new empirical evidence of patterned disadvantage, showing how SPM outcomes map systematically onto groups of schools with similar intake characteristics, offering a foundation for more nuanced, context-sensitive interpretation of accountability measures.

Fourth, SPMs should be reconceptualised as communicative and interpretive infrastructures, designed to support sense-making across diverse user groups rather than simplified ranking. This entails a layered presentation of information, clear narrative contextualisation, explicit guidance on uncertainty and appropriate use and tailored reporting for different audiences, such as parents, school leaders, inspectors and policymakers. Improving the communication of SPMs in this way would reduce misinterpretation, support more informed engagement and strengthen data literacy across the system.

Finally, policymakers should pursue a reconfiguration of accountability grounded in trust, ethics and collaboration. Current arrangements rest on an implicit assumption of distrust and rely on external surveillance, judgement and sanction to secure compliance. This thesis supports an alternative model focused on cultivating trustworthiness through transparency, ethical data use and professional responsibility. Policymakers should adopt an explicitly ethical approach to performance measurement. This involves assessing not only what data can be measured, but what *should* be measured, how data are used and who bears the consequences. Embedding FATE principles within accountability policy would support more legitimate and socially just governance arrangements. Practical steps include embedding moderated self-evaluation, peer review and collaborative accountability at group or area level, redefining inspection as facilitative rather than punitive and diversifying indicators over time to discourage gaming. Such an approach would align accountability more closely with care, fairness and continuous improvement, while acknowledging the interdependence of schools within complex educational systems.

Taken together, these recommendations point towards a fundamentally different role for performance measurement in education: one that prioritises learning over labelling, support over sanction and system improvement over competitive hierarchy.

9.5 Future Research

This thesis opens several avenues for future research. Comparative international studies could examine how alternative accountability systems operate in different policy contexts, providing insight into the conditions under which non-competitive models are politically and institutionally viable. Longitudinal research could explore how changes to performance measurement regimes affect school behaviour, professional practice and student outcomes over time.

Further research is also needed to operationalise and evaluate alternative accountability models, including collective and system-level approaches. The exploratory findings of the cluster analysis at Stage 2 suggest that this warrants further investigation and development. Such approaches have the potential to provide new insights into the challenges faced by groups of schools and how context shapes performance patterns. This includes methodological work on how such models might be designed, as well as empirical studies examining their reception and effects among stakeholders.

Finally, future research could deepen engagement with critical data infrastructures in education, examining how advances in data science, machine learning and visualisation might be harnessed in ways that support democratic accountability rather than surveillance. Building on the frameworks developed in this thesis, such work could contribute to the development of data practices that are not only technically sophisticated but ethically grounded and educationally meaningful.

9.6 Concluding Remarks

This thesis has argued that the persistence of secondary school performance measures in England reflects their political and administrative utility rather than their effectiveness in achieving stated educational aims. By integrating empirical analysis with critical theoretical frameworks, it has shown that improving SPMs requires more than better data: it requires a fundamental rethinking of how performance, accountability and responsibility are conceptualised within the education system. In particular, the empirical contributions of this thesis in the form of detailed analysis of the persistent ideological assumptions and policy logics that underpin SPMs, latent factor modelling, school-level SEM, clustering of schools by intake and performance and detailed insight from a broad range of stakeholders provide a robust, quantified basis for understanding the limits and consequences of current SPMs, and proposing directions for their improvement, marking a substantive contribution to school accountability research. In offering this critique and outlining pathways for reform, the thesis contributes to ongoing efforts to develop accountability arrangements that support equity, trust and genuine educational improvement.

10. References

Acquah, D. (2013). *School Accountability in England: Past, Present and Future*. Centre for Education Research and Policy.

Ainscow, M. (2020). Promoting inclusion and equity in education: Lessons from international experiences. *Nordic Journal of Studies in Educational Policy, 6*(1), 7–16. https://doi.org/10.1080/20020317.2020.1729587

Aitkin, M., & Longford, N. (1986). Statistical Modelling Issues in School Effectiveness Studies. *Journal of the Royal Statistical Society: Series A (General)*, *149*(1), 1–26. https://doi.org/10.2307/2981882

Alasuutari, P., Bickman, L., & Brannen, J. (eds) (2008) *The SAGE handbook of social research methods.* Los Angeles: SAGE.

Alba, J. T. (2024). Insights into Algorithmic Decision-Making Systems via a Decolonial-Intersectional Lens: A Cross-Analysis Case Study. *Digital Society: Ethics, Socio-Legal and Governance of Digital Technology, 3*(3). https://doi.org/10.1007/s44206-024-00144-9

Aldridge, J. M. (2021). Parents' and caregivers' perceptions of the school climate: Development and validation of the Parent and Caregiver Survey (PaCS)'. *Learning Environments Research*, *24*(1), 23–41. https://doi.org/10.1007/s10984-020-09308-z.

Allen, B. (2019, September 29). *The ungameable game*. Becky Allen. https://rebeccaallen.co.uk/2019/09/29/the-ungameable-game/

Allen, R. (2015). *We cannot compare the effectiveness of schools with different types of intakes*. https://ffteducationdatalab.org.uk/2015/05/we-cannot-compare-the-effectiveness-of-schools-with-different-types-of-intakes/

Allen, R., & Burgess, S. (2011). Can school league tables help parents choose schools? *Institute for Fiscal Studies Journal.*

Allen, R., & Vignoles, A. (2007). What should an index of school segregation measure? *Oxford Review of Education*, 33(5), 643–668. https://doi.org/10.1080/03054980701366306

Altrichter, H., & Kemethofer, D. (2015). Collaboration on school improvement under different educational accountability systems in two countries. *Educational Assessment, Evaluation and Accountability, 27*(3), 303–329. https://doi.org/10.1007/s11092-025-09460-1

Ambrosio, J. (2013). Changing the subject: Neoliberalism and accountability in public education. *Educational Studies, 49*(4), 316–333.

Amoore, L. (2011). Data derivatives: On the performative powers of risk algorithms. *Theory, Culture & Society, 28*(6), 30-52.

Amrein, A. L., & Berliner, D. C. (2002). High-stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives, 10*(18).

Amrein-Beardsley, A. (2014). *Rethinking value-added models in education: Critical perspectives on tests and assessment-based accountability*. Routledge.

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society, 20*(3), 973–989. https://doi.org/10.1177/1461444816676645

Angus, L. (1993). The sociology of school effectiveness. *British Journal of Sociology of Education, 14*(3), 333-345. https://doi.org/10.1080/0142569930140309

Anning, A., & Ball, M. (2008). *Improving services for young children: from Sure Start to children's centres* (1st ed.). SAGE.

Apple, M. W. (2001). *Educating the "right" way: Markets, standards, God, and inequality*. Routledge.

Apple, M. W. (2006). Understanding and Interrupting Neoliberalism and Neoconservatism in Education. *Pedagogies: An International Journal, 1*(1), 21–26. https://doi.org/10.1207/s15544818ped0101_4

Arrow, K. J., & Debreu, G. (1954). Existence of an Equilibrium for a Competitive Economy. *Econometrica, 22*(3), 265–290.

Association of School and College. (2025, February 24). What are "high and rising standards" and how do we achieve them? *SecEd.* https://www.sec-ed.co.uk/content/blogs/what-are-high-and-rising-standards-how-do-we-achieve-them/

Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher*, 36(5), 258–267. https://doi.org/10.3102/0013189X07306523

Bacchi, C. (2009). *Analysing policy: What's the problem represented to be?* Pearson.

Ball, S. J. (1993). Education markets, choice and social class: The market as a class strategy in the UK and the USA. *British Journal of Sociology of Education*, 14(1), 3–19.

Ball, S. J. (1994). *Education reform: A critical and post-structural approach*. Open University Press.

Ball, S. J. (2000). Performativities and fabrications in the education economy : towards the performative society? *Australian Educational Researcher*, 27(2), 1–23. https://doi.org/10.1007/BF03219719

Ball, S. J. (2003a). *Class and strategy. In Class Strategies and the Education Market.* Routledge.

Ball, S. J. (2003b). The teacher's soul and the terrors of performativity. *Journal of Education Policy*, 18(2), 215–228. https://doi.org/10.1080/0268093022000043065

Ball, S. J. (2012*). Global education inc.: New policy networks and the neo-liberal imaginary*. Routledge.

Ball, S. J. (2016). *Education, justice and democracy: The struggle over ignorance and opportunity.* Policy Press.

Ball, S. J. (2021). *The education debate (Fourth edition.).* Policy Press.

Ball, S. J., & Olmedo, A. (2013). Care of the self, resistance and subjectivity under neoliberal governmentalities. *Critical Studies in Education, 54*(1), 85–96. https://doi.org/10.1080/17508487.2013.740678

Barbour, R. (2018). *Doing focus groups* (2nd ed.). SAGE.

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law* Review, 104(3), 671–732. https://doi.org/10.15779/Z38BG31

Barocas, S., Hardt, M., & Narayanan, A. (2019). *Fairness and machine learning*. fairmlbook.org.

Barry, N. P. (1987). *The new right*. Croom Helm.

Bartlett, W. (1992). *Quasi-markets and educational reforms: a case study*. Bristol: SAUS Publications.

Bass, B. M., & Riggio, R. E. (2006). *Transformational Leadership*. New York, NY: Psychology Press. https://doi.org/10.4324/9781410617095

Baum, S., Blom, E., & Cohn, J. (2019). U*sing Multiple Metrics to Strengthen Institutional Accountability.*

BBC. (2020, May 14). Coronavirus: Daily data briefings are 'number theatre', says statistician. *BBC News*. https://www.bbc.co.uk/news/uk-52660056

Becker, G. S. (1964). *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education, First Edition*. NBER. https://www.nber.org/books-and-chapters/human-capital-theoretical-and-empirical-analysis-special-reference-education-first-edition

Beer, D. (2016). *Metric power.* Palgrave Macmillan.

Behn, R. D. (2003). Why measure performance? *Public Administration Review*, 63(5), 586–606.

Belfield, C. R., & Levin, H. M. (2002). The Effects of Competition Between Schools on Educational Outcomes: A Review for the United States. *Review of Educational Research*, *72*(2), 279–341. https://doi.org/10.3102/00346543072002279

Bessant, C., Phillips, M., Phillips, N., & Rennick, A. (2024). *Ofsted Big Listen: Research with the public and parents/carers.*

Bevan, G., & Hood, C. (2006). What's Measured Is What Matters: Targets and Gaming in the English Public Health Care System. *Public Administration*, *84*(3), 517–538. https://doi.org/10.1111/j.1467-9299.2006.00600.x

Bhaskar, R. (1975). *A realist theory of science.* London : New York: Routledge.

Bhaskar, R. (1979). *The possibility of naturalism : a philosophical critique of the contemporary human sciences.* Brighton: Harvester Press.

Biesta, G. (2004). Education, accountability and the ethical demand: Can the democratic potential of accountability be regained? *Educational Theory*, 54(3), 233–250. https://doi.org/10.1111/j.0013-2004.2004.00017.x

Biesta, G. (2009). Good education in an age of measurement. *Educational Assessment, Evaluation and Accountability*, 21(1), 33–46. https://doi.org/10.1007/s11092-008-9064-9

Biesta, G. (2010). Pragmatism and the philosophical foundations of mixed methods research. In *Handbook of Mixed Methods in Social and Behavioral Research,* 2nd ed. Edited by Abbas Tashakkori and Charles Teddlie. Thousand Oaks: Sage, pp. 95–117.

Biesta, G. J. J. (2015). *Good Education in an Age of Measurement: Ethics, Politics, Democracy.* Routledge. https://doi.org/10.4324/9781315634319

Bloor, D. (2001). 'What is a Social Construct?', *Facta Philosophica*, 3(2), pp. 141–156. Available at: https://doi.org/10.5840/factaphil20013211.

Bloor, M., Frankland, J., Thomas, M., & Robson, K. (2001). *Focus groups in social research*. SAGE.

Blumberg, A.E., & Feigl, H. (1931). 'Logical Positivism', *The Journal of Philosophy, 28*(11), pp. 281–296. Available at: https://doi.org/10.2307/2015437.

Bollen, K. A. (1989). *Structural Equations with Latent Variables*. John Wiley & Sons, Inc. https://doi.org/10.1002/9781118619179

Bollen, K. A., & Pearl, J. (2013). Eight Myths About Causality and Structural Equation Models. In S. L. Morgan (Ed.), *Handbook of Causal Analysis for Social Research* (pp. 301–328). Springer Netherlands. https://doi.org/10.1007/978-94-007-6094-3_15

Bonnen, B. (1984). *Action at the interface : schools and industry*. The author.

Booher-Jennings, J. (2005). Below the Bubble: "Educational Triage" and the Texas Accountability System. *American Educational Research Journal*, *42*(2), 231–268. https://doi.org/10.3102/00028312042002231

Booth, T. ,& Ainscow, M. (2002). *Index for inclusion: Developing learning and participation in schools*. Centre for Studies on Inclusive Education.

Bourdieu, P. (1986). The forms of capital. In J. Richardson (Ed.), *Handbook of theory and research for the sociology of education* (pp. 241–258). Greenwood Press.

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, *9*(2), 27–40.

Bowker, G. C. (2005). *Memory practices in the sciences*. Cambridge, Mass. ; MIT.

Bowker, G. C., & Star, S. L. (1999). *Sorting things out: Classification and its consequences.* MIT Press.

boyd, d., & Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society*, *15*(5), 662–679.

Boyle, G. J. (1991). Critique of the five-factor model of personality. *Australian Psychologist*, 26(3), 259–267.

Bradbury, H., & Reason, P. (2008). *The SAGE handbook of action research.* SAGE.

Bradley, S., & Migali, G. (2012). The joint evaluation of multiple educational policies: The case of specialist schools and excellence in cities policies in Britain. *Education Economics, 20*(3), 322–342. https://doi.org/10.1080/09645292.2012.678715

Bradley, S., & Taylor, J. (2002). The Effect of the Quasi-Market on the Efficiency-Equity Trade-Off in the Secondary School Sector. *Public Economics eJournal*. https://doi.org/10.1111/1467-8586.00154

Braun, H. I. (2005*). Using student progress to evaluate teachers: A primer on value added models (Policy Information Perspective)*. Educational Testing Service. https://files.eric.ed.gov/fulltext/ED529977.pdf

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology,* 3(2), 77–101. doi: 10.1191/1478088706qp063oa.

Brenner, N., Peck, J., & Theodore, N. (2010). Variegated neoliberalization: Geographies, modalities, pathways. *Global Networks*, *10*(2), 182–222. https://doi.org/10.1111/j.1471-0374.2009.00277.x

Brighouse, H., Ladd, H. F., Loeb, S. & Swift, A. (2018). *Educational goods: Values, evidence and decision-making*. University of Chicago Press.

Brill, F., Grayson, H., Kuhn, L., & O'Donnell, S. (2018). What impact does accountability have on curriculum, standards and engagement in education? *National Foundation for Educational Research*. https://www.nfer.ac.uk/publications/what-impact-does-accountability-have-on-curriculum-standards-and-engagement-in-education/

Brown, D. (1997) 'Induction and Understanding', in D. Brown (ed.) *Hopkins' Idealism: Philosophy, Physics*, Poetry. Oxford University Press. Available at: https://doi.org/10.1093/acprof:oso/9780198183532.003.0002.

Brown, P., & Lauder, H. (1996). Education, globalization and economic development. *Journal of Education Policy, 11*(1), 1–25. https://doi.org/10.1080/0268093960110101

Brown, W. (2015). *Undoing the demos: Neoliberalism's stealth revolution*. Zone Books.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.

Brunsdon, C. (2001). A Bayesian Approach to Schools' Catchment-based Performance Modelling. *Geographical & Environmental Modelling*, 5(1), 9–22. https://doi.org/10.1080/13615930120032590

Bryk, A. S., Sebring, P. B., Allensworth, E., Luppescu, S., & Easton, J. Q. (2010). *Organizing schools for improvement: Lessons from Chicago*. University of Chicago Press.

Bryman, A. (1988). *Quantity and quality in social research*. Routledge.

Bryman, A. (2006). Integrating quantitative and qualitative research: How is it done? *Qualitative Research, 6*(1), 97–113. https://doi.org/10.1177/1468794106058877

Bryman, A. (2008) *The End of the Paradigm Wars?*

Bryman, A. (2012). *Social research methods.* Oxford: Oxford University Press.

Bryman, A. (2016). *Social research methods* (5th ed.). Oxford University Press.

Burgess, S., & Greaves, E. (2021). School Choice and Accountability. In *Oxford Research Encyclopedia of Economics and Finance*. https://doi.org/10.1093/acrefore/9780190625979.013.650

Burgess, S., & Thomson, D. (2023). Beyond the threshold: The implications for pupil achievement of reforming school performance metrics. *Economics of Education Review*, 97, 102477. https://doi.org/10.1016/j.econedurev.2023.102477

Burgess, S., Greaves, E., Vignoles, A., & Wilson, D. (2015). What Parents Want: School Preferences and School Choice. *The Economic Journal*, *125*(587), 1262–1289. https://doi.org/10.1111/ecoj.12153

Butcher, B., & Corfe, O. (2025, September 1). Find the best secondary schools near you. *The Telegraph*. https://www.telegraph.co.uk/education-and-careers/uk-secondary-school-rankings-gcse-maths-english/

Byrne, B. M. (2016). *Structural Equation Modeling With AMOS: Basic Concepts, Applications and Programming, Third Edition (3rd edn*). Routledge. https://doi.org/10.4324/9781315757421

Cabinet Office. (2008). *Excellence and fairness: Achieving world class public services*. The Stationery Office.

Cabinet Office. (2010). *The Coalition: Our programme for government (Coalition Agreement)* (pp. 7–8). Her Majesty's Government. https://sticerd.lse.ac.uk/dps/case/spcc/WP13.pdf

Capper, C.A. (2018). 'Interpretivist Epistemology', in *Organizational Theory for Equity and Diversity*. Routledge.

Carey, K., & Manwaring, R. (2011). *Growth Models and Accountability: A Recipe for Remaking ESEA*.

Carnoy, M., & Loeb, S. (2002). Does External Accountability Affect Student Outcomes? A Cross-State Analysis. *Educational Evaluation and Policy Analysis*, *24*(4), 305–331. https://doi.org/10.3102/01623737024004305

Chaffee, S. H. (1991). *Explication*. Sage Publications, Inc.

Chapman, C. (2015). From one school to many: Reflections on the impact and nature of school federations and chains in England. *Educational Management, Administration & Leadership*, *43*(1), 46–60. https://doi.org/10.1177/1741143213494883

Chiao, V. (2019). Fairness, accountability and transparency: notes on algorithmic decision-making in criminal justice. *International Journal of Law in Context*, *15*(2), 126–139. https://doi.org/10.1017/S1744552319000077

Children's Commissioner for England. (2021). *Good schools*. https://assets.childrenscommissioner.gov.uk/wpuploads/2021/01/cco-good-schools.pdf

Chitty, C. (1989). *Towards a new education system : the victory of the New Right?* London: Falmer.

Chitty, C. (2014). *Education policy in the UK* (3rd ed.). Palgrave Macmillan.

Cho, G., Hwang, H., Sarstedt, M., & Ringle, C. M. (2020). Cutoff criteria for overall model fit indexes in generalized structured component analysis. *Journal of Marketing Analytics*, *8*(4), 189–202. https://doi.org/10.1057/s41270-020-00089-1

Choi, K., Goldschmidt, P., & Yamashiro, K. (2005). Exploring Models of School Performance: From Theory to Practice. *Teachers College Record, 107*(14), 119–146. https://doi.org/10.1177/016146810510701406

Choi, T., & Seon, S. W. (2021). Target Groups on the Mainline: A Theoretical Framework of Policy Layering and Learning Disparity. *Administration & Society*, *53*(4), 595–618.

Chudowsky, N., Chudowsky, V., & Kober, N. (2019). Are we closing the achievement gaps? The role of state assessments and accountability in US education. *Educational Measurement: Issues and Practice*, 38(3), 23–34.

Clark, T. (2019). *How to do your social research project or dissertation*. Oxford: University Press.

Clark, T., Foster, L. and Bryman, A. (2019). *How to do your social research project or dissertation*. Oxford: OUP Oxford.

Clarke, J., & Newman, J. (1997). *The managerial state: Power, politics and ideology in the remaking of social welfare*. Sage.

Clifford, M., Menon, R., Gangi, T., Condon, C., & Hornung, K. (2012). *Measuring school climate for gauging principal performance: A review of the validity and reliability of publicly accessible measures* (Issue Brief). American Institutes for Research.

Coldron, J., Crawford, M., Jones, S., & Simkins, T. (2014). The restructuring of schooling in England: The responses of well-positioned headteachers. *Educational Management Administration & Leadership*, *42*(3), 387–403. https://doi.org/10.1177/1741143214521592

Coleman, J. S. (1966). *Equality Of Educational Opportunity*. https://eric.ed.gov/?id=ED012275

Colman, A. (2021). School leadership, school inspection and the micropolitics of compliance and resistance: Examining the hyper-enactment of policy in an area of deprivation. *Educational Management, Administration & Leadership*, *49*(2), 268–283. https://doi.org/10.1177/1741143219898479

Connell, R. (2013). The neoliberal cascade and education: an essay on the market agenda and its consequences. *Critical Studies in Education*, *54*(2), 99–112.

Connolly, M. (2013). *Achieving social mobility: The role of equality law (Discussion of Coalition social mobility strategy)*. University of Surrey. https://pure.port.ac.uk/ws/portalfiles/portal/2820971/CONNOLLY_2013_cright.IJDL_Achieving_social_mobility.pdf

Couldry, N., & Mejias, U. A. (2019). *The costs of connection: How data is colonizing human life and appropriating it for capitalism*. Stanford University Press.

Cox, B., & Boyson, R. (1975). *Black paper 1975: the fight for education*. London: Dent.

Cox, B., & Boyson, R. (1977). *Black Paper 1977*. Temple Smith.

Crawford, K. (2013). The anxieties of big data. *The Social Science Journal, 50*(1), 10-18.

Creemers, B., & Reezigt, G. (1997). School effectiveness and school improvement: Sustaining links. *School Effectiveness and School Improvement, 8*(4), 396-429. https://doi.org/10.1080/0924345970080402

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE.

Cullen, J. B., & Reback, R. (2006). Tinkering toward accolades: School gaming under a performance accountability system. In *Improving School Accountability* (Vol. 14, pp. 121–151). Emerald Group Publishing.

Cullinane, C. (2020). *Fairer School Admissions*. The Sutton Trust.

Cuttance, P. (1989). Differential school effectiveness. *International Journal of Educational Research, 13*(7), 769–776. https://doi.org/10.1016/0883-0355(89)90027-X

D'Ignazio, C., & Bhargava, R. (2016). *Critical data literacy: Creating a more just world with data*. Medium.

D'Ignazio, C., & Klein, L. F. (2020). *Data Feminism*. MIT Press. https://doi.org/10.7551/mitpress/11805.001.0001

Dalton, C., & Thatcher, J. (2014). What does a critical data studies look like and why do we care? *Society & Space*. Retrieved from https://societyandspace.org

Daniels, H., & Cole, T. (2010). Exclusion from school: Short-term setback or a long term of difficulties?. *European Journal of Special Needs Education*. 25. 115-130. 10.1080/08856251003658652.

Daugbjerg, C., & Swinbank, A. (2016). Three Decades of Policy Layering and Politically Sustainable Reform in the European Union's Agricultural Policy. *Governance, 29*(2), 265–280. https://doi.org/10.1111/gove.12171

Davies, B. (2009). *The Essentials of School Leadership*. 1–224.

Davies, W. (2014). *The Limits of Neoliberalism: Authority, Sovereignty and the Logic of Competition*. Sage Publications.

Davoli, M., & Entorf, H. (2018). *The PISA Shock, Socioeconomic Inequality, and School Reforms in Germany*. (140).

DCSF. (2008). *A School Report Card: consultation document*. DCSF Publications. https://dera.ioe.ac.uk/id/eprint/8621/2/DCSF-01045-2008.pdf

Denzin, N.K., & Lincoln, Y.S. (2011). *The SAGE Handbook of Qualitative Research*. SAGE.

Department for Business, Innovation & Skills & Deputy Prime Minister's Office. (2015). *2010 to 2015 government policy: Social mobility*. GOV.UK. https://www.gov.uk/government/publications/2010-to-2015-government-policy-social-mobility/2010-to-2015-government-policy-social-mobility

Department for Children, Schools and Families. (2008). *21st century schools: A world-class education for every child*. London, England: The Stationery Office.

Department for Children, Schools and Families. (2009). *Your child, your schools, our future: Building a 21st century schools system*. Norwich, England: The Stationery Office.

Department for Education & Ofsted. (2025). *Memorandum of understanding between the Department for Education and Ofsted*. GOV.UK. (https://www.gov.uk/government/publications/memorandum-of-understanding-between-dfe-and-ofsted)

Department for Education and Employment. (1996). *Learning to compete: Education and training for 14–19 year olds*. London, England: HMSO.

Department for Education and Employment. (1996). *Learning to compete: Education and training for 14–19 year olds (Cm 3486)*. Her Majesty's Stationery Office.

Department for Education and Employment. (1996). *Self-government for schools*. London, England: HMSO.

Department for Education and Employment. (1997). *Excellence in schools*. London, England: The Stationery Office.

Department for Education and Employment. (2006). *Specialist schools: An evaluation of progress (Research Report DfES/RR)*. Department for Education and Skills. https://dera.ioe.ac.uk/id/eprint/4531/1/An%20evaluation%20of%20progress%20%28PDF%20format%29.pdf

Department for Education and Skills. (2001). *Schools achieving success*. London, England: The Stationery Office.

Department for Education and Skills. (2001). *Schools building on success*. London, England: The Stationery Office.

Department for Education and Skills. (2003). *14–19 opportunity and excellence*. London, England: The Stationery Office.

Department for Education and Skills. (2004). *Statistics in education: Variation in pupil progress 2003* (National Statistics Bulletin No. 02/04). London: DfES.

Department for Education and Skills. (2005). *14–19 education and skills*. London, England: The Stationery Office.

Department for Education and Skills. (2005). *Higher standards, better schools for all: More choice for parents and pupils*. London, England: The Stationery Office.

Department for Education and Skills. (2005). *Secondary education*. London, England: The Stationery Office.

Department for Education. (1992). *Choice and diversity: A new framework for schools*. London, England: HMSO.

Department for Education. (2010). *School accountability*. London, England: Department for Education.

Department for Education. (2010). *The Importance of teaching: The schools white paper 2010*. Stationery Office.

Department for Education. (2014). *Secondary accountability measures: Progress 8 school performance measure*. GOV.UK. https://www.gov.uk/government/publications/progress-8-school-performance-measure

Department for Education. (2016). *Educational excellence everywhere*. London, England: Department for Education.

Department for Education. (2020). *Key Stage 4 Similar Schools Guide and technical note for local authorities, maintained schools, academies and free schools*. GOV.UK

Department for Education. (2021). *School Admissions Code*. GOV.UK

Department for Education. (2022a). *How to understand school performance – everything you need to know*. https://educationhub.blog.gov.uk/2022/10/19/how-to-understand-school-performance-everything-you-need-to-know/

Department for Education. (2022b). *School and college accountability: 2021 to 2022 academic year*. GOV.UK. https://www.gov.uk/government/publications/school-and-college-accountability-approach-2020-to-2022/school-and-college-accountability-2021-to-2022-academic-year

Department for Education. (2023). *Key stage 4 performance, academic year 2021–22*. Explore Education Statistics.

Department for Education. (2023). *Key stage 4 performance: Methodology*. Explore Education Statistics.

Department for Education. (2023). *Understanding school and college performance measures*. GOV.UK. https://www.gov.uk/government/publications/understanding-school-and-college-performance-measures/understanding-school-and-college-performance-measures

Department for Education. (2024). *Guidance: Secondary school curriculum and qualifications*. London, United Kingdom: Gov.uk.

Department for Education. (2024). *Secondary accountability measures: 2025 guidance for maintained secondary schools, academies and free schools (including Progress 8 and Attainment 8 methodology)*. GOV.UK. https://www.gov.uk/government/publications/progress-8-school-performance-measure/secondary-accountability-measures-2025-guidance-for-maintained-secondary-schools-academies-and-free-schools GOV.UK

Department for Education. (2025a). *Secondary accountability measures. Guide for maintained secondary schools, academies and free schools*. GOV.UK.

Department for Education. (2025b). *Curriculum and assessment review: final report*. GOV.UK. https://www.gov.uk/government/publications/curriculum-and-assessment-review-final-report

Department for Education. (2025c). *Pupil absence in schools in England: 2023 to 2024*. GOV.UK. https://explore-education-statistics.service.gov.uk/find-statistics/pupil-absence-in-schools-in-england/2023-24

Department for Education. (2025d). *School accountability reform – school profiles, improvement and intervention: Government consultation response*. Retrieved 11 November 2025, from https://assets.publishing.service.gov.uk/media/68beea01c771153e08e0dd4f/school-accountability-reform-government-response.pdf

Department for Education. (2025e). *School accountability reform*. Retrieved 25 February 2025, from https://www.gov.uk/government/consultations/school-accountability-reform

Department for Education. (2025f). *Schools, pupils and their characteristics: Academic year 2024/25*. Retrieved from https://explore-education-statistics.service.gov.uk/

Department of Education and Science & Welsh Office. (1988). *National curriculum: Task Group on Assessment and Testing: A report*. Her Majesty's Stationery Office (HMSO).

Department of Education and Science. (1985). *Better schools*. London, England: Her Majesty's Stationery Office.

Deppeler, J., & Ainscow, M. (2016). Using inquiry-based approaches for equitable school improvement. *School Effectiveness and School Improvement, 27*(1), 1-6. https://doi.org/10.1080/09243453.2015.1026671

Desrosières, A. (1998*). The politics of large numbers: A history of statistical reasoning*. Harvard University Press.

Dexter, L. A. (2006). *Elite and specialized interviewing*. ECPR Press.

Dhawka, P., & Dasgupta, S. (2023). The social construction of visualizations: Practitioner challenges and experiences of visualizing race and gender demographic data. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–16.

Diakopoulos, N., & Koliska, M. (2017). Algorithmic Transparency in the News Media. *Digital Journalism, 5*(7), 809–828. https://doi-org.sheffield.idm.oclc.org/10.1080/21670811.2016.1208053

Diamantopoulos, A., Reifler, P., & Zeugner-Roth, K. (2008). Advancing formative measurement models. *Journal of Business Research*.

Didau, D. (2015, April 29). *Trust, accountability and why we need them both*. David Didau. https://learningspy.co.uk/leadership/trust-vs-accountability/

Dijck, J. van. (2014). Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & Society*, *12*(2), 197–208. https://doi.org/10.24908/ss.v12i2.4776

Dillon, D. R., O'Brien, D. G., & Heilman, E. E. (2000). Literacy Research in the Next Millennium: From Paradigms to Pragmatism and Practicality. *Reading Research Quarterly*, *35*(1), 10–26. http://www.jstor.org.sheffield.idm.oclc.org/stable/748284

Dominitz, J., & Sherman, R. P. (2006). Identification and estimation of bounds on school performance measures: A nonparametric analysis of a mixture model with verification. *Journal of Applied Econometrics, 21*(8), 1295–1326. https://doi.org/10.1002/jae.912

Drucker, J. (2011). Humanities approaches to graphical display. *Digital Humanities Quarterly, 5*(1).

Dukes, E. (2025, September 21). *10 worst secondary schools in Wirral revealed in 2025 league table*. LiverpoolWorld. https://www.liverpoolworld.uk/education/secondary-schools-league-table-wirral-2025-5325597

Dulude, E., & Milley, P. (2020). Institutional complexity and multiple accountability tensions: A conceptual framework for analyzing school leaders' interpretation of competing demands. *Policy Futures in Education, 19*(1), 84-96. https://doi-org.sheffield.idm.oclc.org/10.1177/1478210320940134 (Original work published 2021)

Dunlop, C. A., & Radaelli, C. M. (2018). The lessons of policy learning: Types, triggers, hindrances and pathologies. *Policy & Politics*, *46*(2), 255–272. https://doi.org/10.1332/030557318X15230059735521

Dutaut, J. L. (2025, January 18). *Can education learn to stop worrying and love collaboration?* https://schoolsweek.co.uk/can-education-learn-to-stop-worrying-and-love-collaboration/

Dyson, A., & Todd, L. (2010). *Dealing with diversity: Mapping the terrain of inclusive education*. Routledge.

Eagleton-Pierce, M. (2016). *Neoliberalism : the key concepts*. London: Routledge.

Education Policy Institute. (2018). *Faith schools, pupil characteristics and outcomes*. https://epi.org.uk/wp-content/uploads/2018/01/Pupil_characteristics_and_performance_at_faith_schools.pdf

Education Policy Institute. (2022). *The features of effective school groups: Measuring pupil inclusion at school-group leve*l. Education Policy Institute.

Education Policy Institute. (2023). *Annual Report 2023*. https://epi.org.uk/publications-and-research/annual-report-2023/

Education Policy Institute. (2025). *Reforming accountability: widening the debate on performance measures in English schools*. Education Policy Institute. https://epi.org.uk/wp-content/uploads/2025/01/Reforming_accountability_vFINAL.pdf

Ehren, M. C. M., & Shackleton, N. (2016). Risk-based school inspections: Impact of targeted inspection approaches on Dutch secondary schools. *Educational Assessment, Evaluation and Accountability*, *28*(4), 299–321. https://doi.org/10.1007/s11092-016-9242-0

Elmore, R. F. (2004). School Reform from the Inside Out: Policy, Practice, and Performance. In *Harvard Education Press*. Harvard Education Press.

Elsbach, K. D. (1992). Acquiring Organizational Legitimacy through Illegitimate Actions: A Marriage of Institutional and Impression Management Theory. *Academy of Management Journal, 35*, 699-738.

English, F. W. (Ed.). (2005). *Value added indicators*. In Encyclopedia of Educational Leadership and Administration. Sage.

Erichsen, K., & Reynolds, J. (2019). Public school accountability, workplace culture and teacher morale. *Social Science Research, 85*, 102347.

Espeland, W. N., & Sauder, M. (2007). Rankings and reactivity: How public measures recreate social worlds. *American Journal of Sociology*, 113(1), 1–40. https://doi.org/10.1086/517897

Espeland, W. N., & Stevens, M. L. (1998). Commensuration as a social process. *Annual Review of Sociology, 24*, 313–343. https://doi.org/10.1146/annurev.soc.24.1.313

Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. *American Journal of Theoretical and Applied Statistics, 5*(1), 1–4.

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.

European Centre for Disease Prevention and Control. (n.d.). *COVID-19 data dashboard*. https://www.ecdc.europa.eu/en/covid-19-pandemic

European Commission. (2019). *Ethics guidelines for trustworthy AI*. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

Everitt, B., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster analysis* (5th ed.). Wiley.

Eyles, A., & Machin, S. (2015). The introduction of the English school league tables and parental choice. *Journal of the European Economic Association, 13*(2), 1–28.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272–299.

Ferrão, M. E., & Couto, A. P. (2014). The Use of a School Value-Added Model for Educational Improvement: A Case Study from the Portuguese Primary Education System. *School Effectiveness and School Improvement*, *25*(1), 174–190. https://doi.org/10.1080/09243453.2013.785436

FFT Education Datalab. (2018). *Value added measures in performance tables: A recap of the main issues for secondary schools*. https://ffteducationdatalab.org.uk/2018/05/value-added-measures-in-performance-tables-a-recap-of-the-main-issues-for-secondary-schools/

FFT Education Datalab. (2025). *Schools Like Yours*. Retrieved 13 November 2025, from https://schoolslikeyours.ffteducationdatalab.org.uk/

Figlio, D. N., & Getzler, L. S. (2006). Accountability, Ability and Disability: Gaming the System? In T. J. Gronberg & D. W. Jansen (Eds.), *Improving School Accountability* (Vol. 14, p. 0). Emerald Group Publishing Limited. https://doi.org/10.1016/S0278-0984(06)14002-X

Filer, C., & Sinozich, S. (2021). *Valuing diversity in public schools*. 32.

Fischer, F. (2003). Reframing Public Policy: Discursive Politics and Deliberative Practices. *Reframing Public Policy: Discursive Politics and Deliberative Practices*. https://doi.org/10.1023/B:OLIC.0000035505.23131.80

Fitzgerald, S. B., Johhston, L., R., S., R., & Voss, C. (1991). *Performance Measurement in Service Businesses*. CIMA.

Flick, U. (2018). *Doing triangulation and mixed methods*. SAGE.

Flockton, L., Crooks, T., & McDonald, B. (2017). Assessment and accountability in schools: Impacts on teaching and learning. *Assessment in Education: Principles, Policy & Practice, 24*(3), 263–285.

Flude, M. and Hammer, M. (1990). *The Education Reform Act, 1988 : its origins and implications*. London: Falmer.

Fox, J.P. (2004). Modelling response error in school effectiveness research. *Statistica Neerlandica, 58*(2), 138–160. https://doi.org/10.1046/j.0039-0402.2003.00253.x

Fraser, N. (2008). *Scales of justice: Reimagining political space in a globalizing world*. Polity Press.

Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowing*. Oxford University Press.

Friedman, I. A. (1991). High and Low-Burnout Schools: School Culture Aspects of Teacher Burnout. *The Journal of Educational Research, 84*(6), 325–333. https://doi.org/10.1080/00220671.1991.9941813

Friedman, M. (1953). *Essays in Positive Economics*. University of Chicago Press.

Gelman, A., & Hill, J. (2007). Data analysis using regression and multilevel/hierarchical models. Cambridge University Press.

Gentles, S., Charles, C., Ploeg, J., & McKibbon, K. (2015). Sampling in Qualitative Research: Insights from an Overview of the Methods Literature. *The Qualitative Report*. https://www.semanticscholar.org/paper/Sampling-in-Qualitative-Research%3A-Insights-from-an-Gentles-Charles/5ee8002a529d8f4601a9c2ca7c70c739686eda2a

Gewirtz, S., & Cribb, A. (2002). Plural conceptions of social justice: Implications for policy sociology. *Journal of Education Policy, 17*(5), 499–509.

Gewirtz, S., Ball, S. J., & Bowe, R. (1997). Markets, Choice and Equity in Education. *British Journal of Educational Studies*, *45*(1), 114–116.

Ghuman, J. (2018). *Looking into the London Challenge*. The Sutton Trust. https://www.suttontrust.com/news-opinion/all-news-opinion/appg-the-london-challenge/

Gill, T. (2024). *How have students and schools performed on the Progress 8 performance measure?* Cambridge University Repository.

Gillborn, D. (2010a). The colour of numbers: Surveys, statistics and deficit-thinking about race and class. Journal of Education Policy, 25(2), 253–276. https://doi.org/10.1080/02680930903460740

Gillborn, D. (2010b). The 'soft bigotry of low expectations'? Racism and the real problems of school reform. *International Studies in Sociology of Education, 20*(3), 283–299.

Gillborn, D. (2018). *Head start: Ethnic inequalities and the mismeasure of intelligence*. Routledge.

Gillborn, D., & Youdell, D. (2000). *Rationing education: Policy, practice, reform and equity*. Open University Press.

Giovannini, J. (2023). *Children, Families, Schools and Communities*. ROTEL.

Gitelman, L. (Ed.). (2013). *'Raw data' is an oxymoron*. MIT Press.

GL Assessment. (2019). *Closing the gap: How a narrow curriculum disadvantages the disadvantaged. (Research report)*. https://www.teachertoolkit.co.uk/wp-content/uploads/2019/02/GL2370_Closing-the-Gap-report_FINAL_3-January-2019.pdf

GL Assessment. (2025). *PASS*. https://www.gl-assessment.co.uk/products/pass/

Glazzard, J., & Tate, A. (2025). Reclaiming Initial Teacher Education in universities: Moving beyond a technicist model. *Policy Futures in Education, 23*(2), 404–409. https://doi.org/10.1177/14782103241279582

Glennerster, H., Power, A., & Travers, T. (1991). A new era for social policy: a new enlightenment or a new leviathan? Journal of Social Policy, 20 (3). 389-414. ISSN 0047-2794

Goldhaber, D. (2002). The mystery of good teaching. *Education Next.* 2. 50-55.

Goldstein, H. (1987). *Multilevel models in educational and social research*. London, England: Griffin.

Goldstein, H. (1997). Methods in school effectiveness research. *School Effectiveness and School Improvement*, 8(4), 369–395.

Goldstein, H. (2001). *Using pupil performance data for judging schools and teachers: Scope and limitations*. British Educational Research Journal, 27(4), 433–442. https://doi.org/10.1080/01411920120071487

Goldstein, H. (2017a). Measurement and evaluation issues with PISA. In L. Volante (Ed.), *The PISA effect on global educational governance* (pp. 49–58). Routledge. https://doi.org/10.4324/9781315440521-4

Goldstein, H. (2017b). Using value-added models for teacher evaluation and school improvement. In T. Leuders et al. (Eds.), *Value-added modeling and teacher education* (pp. 15–32). Springer.

Goldstein, H., & Leckie, G. (2008). School League Tables: What can they Really Tell Us? *Significance, 5*, 67–69. https://doi.org/10.1111/j.1740-9713.2008.00289.x

Goldstein, H., & Leckie, G. (2011). Limitations of school league tables to inform choice and accountability. *Journal of the Royal Statistical Society: Series A*. https://doi.org/10.1111/j.1467-985X.2010.00688.x

Goldstein, H., & Rasbash, J. (1994). Efficient analysis of mixed hierarchical and cross-classified random structures using a multilevel model. *Journal of the Royal Statistical Society. Series A (Statistics in Society), 157*(3), 397–412. https://doi.org/10.2307/2983327

Goldstein, H., & Spiegelhalter, D. J. (1996). League Tables and Their Limitations: Statistical Issues in Comparisons of Institutional Performance. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *159*(3), 385–409. https://doi.org/10.2307/2983325

Goldstein, H., & Spiegelhalter, D. J. (1996). League Tables and Their Limitations: Statistical Issues in Comparisons of Institutional Performance. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *159*(3), 385–409. https://doi.org/10.2307/2983325

Goldstein, H., & Thomas, S. (2008). Using examination results as indicators of school and college performance. *Journal of the Royal Statistical Society: Series A, 171*(1), 313–334.

Gorard, S. (1997). Market Forces, Choice and Diversity in Education: The Early Impact. *Sociological Research Online, 2(*3), 137-146. https://doi-org.sheffield.idm.oclc.org/10.5153/sro.111

Gorard, S. (2010). Serious doubts about school effectiveness. *British Educational Research Journal, 36*(5), 745–766. https://doi.org/10.1080/01411920903144251

Gorard, S. (2018). *Education policy: Evidence of equity and effectiveness*. Policy Press.

Gorard, S., & See, B. H. (2013, May 1). *Do parental involvement interventions increase attainment? A review of the evidence.* http://www.nuffieldfoundation.org/sites/default/files/files/Do_parental_involvement_interventions_increase_attainment1.pdf

GOV.UK. (2003). *Every Child Matters.*

GOV.UK. (2013, November 22). *Michael Gove: The civil rights struggle of our time*. https://www.gov.uk/government/speeches/the-civil-rights-struggle-of-our-time

GOV.UK. (2024, 28 November*). National funding formula for schools and high needs: 2025 to 2026*. Retrieved 25 September 2025, from https://www.gov.uk/government/publications/national-funding-formula-for-schools-and-high-needs-2025-to-2026

GOV.UK. (2024, December 5). *Plan for Change*. https://www.gov.uk/government/publications/plan-for-change

GOV.UK. (2024, November 7). *Bridget Phillipson's Speech to the Confederation of School Trusts*. https://www.gov.uk/government/speeches/bridget-phillipsons-speech-to-the-confederation-of-school-trusts

GOV.UK. (2025, 29 September). *Parent, pupil and learner voice: March 2025*. Retrieved 11 November 2025, from https://www.gov.uk/government/publications/parent-pupil-and-learner-voice-omnibus-surveys-for-2024-to-2025/parent-pupil-and-learner-voice-march-2025

GOV.UK. (2025, August 13). *Giving every child the best start in life*. https://www.gov.uk/government/publications/giving-every-child-the-best-start-in-life

GOV.UK. (2025, February 3). *Education Secretary speech on new era of school standards*. https://www.gov.uk/government/speeches/education-secretary-speech-on-new-era-of-school-standards

GOV.UK. (2025, October 16). *Secondary accountability measures: 2025 guidance for maintained secondary schools, academies and free schools (Progress 8 school performance measure)*. GOV.UK. https://www.gov.uk/government/publications/progress-8-school-performance-measure/secondary-accountability-measures-2025-guidance-for-maintained-secondary-schools-academies-and-free-schools

Government Digital Service. (2023). *Algorithmic Transparency Recording Standard (ATRS) hub*. GOV.UK. https://www.gov.uk/government/collections/algorithmic-transparency-recording-standard-hub

Grainger, S., Mao, F., & Buytaert, W. (2016). Environmental data visualisation for non-scientific contexts: Literature review and design framework. *Environmental Modelling and Software*, *85*, 299–318. https://doi.org/10.1016/j.envsoft.2016.09.004

Gray, J., & Jesson, D. (1990). The negotiation and construction of performance indicators: Some principles, proposals and problems. *Evaluation & Research in Education*, *4*(2), 93–108. https://doi-org.sheffield.idm.oclc.org/10.1080/09500799009533283

Greany, T., & Higham, R. (2018). *Hierarchy, Markets and Networks*. UCL Institute of Education Press.

Greene, J. C. (2007). *Mixed methods in social inquiry*. Jossey-Bass.

Grek, S. (2009). Governing by numbers: The PISA 'effect' in Europe. *Journal of Education Policy, 24*(1), 23–37. https://doi.org/10.1080/02680930802412669

Griffiths, D. (2020). The ethical issues of learning analytics in their historical context. In D. Burgos & C. Hernández-Leo (Eds.), *Radical Solutions and Open Science* (pp. 39–55). Springer. https://doi.org/10.1007/978-981-15-4276-3_3

Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 105–117). Sage.

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM Computing Surveys, 51*(5), 1–42. https://doi.org/10.1145/3236009

Gunter, H. M., Grimaldi, E., Hall, D., & Serpieri, R. (2016). *New public management and the reform of education: European lessons for policy and practice.* Routledge.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage.

Hallinger, P., & Huber, S. (2012). School leadership that makes a difference: International perspectives. *School Effectiveness and School Improvement, 23*(4), 359-367. https://doi.org/10.1080/09243453.2012.681508

Hamilton, L. S., Stecher, B. M., & Klein, S. P. (2002). Making Sense of Test-Based Accountability in Education. RAND Education.

Hanushek, E. A., & Raymond, M. E. (2004). Does school accountability lead to improved student performance? *Journal of the European Economic Association, 2*(2–3), 406–415.

Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (2002). The effect of special education on the academic achievement of students with disabilities. *Journal of Policy Analysis and Management, 21*(4), 543–563.

Hargreaves, A. (2003). *Teaching in the knowledge society: Education in the age of insecurity*. Open University Press.

Hargreaves, A., & Fullan, M. (2012). *Professional capital: Transforming teaching in every school*. Teachers College Press.

Harris, D. (2008). Toward Policy-Relevant Benchmarks for Interpreting Effect Sizes: Combining Effects With Costs. *Educational Evaluation and Policy Analysis - EDUC EVAL POLICY ANAL*, *31*. https://doi.org/10.3102/0162373708327524

Harris, K. R. (2018). Educational psychology: A future retrospective. *Journal of Educational Psychology, 110*(2), 163–173. https://doi.org/10.1037/edu0000267

Harvey, D. (2007). *A brief history of neoliberalism*. Oxford University Press.

Hatry, H. P. (2006). *Performance Measurement: Getting Results* (2nd ed.). Urban Institute Press.

Hay, C. (1999). The political economy of New Labour. *Political Studies, 47*(2), 333-356.

Heywood, A. (2012). *Political ideologies: An introduction* (5th ed.). Palgrave Macmillan.

Hill, M., & Hupe, P. (2014). *Implementing public policy* (3rd ed.). Sage.

Hill, R. L., Kennedy, H., & Gerrard, Y. (2016). Visualizing Junk: Big Data Visualizations and the Need for Feminist Data Studies. *Journal of Communication Inquiry*, *40*(4), 331–350. https://doi.org/10.1177/0196859916666041

HM Government. (2007). *The children's plan: Building brighter futures*. Norwich, England: The Stationery Office.

HM Government. (2008). *Excellence and fairness: Achieving world class public services*. London, England: The Stationery Office.

HM Government. (2009). *New opportunities: Fair chances for the future*. London, England: The Stationery Office.

HM Treasury. (2003). *Every child matters*. London, England: The Stationery Office.

HM Treasury. (2022). *Managing public money (July 2022 edition).* HM Treasury. https://www.gov.uk/government/publications/managing-public-money)

Hochschild, A. R. (1983). *The managed heart: Commercialization of human feeling.* University of California Press.

Hoff, D. J. (2007). "Growth Models" Gaining in Accountability Debate'. *Education Week*. https://www.edweek.org/policy-politics/growth-models-gaining-in-accountability-debate/2007/12

Honig, M. I. (2006). *New directions in education policy implementation*. SUNY Press.

Hood, C. (1991). A public management for all seasons? *Public Administration, 69*(1), 3–19.

Hood, C. (1995). 'The 'new public management' in the 1980s: Variations on a theme'. *Accounting, Organizations and Society, Volume 20*, Issues 2-3, Pages 93-109.

Hopmann, S. T. (2008). No child, no school, no state left behind: Schooling in the age of accountability. *Journal of Curriculum Studies, 40*(4), 417–456. https://doi.org/10.1080/00220270801989818

House of Commons Children, Schools and Families Committee. (2008). *Testing and assessment: Third report of session 2007–08 (HC 169-I; Vol. 1)*. The House of Commons. https://publications.parliament.uk/pa/cm200708/cmselect/cmchilsch/169/169.pdf

House of Commons Children, Schools and Families Committee. (2010). *School accountability* (HC 88-I). The Stationery Office.

House of Commons Education and Skills Committee. (2005). *Secondary education: Fifth report of session 2004–05* (HC 86-I). The Stationery Office.

House of Commons Education Committee. (2017). *Ofsted's work with schools (HC 112, Session 2016–2017)*. The Stationery Office. (https://committees.parliament.uk/publications/2666/documents/26658/default/)

Hout, M., & Elliott, S. W. (2011). *Incentives and test-based accountability in education*. Russell Sage Foundation.

Howland, G. (2015). Structural reform: The experience of ten schools driving the development of an all-age hard federation across a market town in northern England. *Management in Education, 29*(1), 25–30. https://doi.org/10.1177/0892020614559238

Howlett, M. (2011). *Designing public policies: Principles and instruments*. Routledge.

Hu, L. and Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), pp. 1–55. Available at: https://doi.org/10.1080/10705519909540118.

Hursh, D. (2000). Neoliberalism and the control of teachers, students and learning: The rise of standards, standardization and accountability. *Cultural Logic: A Journal of Marxist Theory & Practice, 7.*

Hursh, D. (2007). Assessing No Child Left Behind and the Rise of Neoliberal Education Policies. *American Educational Research Journal, 44*(3), 493–518.

Hutchings, M. (2015). *Exam factories? The impact of accountability measures on children and young people*. National Union of Teachers.

Hutchinson, J. (2018). *Educational outcomes of children with English as an additional language*. Education Policy Institute.

Iliadis, A., & Russo, F. (2016). Critical data studies: An introduction. *Big Data & Society, 3*(2), 2053951716674238. https://doi.org/10.1177/2053951716674238

Ingersoll, R. M. (2001). Teacher turnover and teacher shortages: An organizational analysis. *American Educational Research Journal, 38*(3), 499–534. https://doi.org/10.3102/00028312038003499

Innes, J. E. (1990). *Knowledge and Public Policy: The Search for Meaningful Indicators*. Transaction Publishers.

Inuwa-Dutse, I. (2023). *FATE in AI: Towards algorithmic inclusivity and accessibility*. https://doi.org/10.48550/arXiv.2301.01590. arXiv preprint

Jackson, L. J., Trebitz, A. S., & Cottingham, K. L. (2000). An introduction to the practice of ecological modeling. *BioScience*, *50*, 694–706.

Jakeman, A. J., Letcher, R. A., & Norton, J. P. (2006). Ten iterative steps in development and evaluation of environmental models. *Environmental Modelling and Software*, *21*, 602–614.

Jencks, C. (1975). *Inequality : a reassessment of the effect of family and schooling in America*. Harmondsworth: Penguin.

Jennings, J. L., & Corcoran, S. P. (2009). Reflections on growth models for school accountability. *Phi Delta Kappan, 90*(9), 635.

Jerrim, J., & Sims, S. (2022). School accountability and teacher stress: International evidence from the OECD TALIS study. *Educational Assessment, Evaluation and Accountability, 34*(1), 5–32. https://doi.org/10.1007/s11092-021-09360-0

Jessop, B. (2002). *The future of the capitalist state*. Polity Press.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature *Machine Intelligence, 1*(9), 389–398. https://doi.org/10.1038/s42256-019-0088-2 ScienceDirect

Jones, K., & Bird, K. (2000). 'Partnership' as Strategy: Public-private relations in Education Action Zones. *British Educational Research Journal, 26*(4), 491–506. https://doi.org/10.1080/713651572

Jöreskog, K., & Wallentin, F. (1996). Nonlinear structural equation models: The Kenny-Judd model with interaction effects. In *Advanced structural equation modeling: Issues and techniques. Lawrence Erlbaum Associates*.

Kaiser, H. F. (1960). The Application of Electronic Computers to Factor Analysis. *Educational and Psychological Measurement*, 20, 141-151. http://dx.doi.org/10.1177/001316446002000116

Kamanzi, P. C. (2022). School market and the democratization of education: One step forward, two steps back. The case of the Canadian Province of Quebec'. *International Review of Sociology*, *32*(1), 107–127. https://doi.org/10.1080/03906701.2021.2015933.

Kane, M. T. (2006). Validation. In *Educational Measurement* (4th ed.).

Kane, T. J., & Staiger, D. O. (2002). The promise and pitfalls of using imprecise school accountability measures. *Journal of Economic Perspectives, 16*(4), 91–114. https://doi.org/10.1257/089533002320951102

Kaufman, L., & Rousseeuw, P. J. (Eds.). (1990). *Finding Groups in Data*. John Wiley & Sons, Inc. https://doi.org/10.1002/9780470316801

Kelly, A., & Downey, C. (2010). Value-added measures for schools in England: Looking inside the 'black box' of complex metrics. *Educational Assessment, Evaluation and Accountability*, *22*, 181–198. https://doi.org/10.1007/s11092-010-9100-4

Kelly, G. (2011). *Editorial - Fuelled by optimism, not blame, London Challenge is testament to what can be done*: News. The Times Educational Supplement Cymru, 348, 2.

Kelman, S., & Friedman, J. N. (2009). Performance improvement and performance dysfunction: An empirical examination of distortionary impacts of the emergency room wait-time target in the English NHS. *Journal of Public Administration Research and Theory, 19*(4), 917–946.

Kennedy, H. (2022). Original Seeing Data Research. *Seeing Data*. Retrieved 16 August 2022, from https://seeingdata.org/original-seeing-data-research/

Kennedy, H., & Hill, R. L. (2018). The Feeling of Numbers: Emotions in Everyday Engagements with Data and Their Visualisation. *Sociology*, *52*(4), 830–848. https://doi.org/10.1177/0038038516674675

Kidson, M., & Norris, E. (2014). *Implementing the London Challenge*. Institute of Government and Joseph Rowntree Foundation.

Kitchin, R. (2014a). Big data, new epistemologies and paradigm shifts. *Big Data & Society, 1*(1), 1–12.

Kitchin, R. (2014b). *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences*. SAGE Publications Ltd. https://doi.org/10.4135/9781473909472

Kitchin, R. (2022). *The data revolution: A critical analysis of big data, open data and data infrastructures* (Second). SAGE.

Kitchin, R., & Dodge, M. (2011). *Code/space: Software and everyday life*. MIT Press.

Kitchin, R., & Lauriault, T. P. (2014). Towards critical data studies: Charting and unpacking data assemblages and their work. *The Programmable City Working Paper, 2*. Retrieved from https://mural.maynoothuniversity.ie/5683/1/KitchinLauriault_CriticalDataStudies_ProgrammableCity_WorkingPaper2_SSRN-id2474112.pdf

Kitzinger, J. (1995). Qualitative research: Introducing focus groups. *BMJ, 311*, 299–302.

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.

Knafo, S., Dutta, S.J., Lane, R., & Wyn-Jones, S. (2019).The managerial lineages of neoliberalism. *New Political Economy, Vol. 24.* No. 2, pp. 235-251.

Koretz, D. (2017). *The testing charade: Pretending to make schools better*. Chicago: University of Chicago Press.

Kroll, J. A., Huey, J., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2017). Accountable algorithms. *University of Pennsylvania Law Review, 165*(3), 633–706.

Krueger, R. A., & Casey, M. A. (2015). *Focus groups: A practical guide for applied research* (5th ed.). Sage.

Kuhn, T. (1962). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Kyndt, E. and Onghena, P. (2014). The integration of work and learning: Tackling the complexity by means of structural equation modeling.', in *Professional and Practice-based Learning*, pp. 255–291. Available at: https://doi.org/10.1007/978-94-007-7012-6_14.

Lavy, V. (2015). Do differences in schools' instruction time explain international achievement gaps? Evidence from developed and developing countries*. The Economic Journal, 125*(588), F397–F424. https://doi.org/10.1111/ecoj.12233

Lawton, D. (1989). *The Education Reform Act : choice and control*. London: Hodder & Stoughton.

Lawton, D. (1992). *Education and politics in the 1990s : conflict or consensus?* London: Falmer.

Le Grand, J. (2007). *The other invisible hand: Delivering public services through choice and competition*. Princeton University Press.

Leckie, G. (2008). *Modelling the Effects of Pupil Mobility and Neighbourhood on School Differences in Educational Achievement*. CMPO, University of Bristol.

Leckie, G. (2009). The complexity of school and neighbourhood effects and movements of pupils on school differences in models of educational achievement. *Journal of the Royal Statistical Society: Series A, 172*(3), 537–554. https://doi.org/10.1111/j.1467-985X.2008.00577.x

Leckie, G., & Goldstein, H. (2009). The limitations of using school league tables to inform school choice. *Journal of the Royal Statistical Society: Series A, 180*(2), 475–492.

Leckie, G., & Goldstein, H. (2017a). The evolution of school league tables in England 1992–2016: 'Contextual value-added', 'expected progress' and 'progress 8'. *British Educational Research Journal, 43*(2), 193–212. https://doi.org/10.1002/berj.3264

Leckie, G., & Goldstein, H. (2017b). The limitations of using school league tables to inform school choice. *Journal of the Royal Statistical Society: Series A, 180*(2), 475–492.

Leckie, G., & Goldstein, H. (2018a). *Monitoring School Performance Using Value-Added and Value-Table Models: Lessons from the UK* (pp. 251–260). https://doi.org/10.1007/978-3-319-73906-9_23

Leckie, G., & Goldstein, H. (2018b). *Should we adjust for pupil background in school value-added models? A study of Progress 8 and school accountability in England*. arXiv. https://arxiv.org/abs/1811.09240

Leckie, G., & Goldstein, H. (2019a). Should we adjust for pupil background in school value-added models? A study of Progress 8 and school accountability in England. *Journal of the Royal Statistical Society: Series A*. Advance online publication. https://doi.org/10.1111/rssa.12446

Leckie, G., & Goldstein, H. (2019b). The evolution of school league tables in England, 1992–2016: contextual value-added, expected progress and Progress 8. *British Educational Research Journal, 45*(4), 905–926.

Leckie, G., & Goldstein, H. (2019c). The importance of adjusting for pupil background in school value-added models: A study of Progress 8 and school accountability in England. *British Educational Research Journal, 45*(3), 518–537. https://doi.org/10.1002/berj.3511

Leckie, G., & Prior, L. (2022). A comparison of value-added models for school accountability. *School Effectiveness And School Improvement, 33*(3), 431–455. https://doi.org/10.1080/09243453.2022.2032763

Lehtiniemi, T., & Ruckenstein, M. (2019). The social imaginaries of data activism. *New Media & Society, 21*(4), 864–881.

Lessof, C., Cook, L., Page, C., Sammons, P., Smees, R., Toth, K., & Thomas, S. (2018a). *Understanding progress 8: How progress is measured in secondary schools*. Department for Education.

Lessof, C., Ross, A., Brind, R., Harding, C., & Bell, E. (2018b). *Understanding KS4 attainment and progress: Evidence from LSYPE2.* Department for Education.

Levačić, R. (2000). *The Impact of Quasi-Markets and Performance Regulation on Socially Disadvantaged Schools*. https://www.academia.edu/57353367/The_Impact_of_Quasi_Markets_and_Performance_Regulation_on_Socially_Disadvantaged_Schools_Working_Paper

Levačić, R. (2004). Competition and the performance of english secondary schools: Further evidence. *Education Economics*, *12*(2), 177–193. https://doi.org/10.1080/0964529042000239186

Levačić, R., & Woods, P. A. (2013). Raising School Performance in the League Tables (Part 1): Disentangling the effects of social disadvantage'. *British Educational Research Journal*, *28*(2), 207–226. https://doi.org/10.1080/01411920120122158.

Lewis-Beck, M., Bryman, A., & Liao, T.F. (2004). *The SAGE Encyclopedia of Social Science Research Methods*. Sage Publications, Inc. Available at: https://doi.org/10.4135/9781412950589.

Lighthouse. (2022). *Lighthouse School Analysis*. https://www.lighthousedata.org/all-projects

Lincoln, Y., & Guba, E. G. (1985). *Naturalistic inquiry*. Newbury Park, CA: Sage.

Lingard, B. (2011). Policy as numbers. *Journal of Education Policy, 26*(4), 379–393.

Lingard, B., & Sellar, S. (2013). Globalization, edu-business and network governance: The policy sociology of PISA. *Journal of Education Policy, 28*(5), 710–730. https://doi.org/10.1080/02680939.2013.776828

Lingard, B., Sellar, S., & Lewis, S. (2017). Accountabilities in schools and school systems. *Oxford Review of Education, 43*(1), 1–16.

Lipman, P. (2011). *The new political economy of urban education*. Routledge.

LocalEd2025. (2025). *Pilot 3: Towards a new model of accountability. Educating For the Future.* Retrieved 11 November 2025, from https://localed2025.org.uk/pilots/pilot-3/

Loeb, S., & Figlio, D. (2011). School accountability. In E. A. Hanushek, S. Machin & L. Woessmann (Eds.), *Handbook of the economics of education (Vol. 3, pp. 383–421).* Elsevier.

Loehle, C. (2011). Complexity and the problem of ill-posed questions in ecology. *Ecological Complexity*, *8*(1), 60–67. https://doi.org/10.1016/j.ecocom.2010.11.004

Long, R. (2023). *School inspections in England: Ofsted (House of Commons Library Briefing Paper CBP-07091)*. House of Commons Library. (https://researchbriefings.files.parliament.uk/documents/CBP-7091/CBP-7091.pdf)

López-Torres, L., Johnes, J., Elliott, C., & Polo, C. (2021). The effects of competition and collaboration on efficiency in the UK independent school sector. *Economic Modelling*, *96*, 40–53. https://doi.org/10.1016/j.econmod.2020.12.020

Loreman, T., Forlin, C., & Sharma, U. (2014). Measuring indicators of inclusive education: A systematic review of the literature. In J. K. Bakken (Ed.), *Advances in learning and behavioral disabilities: Vol. 3* (pp. 169–200). Emerald Group Publishing Limited.

Lubienski, C. (2007). Marketing Schools: Consumer Goods and Competitive Incentives for Consumer Information. *Education and Urban Society, 40*(1), 118-141. https://doi.org/10.1177/0013124507303994

Lubienski, C., & Lubienski, S. T. (2006). Charter, private, public schools and academic achievement: New evidence from NAEP mathematics data. National Center for the Study of Privatization in Education, Columbia University.

Lupton, D., & Michael, M. (2017). *Depends on Who's Got the Data': Public Understanding of Personal Digital Dataveillance*. Surveillance and Society.

Lupton, R. (2005). Social justice and school improvement: Improving the quality of schooling in the poorest neighbourhoods. *British Educational Research Journal, 31*(5), 589–604. https://doi.org/10.1080/01411920500240759

Lyon, A. R., Corbin, C. M., Brown, E. C., et al. (2022). Leading the charge in the education sector: Development and validation of the School Implementation Leadership Scale (SILS). *Implementation Science, 17*, Article 48. https://doi.org/10.1186/s13012-022-01222-7

Machin, S., & Vignoles, A. (2018). *Education policy in the UK*. London, United Kingdom: Routledge.

Maisuria, A. (2014). The Neo-Liberalisation Policy Agenda and its Consequences for Education in England: A Focus on Resistance Now and Possibilities for the Future. *Policy Futures in Education, 12*(2), 286–296. https://doi.org/10.2304/pfie.2014.12.2.286

Mansell, W., James, M., & the Assessment Reform Group. (2009). *Assessment in schools: Fit for purpose? A commentary by the Teaching and Learning Research Programme*. Economic and Social Research Council, Teaching and Learning Research Programme.

Mansur, P., & Thorsuwan, S. (2008). Identity Construction of Young Professionals in a Late Modern Emotion-Intensive Context: Multiple Faces of Family Discourse in a Big Four Auditing Firm.

Marshall, A. (2013). *Principles of Economics* (8th ed.). Palgrave Macmillan. (Original work published 1890).

Martinez-Arias, R., Gaviria, J.-L., & Morera, M. (2009). Concept and evolution of educational value-added models. *Revista de Educacion*, *348*, 15–45.

Mayston, D. J., & Jesson, D. S. J. (1991). Educational Performance Assessment: A new framework of analysis. *Policy & Politics*, *19*(2), 99–108. https://doi.org/10.1332/030557391782454296

McCorkel, J., & Meyers, K. (2003). What difference does difference make? Position and privilege in the field. Qualitative Sociology, 26(2), 199–231. https://doi.org/10.1023/A:10229 67012774

McDonald-Kerr, L. and Boyce, G. (2024). Neoliberalism and New Public Management: decision-making and accountability in a public–private partnership. *Accounting Auditing & Accountability Journal, 37(*1), pp. 199–226. doi: 10.1108/AAAJ-01-2020-4392.

McGovern, K. (2014). *Greenhouse Schools in Boston*. TNTP. https://eric.ed.gov/?id=ED537054

McInerny, G. J., Chen, M., Freeman, R., Gavaghan, D., Meyer, M., Rowland, F., Spiegelhalter, D. J., Stefaner, M., Tessarolo, G., & Hortal, J. (2014). Information visualisation for science and policy: engaging users and avoiding bias. *Trends in Ecology & Evolution, 29*(3), 148–157. https://doi.org/10.1016/j.tree.2014.01.003 (Warwick Research Archive Portal)

McShane, J. (2020). *We know off rolling happens: Why are we still doing nothing? Support for Learning*. https://doi.org/10.1111/1467-9604.12309

Meadows, D. H. (2008). *Thinking in systems: A primer*. Chelsea Green Publishing.

Medina, T. R., Smith, S. N., & Long, J. S. (2009). Measurement Models Matter: Implicit Assumptions and Cross-national Research. *International Journal of Public Opinion Research*, *21*(3), 333–361. https://doi.org/10.1093/ijpor/edp037

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys, 54*(6), 1–35. https://doi.org/10.1145/3457607

Meijer, A. (2014). Transparency. In M. Bovens, R. E. Goodin & T. Schillemans (Eds.), *The Oxford Handbook of Public Accountability*. Oxford University Press.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed.). American Council on Education.

Messick, S. (1995). Validity of psychological assessment. *American Psychologist, 50*(9), 741–749.

Meyer, J. W., & Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology, 83*(2), 340–363. https://doi.org/10.1086/226550

Mezzanotte, C., & Calvel, C. (2023). *Indicators of inclusion in education: A framework for analysis (OECD Education Working Papers, No. 300)*. OECD Publishing. https://doi.org/10.1787/d94f3bd8-en

Microsoft Research Lab (2025). *FATE: Fairness, Accountability, Transparency & Ethics in AI*. Retrieved 5 March 2025, from https://www.microsoft.com/en-us/research/theme/fate/

Mikecz, R. (2012). Interviewing elites: Addressing methodological issues. *Qualitative Inquiry, 18*(6), 482–493.

Millar, F. (2018). *The Best for My Child: Did the Schools Market Deliver?* John Catt Educational Limited. https://books.google.co.uk/books?id=v8wStAEACAAJ

Mills, B. (2019, November 23). *Schools should be judged on pupil inclusion, not just exams, think tank says*. Schools Week. https://schoolsweek.co.uk/schools-should-be-judged-on-pupil-inclusion-not-just-exams-think-tank-says/

Mirowski, P., & Plehwe, D. (Eds.). (2009). The road from Mont Pèlerin: The making of the neoliberal thought collective. Harvard University Press.

Mizrahi, S. (2021). Performance Funding and Management in Higher Education: The Autonomy Paradox and Failures in Accountability'. *Public Performance and Management Review*, *44*(2), 294–320. https://doi.org/10.1080/15309576.2020.1806087.

Molnar, C. (2019). *Interpretable machine learning: A guide for making black box models explainable*. https://christophm.github.io/interpretable-ml-book/

Monbiot, G. (2016, April 15). *Neoliberalism – the ideology at the root of all our problems*. The Guardian. https://www.theguardian.com/books/2016/apr/15/neoliberalism-ideology-problem-george-monbiot

Monbiot, G., & Hutchison, P. (2024). *The invisible doctrine: the secret history of neoliberalism (& how it came to control your life).* Allen Lane.

Moon, K. (2014). A Guide to Understanding Social Science Research for Natural Scientists. *Conservation Biology, (5),* pp. 1167–1177. Available at: https://doi.org/10.1111/cobi.12326.

Moore, M. H. (1995). *Creating public value: Strategic management in government*. Harvard University Press.

Moran, M. (2003). *The British Regulatory State: High Modernism and Hyper-Innovation*. Oxford University Press.

Morgan, D. L. (1996). Focus groups. *Annual Review of Sociology, 22*, 129–152.

Morgan, D. L. (2014). *Integrating Qualitative and Quantitative Methods: A Pragmatic Approach*. Thousand Oaks: Sage.

Mulholland, H. (2012). *Michael Gove accuses unions of undermining respect for teaching*. The Guardian. https://www.theguardian.com/politics/2012/oct/09/michael-gove-unions-respect-teaching

Mundy, K., Green, A., Lingard, B., & Verger, A. (2016*). The handbook of global education policy*. Wiley-Blackwell.

Murphy, R. (n.d.). *It's horrible to watch a health crisis in the making*. Retrieved 29 May 2025, from https://www.taxresearch.org.uk/Blog/2025/05/29/its-horrible-to-watch-a-health-crisis-in-the-making/

Murray, K., & Howe, K. R. (2017). Neglecting democracy in education policy: A-F school report card accountability systems. *Education Policy Analysis Archives*, *25*, 109–109. https://doi.org/10.14507/epaa.25.3017

Murtagh, F., & Legendre, P. (2014). Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? *Journal of Classification*, *31*(3), 274–295. https://doi.org/10.1007/s00357-014-9161-z

Naples, N. A. (2003). *Feminism and method: Ethnography, discourse analysis and activist research*. Routledge.

Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A., & Giovannini, E. (2008). *Handbook on Constructing Composite Indicators: Methodology and User Guide*. OECD.

National Assessment and Testing: The TGAT Report. (1989). *British Journal of Sociology of Education, 10*(2), 233–251. https://doi.org/10.1080/014256989

National Audit Office. (2018). *Ofsted's inspection of schools (HC 1109, Session 2017–2019).* NAO. (https://www.nao.org.uk/reports/ofsteds-inspection-of-schools)

National Education Union. (2024, January 18). *Ofsted safeguarding risk*. https://neu.org.uk/latest/press-releases/ofsted-safeguarding-risk

National Research Council. (2011). *Incentives and test based accountability in education.* Washington, DC: The National Academies Press.

Neely, A., Adams, C., & Kennerley, M. (2002). *The Performance Prism: The Scorecard for Measuring and Managing Business Success*. Pearson.

Neumann, E., Gewirtz, S., Maguire, M., & Towers, E. (2020). Neoconservative education policy and the case of the English Baccalaureate. *Journal of Curriculum Studies, 52*(5), 702–719. https://doi.org/10.1080/00220272.2019.1708466

Nevile, A., Malbon, E., Kay, A., & Carey, G. (2019). The implementation of complex social policy: Institutional layering and unintended consequences in the National Disability Insurance Scheme. *Australian Journal of Public Administration, 78*(4), 562-576.

Newton, P. E. (2016). Fairness and justice in assessment. *Educational Measurement: Issues and Practice, 35*(3), 5–14. https://doi.org/10.1111/emip.12110

Nguyen, T.-T., Pham, H.-H., Cao, Q.-T., Nguyen, X. A., & Do, M. T. (2021). Investigating the impacts of core educational quality on the satisfaction and loyalty of parents of secondary school students: The mediating role of transformative quality. *Cogent Education*, *8*(1), 1911283. https://doi.org/10.1080/2331186X.2021.1911283

NHS England. (2023, November 21). *One in five children and young people had a probable mental disorder in 2023*. https://www.england.nhs.uk/2023/11/one-in-five-children-and-young-people-had-a-probable-mental-disorder-in-2023/

Niemann, D., Martens, K., & Teltemann, J. (2017). PISA and its consequences: Shaping education policies through international comparisons'. *European Journal of Education*, *52*(2), 175–183. https://doi.org/10.1111/ejed.12220.

Noble, S. U. (2018). Algorithms of oppression: How search engines reinforce racism. New York University Press.

Northern Powerhouse Partnership. (2024, October 30). *Northern Powerhouse launches third edition of Fairer Schools Index*. https://www.northernpowerhousepartnership.co.uk/northern-powerhouse-launches-third-edition-of-fairer-schools-index/

Nuttall, D. L., Goldstein, H., Prosser, R., & Rasbash, J. (1989). Differential school effectiveness. International Journal of Educational Research, 13(7), 769–776. https://doi.org/10.1016/0883-0355(89)90027-X

Nye, P. (2020). *How pupil characteristics interact to influence education outcomes*. https://ffteducationdatalab.org.uk/2020/09/how-pupil-characteristics-interact-to-influence-education-outcomes/

Nye, P., & Thomson, D. (2019). *Who's Left 2019, part one: The disadvantage gap is bigger than we thought*. https://ffteducationdatalab.org.uk/2019/12/whos-left-2019-part-one-the-disadvantage-gap-is-bigger-than-we-thought/

O'Higgins, A., Sebba, J., & Luke, N. (2015). *What is the relationship between being in care and the educational outcomes of children?* Rees Centre, University of Oxford. https://www.education.ox.ac.uk/wp-content/uploads/2019/05/285231.pdf ORA

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.

O'Neill, O. (2002). *A question of trust*. Cambridge University Press.

O'Neill, O. (2018). Linking Trust to Trustworthiness. *International Journal of Philosophical Studies, 26*(2), 293–300. https://doi.org/10.1080/09672559.2018.1454637

OECD. (2016). *PISA 2015 results (Volume I): Excellence and equity in education*. OECD Publishing. https://doi.org/10.1787/9789264266490-en

OECD. (2019). *TALIS 2018 results (Volume I): Teachers and school leaders as lifelong learners*. OECD Publishing.

Ofsted. (2023). *Education inspection framework*. GOV.UK.

Ofsted. (2023). *School inspection handbook*. GOV.UK.

Ofsted. (2023). *School monitoring handbook*. GOV.UK.

Office for Statistics Regulation. (2022, July 13). *Assessment of compliance with the Code of Practice for Statistics: Key Stage 4 performance statistics for England*. UK Statistics Authority. https://osr.statisticsauthority.gov.uk/publication/assessment-of-compliance-with-the-code-of-practice-for-statistics-key-stage-4-performance-statistics-for-england

Office of Qualifications and Examinations Regulation. (2022). *Infographics for GCSE results, 2022*. GOV.UK.

Office of Qualifications and Examinations Regulation. (n.d.). *Analytics*. GOV.UK.

Ofsted. (2005). *Excellence in Cities. Managing associated initiatives to raise standards*. Ofsted

Ofsted. (2017). *School inspection handbook: Handbook for inspecting schools in England under section 5 of the Education Act 2005*. GOV.UK. https://www.gov.uk/government/publications/school-inspection-handbook eif

Ofsted. (2019a). *Education inspection framework 2019: Inspecting the substance of education*. Available from: https://www.gov.uk/government/consultations/education-inspection-framework-2019-inspecting-the-substance-of-education

Ofsted. (2019b). *Exploring the issue of off rolling*. GOV.UK. https://www.gov.uk/government/publications/off-rolling-exploring-the-issue

Ofsted. (2019c). Inspection of Haydock High School. (Report No. 50134809). https://files.ofsted.gov.uk/v1/file/50134809

Ofsted. (2022). *Corporate governance framework*. GOV.UK. Available from: https://www.gov.uk/government/publications/ofsted-corporate-governance-framework

Ofsted. (2023). *Schools commentary: Alignment between inspection grades and headline data measures*. GOV.UK. Retrieved 27 October 2025, from

https://www.gov.uk/government/publications/school-inspections-statistical-commentaries-2022-to-2023/schools-commentary-alignment-between-inspection-grades-and-headline-data-measures

Ofsted. (2023*). State-funded schools inspections and outcomes: Methodology and quality report.* London, United Kingdom: Gov.uk.

Ofsted. (2024a). *Findings of Ofsted's Big Listen public consultation*. GOV.UK. https://www.gov.uk/government/publications/ofsted-big-listen-supporting-documents/findings-of-ofsteds-big-listen-public-consultation

Ofsted. (2024b). *Hearing feedback, accepting criticism and building a better Ofsted: the response to the Big Listen*. GOV.UK. https://www.gov.uk/government/consultations/ofsted-big-listen/outcome/hearing-feedback-accepting-criticism-and-building-a-better-ofsted-the-response-to-the-big-listen

Ofsted. (2025, October 31). *Ofsted Parent View: Management information*. https://www.gov.uk/government/statistical-data-sets/ofsted-parent-view-management-information

Ofsted. (2025a). *Annual report and accounts 2024–25 (HC 123)*. The Stationery Office. (https://www.gov.uk/government/publications/ofsted-annual-report-and-accounts-2024-to-2025)

Ofsted. (2025b). *Improving the way Ofsted inspects education: Consultation document.* Retrieved 25 February 2025, from https://www.gov.uk/government/consultations/improving-the-way-ofsted-inspects-education/improving-the-way-ofsted-inspects-education-consultation-document

Ofsted. (2025c). *Improving the way Ofsted inspects education: Consultation document*. Retrieved 21 July 2025, from https://www.gov.uk/government/consultations/improving-the-way-ofsted-inspects-education/improving-the-way-ofsted-inspects-education-consultation-document

Ofsted. (2025d). *Ofsted confirms changes to education inspection and unveils new-look report cards.* Retrieved 11 November 2025, from https://www.gov.uk/government/news/ofsted-confirms-changes-to-education-inspection-and-unveils-new-look-report-cards

Ofsted. (2025e). *School inspection data summary report (IDSR) guide*. GOV.UK. Retrieved 13 November 2025, from https://www.gov.uk/guidance/school-inspection-data-summary-report-idsr-guide

Ofsted. (2025f). *The role of data on school and further education inspections*. Ofsted.

Ofsted. (2026). *Methodology — Explore an area*. Ofsted.

Ogden, V. J. (2008). Policy impact and policy process in transforming urban schools: A single school case study of the London Challenge. *Educate~, 8*(2), 36–49.

Olson, M. (2010). *The logic of collective action: Public goods and the theory of groups*. Harvard University Press.

Ord, J. K. (1979). *Time Series and Spatial Patterns in Ecology* (C. RM & O. JK, Eds.). International Cooperative Publishing House.

Ordofa, B., & Asgedom, A. (2022). *School accountability and its relationship with learning outcomes: A systematic literature review*. Social Sciences & Humanities Open.

Ozga, J. (2009). Governing education through data in England. *Journal of Education Policy, 24*(2), 149–162.

Ozga, J. (2019). Accountability as a policy technology: Accounting for education performance in Europe. *International Review of Administrative Sciences, 79*(2), 292–309. https://doi.org/10.1177/0020852313477762

Paletta, A. (2019). How do school leaders respond to the growing intrusiveness of accountability policies? Evidence from Italy'. *Journal of Educational Administration and History*, *51*(4), 381–401. https://doi.org/10.1080/00220620.2019.1607266.

Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health and Mental Health Services Research, 42*(5), 533–544.

Pangrazio, L., Auld, G., Lynch, J., Sawatzki, C., Duffy, G., Hannigan, S., & O'Mara, J. (2024). *Data justice in education: Toward a research agenda*. *Educational Philosophy and Theory*. https://doi.org/10.1080/00131857.2024.2320196

Parcerisa, L., Verger, A., Pagès, M., & Browes, N. (2022). Teacher autonomy in the age of performance-based accountability: A review based on teaching profession regulatory models. *Education Policy Analysis Archives, 30*(100).

Parsons, S., & Demie, F. (2019). The link between special educational needs and English as an additional language and the educational progress of students. *Educational Psychology, 39*(7), 896–912.

Parsons, S., & Platt, L. (2014). The early academic progress of children with special educational needs. *Journal of Educational Psychology, 106*(2), 481–493.

Patton, M. Q. (2015). *Qualitative research & evaluation methods: Integrating theory and practice* (4th ed.). Sage.

Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. Sage.

Payne, G. (2012). A New Social Mobility? The Political Redefinition of a Sociological Problem. *Contemporary Social Science. 7* (1): 55–71. doi:10.1080/21582041.2011.652360. S2CID 144041004.

Payne, G. (2017). *The New Social Mobility: How the Politicians Got It Wrong*. Bristol: Policy Press. ISBN 978-1-447-31065-5.

Peck, J., & Tickell, A. (2002). Neoliberalizing Space. *Antipode, 34*(3), 380–404. https://doi.org/10.1111/1467-8330.00247

Perry, T. (2016). English Value-Added Measures: Examining the Limitations of School Performance Measurement. *British Educational Research Journal*, *42*(6), 1056–1080. https://doi.org/10.1002/berj.3247

Perry, T. (2019). 'Phantom' compositional effects in English school value-added measures: The consequences of random baseline measurement error. *Research Papers in Education*, *34*(2), 239–262. https://doi.org/10.1080/02671522.2018.1424926

Perryman, J. (2009). Inspection and the fabrication of professional and performative processes. *Journal of Education Policy, 22*(5), 545–563. https://doi.org/10.1080/02680930701541735

Perryman, J. (2022). *Teacher Retention in an Age of Performative Accountability: Target Culture and the Discourse of Disappointment*. Routledge. https://doi.org/10.4324/9780429344121

Perryman, J., & Calvert, G. (2020). What motivates people to teach and why do they leave? Accountability, performativity and teacher retention. *British Journal of Educational Studies, 68*(1), 3–23. https://doi.org/10.1080/00071005.2019.1589417

Pollitt, C. (2007a). Convergence or divergence: What has been happening in Europe? In C. Pollitt, S. van Thiel & V. Homburg (Eds.), *New public management in Europe: Adaptation and alternatives*. Palgrave Macmillan.

Pollitt, C. (2007b). The New Public Management: An Overview of Its Current Status. *Administratie si Management Public, 8(*1), pp. 110–115.

Pope, T., & Tymms, P. (2017). *How could Ofsted and the Department for Education reform school inspection?* Institute for Government. (https://www.instituteforgovernment.org.uk/publication/school-inspection)

Porter, T. M. (1995). *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton University Press.

Power, M. (1997). *The audit society: Rituals of verification*. Oxford University Press.

Prebble, T. (2005). *Impact of student support services and academic development programmes on student outcomes in undergraduate tertiary study: A synthesis of the research : report to the Ministry of Education*. Ministry of Education.

Pressman, J. L., & Wildavsky, A. (1984). *Implementation* (3rd ed.). University of California Press.

Prideaux, S. (2005). *Not so new Labour: A sociological critique of New Labour's policy and practice.* The Policy Press.

Pring, R. (2012). Putting persons back into education. *Oxford Review of Education*, *38*(6), 747–760. https://doi.org/10.1080/03054985.2012.744193

Prior, L. J., Jerrim, J., Thomson, D., & Leckie, G. (2021). A review and evaluation of secondary school accountability in England: Statistical strengths, weaknesses and challenges for 'Progress 8'. *Review of Education, 9*(3), e3299. https://doi.org/10.1002/rev3.3299

Prior, L., Evans, C., Merlo, J., & Leckie, G. (2023*). Sociodemographic inequalities in student achievement: An intersectional multilevel analysis of individual heterogeneity and discriminatory accuracy (MAIHDA) with application to students in London, England* (No. arXiv:2211.06321). arXiv. https://doi.org/10.48550/arXiv.2211.06321

Prior, L., Goldstein, H., & Leckie, G. (2020). School value-added models for multivariate academic and non-academic outcomes: A more rounded approach to using student data to inform school accountability. *arXiv*.

Prior, L., Jerrim, J., Thomson, D., & Leckie, G. (2021). A review and evaluation of secondary school accountability in England: Statistical strengths, weaknesses and challenges for `Progress 8' raised by COVID-19. *Review of Education, 9*(3). https://doi.org/10.1002/rev3.3299

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). SAGE Publications.

Ray, A., McCormack, T., & Evans, H. (2009). Value added in English schools. *Education Finance and Policy, 4*(4), 415–438. https://doi.org/10.1162/edfp.2009.4.4.415

Reardon, S. F. (2011). The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. In R. Murnane & G. Duncan (Eds.), *Whither opportunity? Rising inequality, schools and children's life chances* (pp. 91–116). Russell Sage Foundation.

Reay, D. (2012). What would a socially just education system look like?: Saving the minnows from the pike. *Journal of Education Policy, 27*(5), 587–599. https://doi.org/10.1080/02680939.2012.710015

Reay, D. (2017). *Miseducation: Inequality, education and the working classes*. Policy Press.

Rethinking Assessment. (2024). *Balanced Scorecards: Towards Broader and Better Measures of School Success.* Retrieved 11 November 2025, from https://rethinkingassessment.com/rethinking-blogs/balanced-scorecards-towards-broader-and-better-measures-of-school-success/

Revise Sociology. (2015). *Positivism and Interpretivism in Social Research*. Available at: https://revisesociology.com/2015/05/18/positivism-interpretivism-sociology/ (Accessed: 23 February 2024).

Rhodes, C., & Brundrett, M. (2009). Leadership development and school improvement. *Educational Review, 61*(4), 361-374. https://doi.org/10.1080/00131910903403949

Richardson, J., & Treadaway, M. (2016). https://ffteducationdatalab.org.uk/2016/11/apples-to-apples-are-grammar-schools-really-as-effective-as-they-seem/

Richardson, J., & Treadaway, M. (2016). https://ffteducationdatalab.org.uk/2016/11/apples-to-apples-are-grammar-schools-really-as-effective-as-they-seem/

Rizvi, F., & Lingard, B. (2010). *Globalizing education policy*. Routledge.

Robbins, L. (2014). *An Essay on the Nature and Significance of Economic Science*. Macmillan. (Original work published 1932).

Roberson, C. C. B. (2025). *Biases and inequality in school systems: A literature review on disproportionality in special education and school discipline (Unpublished report).* SREE/Oak Foundation.

Roberts, J. (2009). No one is perfect: The limits of transparency and an ethic for "intelligent" accountability. *Accounting, Organizations and Society, 34*(8), 957–970. https://doi.org/10.1016/j.aos.2009.04.005

Robinson, D. (2024). *Access to extra-curricular provision and the association with outcomes*. Education Policy Institute.

Robinson, O. C. (2014). Sampling in interview-based qualitative research: A theoretical and practical guide. *Qualitative Research in Psychology, 11*(1), 25–41.

Rose, N. (1999). *Powers of freedom: Reframing political thought*. Cambridge University Press.

Rose, N., & Miller, P. (2008). *Governing the present: Administering economic, social and personal life*. http://www.polity.co.uk/

Rosenbaum, E. (2022). Mental models and institutional inertia. *Journal of Institutional Economics, 18*(3), 361–378. https://doi.org/10.1017/S174413742100059X

Ruppert, E., Savage, J., & Shkolnik, D. (2018). Data politics. *Big Data & Society, 5*(2), 205395171878891.

Russo, F., & Iliadis, A. (2016). Critical data studies: An introduction. *Big Data & Society, 3*(2). https://doi.org/10.1177/2053951716674238 Directory of Open Access Journals.

Sahlberg, P. (2011). *Finnish lessons: What can the world learn from educational change in Finland?* Teachers College Press.

Sahlberg, P. (2016). The global educational reform movement and its impact on schooling. In K. Mundy, A. Green, B. Lingard & A. Verger (Eds.), *The handbook of global education policy* (pp. 128–144). Wiley-Blackwell.

Sammons, P., Mortimore, P., & Thomas, S. (1996). School effectiveness and school improvement in the United Kingdom. *School Effectiveness and School Improvement*, 7(2), 95–123.

Sammons, P., Nuttall, D., & Cuttance, P. (1993). Differential school effectiveness: Results from a reanalysis of the Inner London Education Authority's Junior School Project data. *British Educational Research Journal, 19*(4), 381–405. https://doi.org/10.1080/0141192930190407

Samuelson, P. A., & Nordhaus, W. D. (2009). *Economics* (19th ed.). McGraw-Hill.

Saunders, L. (1998). *'Value added' measurement of school effectiveness: An overview*. National Foundation for Educational Research.

Schäfer, M. T., Es, K. v., & Lauriault, T. P. (Eds.). (2024). *Collaborative research in the datafied society: Methods and practices for investigation and intervention*. Amsterdam University Press.

Schedler, K., & Proeller, I. (2000). *New Public Management*. UTB.

Scheerens, J., & Bosker, R. J. (1997). *The foundations of educational effectiveness.* Oxford: Pergamon.

Scheerens, J., & Creemers, B. (1989). Conceptualizing school effectiveness. International Journal of *Educational Research, 13*(7), 691-706. https://doi.org/10.1016/0883-0355(89)90022-0

Scherer, R., & Nilsen, T. (2019). Closing the gaps? Differential effectiveness and accountability as a road to school improvement. *School Effectiveness and School Improvement, 30*(2), 255-260. https://doi.org/10.1080/09243453.2019.1623450

Scherer, R., & Nilsen, T. (2019). Closing the gaps? Differential effectiveness and accountability as a road to school improvement. School Effectiveness and School Improvement, 30(3), 255–260. https://doi.org/10.1080/09243453.2019.1623450

Schmidt, V. A. (2008). Discursive institutionalism: The explanatory power of ideas and discourse. *Annual Review of Political Science, 11*, 303–326. https://doi.org/10.1146/annurev.polisci.11.060606.135342

Schneider, A., & Ingram, H. (1997). *Policy design for democracy*. University Press of Kansas.

School approaches to the education of EAL students. (2012). ERIC Document Reproduction Service. https://files.eric.ed.gov/fulltext/EJ1410375.pdf

School Performance Tables. (2022). *Component measures*. https://www.schoolperformancetables.com/documentation/component-measures

SchoolDash. (2025). *Dashboards*. Retrieved 13 November 2025, from https://www.schooldash.com/dashboards.html

Schools BI. (2025). *Schools BI - Data Analytics for Schools and MATs*. Retrieved 23 October 2025, from https://www.schoolsbi.co.uk/

Schultz, T. W. (1961). Investment in Human Capital. *The American Economic Review*, *51*(1), 1–17. http://www.jstor.org.sheffield.idm.oclc.org/stable/1818907

Schwandt, T. A. (1997). *Qualitative inquiry: A dictionary of terms*. Sage Publications, Inc.

Sculley, D., & Pasanek, B. (2008). Meaning and mining: The impact of implicit assumptions in data mining for the humanities. *Literary and Linguistic Computing, 23*(4), 409–424. https://doi.org/10.1093/llc/fqn019 (OUP Academic)

Secretary of State for Education and Employment. (1996). Self-Government for Schools (Cm. 3315). Her Majesty's Stationery Office. https://www.education-uk.org/documents/official-papers/1996a-wp-self-government-for-schools.html

Secretary of State for Education. (2012, November 14). *Secretary of State for Education Michael Gove gives speech to IAA [Speech].* GOV.UK. https://www.gov.uk/government/speeches/secretary-of-state-for-education-michael-gove-gives-speech-to-iaa

Selbst, A. D., Boyd, D., Friedler, S. A., Vandenberghe, F., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. *Proceedings of the Conference on Fairness, Accountability and Transparency (FAT*),* 59–68. https://doi.org/10.1145/3287560.3287598

Sellar, S., & Lingard, B. (2013). The OECD and global governance in education. *Journal of Education Policy, 28*(5), 710–725. https://doi.org/10.1080/02680939.2013.779791

Sen, A. (2009). *The idea of justice*. Harvard University Press.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). Experimental and quasi-experimental designs for generalized causal inference. Houghton Mifflin.

Sharma, S., Henderson, J., & Ghosh, J. (2020, February). CERTIFAI. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*. https://doi.org/10.1145/3375627.3375812

Shepherd, K., & Adams, R. (2025). Beyond the score: Validity, equity and interpretation in diverse assessment contexts. *Educational Assessment, Evaluation and Accountability*, 37, 163–169.

Shermer, E.B. (2014). *The Neoliberal turn: Historical and ideological perspectives on the resurgence of free markets*. Routledge.

Shore, C., & Wright, S. (2011). *Policy worlds: Anthropology and the analysis of contemporary power*. Berghahn.

Shore, C., & Wright, S. (2015). Governing by numbers. *Social Anthropology, 23*(1), 22–28.

Shore, C., & Wright, S. (2024). *Audit culture : how indicators and rankings are reshaping the world (First edition.)*. Pluto Press. https://doi.org/10.2307/jj.10819589

Siegel, R., König, C. J., & Lazar, V. (2022). The impact of electronic monitoring on employees' job satisfaction, stress, performance and counterproductive work behaviour: A meta analysis. *Computers in Human Behavior Reports, 8*, 100227. https://doi.org/10.1016/j.chbr.2022.100227

Silvert, W. L. (1981). Principles of ecosystem modelling. In ed. (Ed.), *Longhurst AR*. Academic Press.

Singhal, A., Neveditsin, N., Tanveer, H., & Mago, V. (2024). Toward Fairness, Accountability, Transparency and Ethics in AI for Social Media and Health Care: Scoping Review. JMIR Medical Informatics, 12, e50048–e50048. https://doi.org/10.2196/50048

Sivo, S. A., Fan, X., Witta, E. L., & Willse, J. T. (2006). The Search for 'Optimal' Cutoff Properties: Fit Index Criteria in Structural Equation Modeling. *The Journal of Experimental Education*, *74*(3), 267–288. https://doi.org/10.3200/JEXE.74.3.267-288

Slater, R., & Teddlie, C. (1992). Toward a theory of school effectiveness and leadership. *School Effectiveness and School Improvement, 3*(4), 247-257. https://doi.org/10.1080/0924345920030402

Smith, P. (1995). On the unintended consequences of publishing performance data in the public sector. *International Journal of Public Administration, 18*(2–3), 277–310.

Smith, W. C., & Holloway, J. (2020). School testing culture and teacher satisfaction. *Educational Assessment, Evaluation and Accountability, 32*(4), 461–479.

Sokol, K., Hepburn, A., Poyiadzi, R., Clifford, M., Santos-Rodriguez, R., & Flach, P. (2020). FAT Forensics: A Python Toolbox for Implementing and Deploying Fairness, Accountability and Transparency Algorithms in Predictive Systems. *Journal of Open Source Software*, *5*(49), 1904. https://doi.org/10.21105/joss.01904

Spiegelhalter, D. (2019). *The art of statistics: How to learn from data*. Basic Books. (UK edition published by Pelican) (Penguin Random House)

Spillane J. P. (2012). Data in practice: Conceptualizing the data-based decision-making phenomena. *American Journal of Education*, 118, 113-141.

Spillane, J. P., & Burch, P. (2006). The institutional environment of instructional practice in districts and schools. In J. Hannaway & A. J. Daly (Eds.), *Smart districts, savvy schools* (pp. 87–105). Corwin.

Stewart, W. (2011). The challenge now is to hang on to this success: Comment. *The Times Educational Supplement, 4934*, 34.

Stiglitz, J. E. (2002). *Information and the Change in the Paradigm in Economics (NBER Working Paper No. 8977)*. National Bureau of Economic Research.

Stone, D. A. (2012). *Policy paradox: The art of political decision making* (3rd ed.). W. W. Norton.

Strand, S, Malmberg, D. L., & Hall, D. J. (2015). *English as an Additional Language (EAL) and educational achievement in England: An analysis of the National Pupil Database*. University of Oxford.

Strand, S. (2010). Do some schools narrow the gap? Differential school effectiveness by ethnicity, gender, poverty, and prior achievement. *School Effectiveness and School Improvement, 21*(3), 289–314. https://doi.org/10.1080/09243451003732651

Strand, S. (2010). Do some schools narrow the gap? Differential school effectiveness by ethnicity, gender, poverty, and prior achievement. *School Effectiveness and School Improvement*, 21(3), 289–314.

Strand, S. (2014). Ethnicity, gender, social class and achievement gaps at age 16: Intersectionality and "getting it" for the white working class. *Research Papers in Education*, 29(2), 131–171. https://doi.org/10.1080/02671522.2013.767370

Strand, S. (2016). Do some schools narrow the gap? Differential school effectiveness revisited. *Review of Education, 4*(2), 107–144. https://doi.org/10.1002/rev3.3054

Strand, S. (2016). Ethnicity, deprivation and educational achievement at age 16 in England: Trends over time. Oxford Review of Education, 42(5), 637–659.

Strand, S., & Hessel, A. (2018). *English language proficiency and attainment of EAL pupils in England*. Education Policy Institute.

Strathern, M. (2000). *Audit cultures: Anthropological studies in accountability, ethics and the academy*. Routledge.

Tashakkori, A. and Teddlie, C. (1998). Mixed Methodology: Combining Qualitative and Quantitative Approaches. *Applied Social Research Methods Series, 46*; Thousand Oaks: Sage Publications.

Tashakkori, A., & Teddlie, C. (2010). *Mixed methods in social & behavioral research* (2nd ed.). SAGE.

Task Group on Assessment and Testing. (1987). National Curriculum: Task Group on Assessment and Testing. A Report. Department of Education and Science and the Welsh Office

Taut, S., Valencia, E., Palacios, D., Santelices, M. V., Jiménez, D., & Manzi, J. (2016). Teacher performance and student learning: Linking evidence from two national assessment programmes. *Assessment in Education: Principles, Policy & Practice*, *23*(1), 53–74. https://doi.org/10.1080/0969594X.2014.961406

Taylor, L. (2017). What is data justice? The case for connecting digital rights and freedoms globally. *Big Data & Society, 4*(2), 1–14.

Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed methods research*. SAGE.

Tekwe, C. D., Carter, R. L., Ma, C.-X., Algina, J., Lucas, M. E., Roth, J., Ariet, M., Fisher, T., & Resnick, M. B. (2004). An Empirical Comparison of Statistical Models for Value-Added Assessment of School Performance. *Journal of Educational and Behavioral Statistics, 29*(1), 11–35. https://doi.org/10.3102/10769986029001011

The Guardian. (2025). *Bridget Phillipson: 'We must think very differently' on school Send support Special educational needs*. Retrieved 22 July 2025, from https://www.theguardian.com/education/2025/mar/14/different-thinking-needed-send-funding-bridget-phillipson-says

The Guardian. (2025, November 12). *All schools in England to be given AI-generated pupil attendance targets.* Retrieved 12 November 2025, from https://www.theguardian.com/education/2025/nov/12/all-schools-in-england-to-be-given-ai-generated-pupil-attendance-targets

The Independent. (2025). *Work needed to make schools inclusive for pupils with special needs – adviser*. Retrieved 22 July 2025, from https://www.independent.co.uk/news/uk/home-news/work-government-england-bridget-phillipson-schools-b2781970.html

Thomas, S., Sammons, P., Mortimore, P., & Smees, R. (1997). Differential secondary school effectiveness: Comparing the performance of different pupil groups. *British Educational Research Journal, 23*(4), 451–469. https://doi.org/10.1080/0141192970230405

Thompson, N. (2008). Hollowing Out the State: Public Choice Theory and the Critique of Keynesian Social Democracy. *Contemporary British History, 22*(3), pp. 355–382. Available at: https://doi.org/10.1080/13619460701731913.

Thrupp, M. (1999). *Schools making a difference : let's be realistic! : school mix, school effectiveness and the social limits of reform*. Buckingham: Open University Press.

Thrupp, M. (2001). Sociological and political concerns about school effectiveness research: Time for a new research agenda. *School Effectiveness and School Improvement, 12*(1), 40-7. https://doi.org/10.1076/sesi.12.1.7.3459

Thrupp, M., & Lupton, R. (2006). Taking school contexts more seriously: The social justice challenge. *British Journal of Educational Studies, 54*(3), 308–328. https://doi.org/10.1111/j.1467-8527.2006.00348.x

Timmermans, A., Bosker, R., Wolf, I., Doolaard, S., & Werf, M. (2013). Value added based on educational positions in Dutch Secondary Education. *British Educational Research Journal*, *40*. https://doi.org/10.1002/berj.3126

Todd, P. E., & Wolpin, K. I. (2003). On the Specification and Estimation of the Production Function for Cognitive Achievement. *The Economic Journal*, *113*(485), F3–F33. https://doi.org/10.1111/1468-0297.00097

Tracy, S. J. (2010). Qualitative Quality: Eight "Big-Tent" Criteria for Excellent Qualitative Research. *Qualitative Inquiry*, *16*(10), 837-851.

Troncoso, P. (2019). A two-fold indicator of school performance and the cost of ignoring it. *International Journal of Educational Research, 95*, 153–167. https://doi.org/10.1016/j.ijer.2019.02.002

Turner, J.H. (2001). Positivism: Sociological. In N.J. Smelser and P.B. Baltes (eds) *International Encyclopedia of the Social & Behavioral Sciences*. Oxford: Pergamon, pp. 11827–11831. Available at: https://doi.org/10.1016/B0-08-043076-7/01941-0.

Turner, R. S. (2007). The 'rebirth of liberalism': The origins of neo-liberal ideology. *Journal of Political Ideologies, 12*(1), 67–83. https://doi.org/10.1080/13569310601095614

UK Government. (2018). *Off rolling in English schools (House of Commons Library Briefing Paper CBP 8444)*. https://researchbriefings.files.parliament.uk/documents/CBP-8444/CBP-8444.pdf#

UK Government. (2025). *Data and AI ethics framework*. GOV.UK. https://www.gov.uk/government/publications/data-ethics-framework/data-and-ai-ethics-framework

UK Health Security Agency. (n.d.). Coronavirus (COVID-19) surveillance dashboard. https://ukhsa-dashboard.data.gov.uk/respiratory-viruses/covid-19

UK Parliament. (1987, December 1). *Education Reform Bill: Second reading*. Hansard, House of Commons Debates, Vol. 123.

UK Parliament. (1988, July 18). *Education Reform Bill (allocation of time)*. Hansard, House of Commons Debates.

UK Parliament. (1988, June 22). *Education Reform Bill*. Hansard, House of Commons Debates.

UK Parliament. (1988, May 3). *Education Reform Bill*. Hansard, House of Commons Debates.

UK Parliament. (1991, July 12). *Education (Information on School Examination Results) (England) Regulations 1991*. Hansard, House of Commons Debates.

UK Parliament. (1996, June 25). *Schools (self-government)*. Hansard, House of Commons Debates.

UK Parliament. (1996, November 13). *Educational choice [House of Lords debate]*. Hansard. https://api.parliament.uk/historic-hansard/lords/1996/nov/13/educational-choice

UK Statistics Authority. (2015). *Assessment report: school level examination statistics for England*. Office for Statistics Regulation. https://uksa.statisticsauthority.gov.uk/wp-content/uploads/2015/11/images-assessmentreport289schoollevelexaminationstatisticsforenglan_tcm97-44097.pdf

UNESCO. (1995). *UNESCO Prize 1995 for Peace Education—UNESCO Digital Library*. Retrieved 23 January 2026, from https://unesdoc.unesco.org/ark:/48223/pf0000109611

UNESCO. (2017a). *A guide for ensuring inclusion and equity in education*. UNESCO Publishing.

UNESCO. (2017b). *A holistic approach to student well-being in education systems*. UNESCO Publishing.

United Kingdom Parliament. (1988). *Education Reform Act 1988* (c. 40). HMSO.

United Kingdom Parliament. (1992). *Education (Schools) Act 1992* (c. 38). HMSO.

Valdrighi, G., Ribeiro, A. M., Pereira, J. S. B., Guardieiro, V., Hendricks, A., Filho, D. M., Garcia, J. D. N., Bocca, F. F., Veronese, T. B., Wanner, L., & Raimundo, M. M. (2024). *Best Practices for Responsible Machine Learning in Credit Scoring.* https://doi.org/10.48550/arxiv.2409.20536

Valli, L., & Buese, D. (2007). The changing roles of teachers in an era of high-stakes accountability. *American Educational Research Journal, 44*(3), 519–558. https://doi.org/10.3102/0002831207306859

Van der Velden, J., & Harambam, J. (2016). Beyond data as representation: The performativity of big data in surveillance. *Surveillance & Society, 14*(4), 516–528.

Verger, A., Fontdevila, C., & Parcerisa, L. (2019). Reforming governance through policy instruments: How and to what extent standards, tests and accountability in education spread worldwide. *Discourse: Studies in the Cultural Politics of Education, 40*(2), 248–270. https://doi.org/10.1080/01596306.2019.1569882

Verger, A., Parcerisa, L., & Fontdevila, C. (2019). The growth and spread of large-scale assessments and test-based accountabilities: A political sociology of global education reforms. *Educational Review, 71*(1), 5–30. https://doi.org/10.1080/00131911.2019.1522045

Verger, A., Parcerisa, L., Pagès, M., & Camphuijsen, M. (2024). School autonomy with accountability as a cross-national policy model: Diverse adoptions, practices and impacts. *Nordic Journal of Studies in Educational Policy, 10*(3), 167–179. https://doi.org/10.1080/20020317.2024.2430706

Verger, A., Steiner-Khamsi, G., & Lubienski, C. (2017). The emerging global education industry: Analysing market-making in education through market sociology. *GLOBALISATION SOCIETIES AND EDUCATION*, *15*(3), 325–340. (WOS:000405606700004). https://doi.org/10.1080/14767724.2017.1330141

Verger, A., Steiner-Khamsi, G., & Lubienski, C. (Eds.). (2016). *The Global Education Industry: World Yearbook of Education 2016*. Routledge.

von Stumm, S., Smith-Woolley, E., Cheesman, R., Pingault, J.-B., Asbury, K., Dale, P. S., Allen, R., Kovas, Y., & Plomin, R. (2021). School quality ratings are weak predictors of students' achievement and well-being. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *62*(3), 339–348. https://doi.org/10.1111/jcpp.13276

Vono de Vilhena, D., & Loi, S. (2020). *Exclusion through statistical invisibility* (Working Paper No. 13). Humboldt-Universität zu Berlin, SFB 1265 Re-Figuration von Räumen.

Vujnovic, M., & Kruckeberg, D. (2016). Pitfalls and promises of transparency in the digital age. *Public Relations Inquiry, 5*(2), 121-143. https://doi-org.sheffield.idm.oclc.org/10.1177/2046147X16635227 (Original work published 2016)

Wei, R. C., Darling Hammond, L., Andree, A., Richardson, N., & Orphanos, S. (2009). *Professional learning in the learning profession: A status report on teacher development in the United States and abroad*. National Staff Development Council.

Weiss, C. H. (1995). *Nothing as practical as good theory: Exploring theory-based evaluation for comprehensive community initiatives*. New Directions for Evaluation, 1995(65), 65–92.

West, A., Mattei, P., & Roberts, J. (2011). Accountability and Sanctions in English Schools. *British Journal of Educational Studies*, *59*(1), 41–62. https://doi.org/10.1080/00071005.2010.529416

West, A., Mattei, P., & Roberts, J. (2011). Accountability and Sanctions in English Schools. *British Journal of Educational Studies*, *59*(1), 41–62. https://doi.org/10.1080/00071005.2010.529416

Whitty, G. (1997). On the changing relationships between the state, civil society, and changing notions of teacher professionalism. In A. H. Halsey, H. Lauder, P. Brown & A. S. Wells (Eds.), *Education: Culture, economy and society* (pp. 299–310). Oxford, UK: Oxford University Press.
Whitty, G. (2002). Making sense of education policy. Sage.

Whitty, G., & Anders, J. (2013). *(How) did New Labour narrow the achievement and participation gap? LLAKES Research Paper (PDF).* Retrieved [date], from https://www.llakes.ac.uk/wp-content/uploads/2021/03/46.-Whitty-and-Anders-final.pdf

Whitty, G., & Anders, J. (2014). *(How) did New Labour narrow the achievement and participation gap?*

Whitty, G., Power, S., & Halpin, D. (1998). *Devolution and choice in education: The school, the state and the market.* Open University Press/ACER Press.

Wieland, R., Ravensbergen, S., Gregr, E. J., Satterfield, T., & Chan, K. M. A. (2016). Debunking trickle-down ecosystem services: The fallacy of omnipotent, homogeneous beneficiaries. *Ecological Economics*, *121*, 175–180. https://doi.org/10.1016/j.ecolecon.2015.11.007

Wilkins, A. (2018). Neoliberalism, citizenship and education: A policy discourse analysis. In A. Peterson, G. Stahl & H. Soong (Eds.), *The Palgrave Handbook of Citizenship and Education*. Palgrave.

Wilkins, A., Collet-Sabé, J., Gobby, B., & Hangartner, J. (2019). Translations of new public management: a decentred approach to school governance in four OECD countries. *Globalisation, Societies and Education, 17*(2), 147–160. https://doi.org/10.1080/14767724.2019.1588102

Wilkinson, D., Bryson, A., & Stokes, L. (2018). Assessing the Variance in Pupil Attainment: How Important is the School Attended? *National Institute Economic Review, 243*, R4–R16. doi:10.1177/002795011824300110

Williams, A. (2025). *Epistemic Closure and the Irreversibility of Misalignment: Modeling Systemic Barriers to Alignment Innovation*. https://doi.org/10.48550/arxiv.2504.02058

Williams, B. (2024). The 'New Right' and its legacy for British conservatism*. Journal of Political Ideologies*, 29(1), 121–144. https://doi.org/10.1080/13569317.2021.1979139

Williamson, B. (2017). *Big Data in Education: The digital future of learning, policy and practice*. SAGE Publications Ltd. https://doi.org/10.4135/9781529714920

Wills, G., & van der Berg, S. (2020). Measuring school leadership and management and linkages with literacy: Evidence from rural and township primary schools in South Africa. *Educational Management Administration & Leadership*. https://doi.org/10.1177/1741143220915923

Wolf, A. (2011). *Review of vocational education – The Wolf Report*. Department for Education.

Woods, P. A., & Levačić, R. (2002). Raising School Performance in the League Tables (Part 2): Barriers to responsiveness in three disadvantaged schools. *British Educational Research Journal*, *28*(2), 227–247. https://doi.org/10.1080/01411920120122167

Worth, J., McLean, D., & Faulkner-Ellis, H. (2023). Teacher labour market in England: Annual report 2023. *National Foundation for Educational Research*. https://www.nfer.ac.uk/media/uash4sbq/teacher_labour_market_in_england_annual_report_2023.pdf

Wragg, E. C. (1988). *Education in the market place: The ideology behind the 1988 Education Bill.* National Union of Teachers.

Wright, S., & Shore, C. (1999). Audit culture and anthropology: Neo-liberalism in British higher education. *Journal of the Royal Anthropological Institute*, 5(4), 557–575. https://doi.org/10.2307/2661148

Yanow, D. (2000). *Conducting interpretive policy analysis*. SAGE.

Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). SAGE.

Zvoch, K., & Stevens, J. J. (2008). Measuring and Evaluating School Performance: An Investigation of Status and Growth-Based Achievement Indicators. *Evaluation Review, 32*(6), 569-595. https://doi-org.sheffield.idm.oclc.org/10.1177/0193841X08320398

Appendix 1. Coding frequency within documents

Socio-political context (Research Question 1)

Neoliberalism

Name	Files	References
autonomy and freedom	26	256
supported autonomy	2	11
economic focus	31	664
business and enterprise	18	181
efficiency	25	149
finance (local control of)	13	58
requirements of the economy	11	27
value-added	15	84
quasi-market	34	1694
choice	29	1449
Competition	30	235
League Tables	23	196
Inform parents	6	9
standards	36	1761
quality	13	211
raise standards	23	507
self-improving	2	6
target setting	12	41

Neoconservatism

Name	Files	References
standards of behaviour	9	80
discipline	9	72
social mobility	2	7

Social Democracy

Name	Files	References
collaboration	19	286
disadvantage	22	252
equality of opportunity	20	88
equity	8	20
social justice	6	12

Aims and purposes of SPMs (Research Question 2)

Name	Files	References
purpose	27	1621
Publication	10	39
Performance	26	940
Link to Ofsted Inspection	8	49
Intervention	3	3
Inform parents	12	38
Behaviours	7	45
Accountability	14	476
Methodology	9	70
Focus	5	19
Limitations	8	70
League Tables (opposed)	2	7
Covid-19	3	17
interpretation	4	21

Appendix 2. Timeline to show political context of production of documents.

Secretary of State for Education	Start date	End date	Party	Prime Minister	Documents
Fred Mulley	10 June 1975	9 September 1976	Labour	James Callaghan	
Shirley Williams	10 September 1976	4 May 1979	Labour	James Callaghan	
Mark Carlisle	5 May 1979	14 September 1981	Conservative	Margaret Thatcher	
Keith Joseph	14 September 1981	20 May 1986	Conservative	Margaret Thatcher	
Kenneth Baker	21 May 1986	23 July 1989	Conservative	Margaret Thatcher	Hansard Education Reform Bill - Hansard - UK Parliament - 2nd Reading 1 December 1987 Hansard Education Reform Bill 03 May 1988 Hansard Education Reform Bill 22 June 1988 Hansard Education Reform Bill (Allocation Of Time) 18 August 1988 UK Gov The 1988 Education Reform Act
John MacGregor	24 July 1989	1 November 1990	Conservative	Margaret Thatcher	
Kenneth Clarke	2 November 1990	9 April 1992	Conservative	Margaret Thatcher and John Major	Hansard Education (Information On School Examination Results) (England) Regulations 1991 UK Gov Education (Schools) Act 1992
John Patten	10 April 1992	20 July 1994	Conservative	John Major	
Gillian Shephard	20 July 1994	1 May 1997	Conservative	John Major	**1996** White Paper *Learning to Compete: Education and Training for 14-19 Year Olds* (December): set out 'the Government's vision of the education and training world we shall need to meet the demands of the next century'.

David Blunkett	1 May 1997	8 June 2001	Labour	Tony Blair	**1997** White Paper *Excellence in schools*: formed the basis of the 1998 School Standards and Framework Act.
Estelle Morris	8 June 2001	24 October 2002 (resigned)	Labour	Tony Blair	**2001** White Paper *Schools: achieving success*: formed the basis of the 2002 Education Act.
Charles Clarke	24 October 2002	15 December 2004	Labour	Tony Blair	
Ruth Kelly	15 December 2004	5 May 2006	Labour	Tony Blair	**2005** ***Secondary Education*** **(17 March): report by the Commons Education Select Committee critical of the academies programme and the government's plans for 'independent specialist schools'.** **2005** White Paper *Higher Standards, Better Schools for All*: proposed independent trust schools; formed the basis of the 2006 Education and Inspections Bill.
Alan Johnson	5 May 2006	28 June 2007	Labour	Tony Blair	
Ed Balls	28 June 2007	11 May 2010	Labour Co-op	Gordon Brown	**2010** *School Accountability* (7 January): report by the House of Commons Children, Schools and Families Committee on school self-evaluation, inspection, league tables etc.
Michael Gove (tenure)	12 May 2010	15 July 2014	Conservative	David Cameron (Coalition)	**2010** White Paper *The Importance of Teaching*: wide-ranging document covering teaching, leadership, behaviour, new schools, accountability etc.
Nicky Morgan	15 July 2014	13 July 2016	Conservative	David Cameron (II)	**2016** White paper *Educational Excellence Everywhere*: set out the government's proposals for turning all schools into academies.

Justine Greening	14 July 2016	8 January 2018	Conservative	Theresa May (I)	
Damian Hinds	8 January 2018	24 July 2019	Conservative	Theresa May (II)	
Gavin Williamson	24 July 2019	15 September 2021	Conservative	Boris Johnson (I)	
Nadhim Zahawi	15 September 2021	5 July 2022	Conservative	Boris Johnson (II)	
Michelle Donelan	5 July 2022	7 July 2022	Conservative	Boris Johnson (II)	
James Cleverly	7 July 2022	6 September 2022	Conservative	Boris Johnson (II)	Ofqual Infographics for GCSE results, 2022 OSR Assessment_Report_KS4_Statistics (July 2022)
Kit Malthouse	6 September 2022	25 October 2022	Conservative	Liz Truss (Truss ministry)	How to understand school performance – everything you need to know - The Education Hub Key stage 4 performance, Methodology – Explore education statistics Key stage 4 performance, Academic year 2021~22 – Explore education statistics
Gillian Keegan	25 October 2022	5 July 2024	Conservative	Rishi Sunak (Sunak ministry)	DfE Secondary accountability measures DfE Understanding school and college performance measures Ofqual Analytics Ofsted Education inspection framework for September 2023 Ofsted School inspection handbook for September 2023 Ofsted School monitoring handbook for September 2023
Bridget Phillipson	5 July 2024	incumbent	Labour	Sir Kier Starmer	

Appendix 3. Details of data elements used in quantitative analyses

Headline Secondary School Performance Measures

These are the six headline measures used by the government to describe outcomes for state-funded secondary schools in England in 2021/2022, based on pupils at the end of Key Stage 4 (around age 16, Year 11).

Measure	What it measures	How it is calculated
Attainment 8	Overall academic attainment at age 16	Each pupil receives a score based on performance across 8 qualifications (English and mathematics count double, plus other approved subjects). School scores are averaged across pupils.
Progress 8	Academic progress made during secondary school	Compares pupils' key stage 4 results to those of similar pupils nationally who had the same prior attainment at the end of primary school (KS2). School scores are averaged. A score of 0 = average progress nationally.
% Grade 5+ in English and Maths ("Basics")	Proportion of pupils reaching a benchmark in core subjects	Percentage of pupils achieving Grade 5 or above in both English and mathematics GCSEs.
% EBacc Entry	Participation in a broad academic curriculum	Percentage of pupils entered for the subjects that make up the English Baccalaureate (EBacc): English, mathematics, sciences, a language, and history or geography.
EBacc Average Point Score (EBacc APS)	Achievement in EBacc subjects	Average point score achieved by pupils across the EBacc subject set (English, maths, sciences, language, history/geography).
Pupil destinations	The extent to which pupils successfully transition into further education, employment, or training after completing compulsory secondary schooling (Year 11 / Key Stage 4).	For each school, the percentage of pupils who enter and sustain a destination in the academic year after leaving Year 11. A destination is counted if the pupil remains in education, employment, or an apprenticeship for six months (typically October–March for education/employment; apprenticeship rules differ slightly).

Wider School Context and Intake Measures

These indicators are commonly used to describe school context and pupil intake characteristics. They are not performance measures but are often used when analysing differences between schools.

Measure	What it measures	How it is calculated (simplified)
KS2 Average Point Score (KS2 APS)	Prior attainment of pupils before entering secondary school	Average attainment score achieved by pupils at the end of Key Stage 2 (primary school, age 11) before they entered secondary education.
% of pupils with Free School Meals ever in last six years (FSM6)	Socio-economic disadvantage	Percentage of pupils who have been eligible for Free School Meals at any point in the previous six years. Eligibility is linked to household and benefit status.
% English as an Additional Language (EAL)	Linguistic background of pupils	Percentage of pupils whose first language is not English.
% Special Educational Needs (SEN)	Additional learning needs among pupils	Percentage of pupils identified as requiring additional educational support due to learning difficulties, disabilities or other educational needs.
School admission policy (selective or non-selective)	Whether admission is based on academic selection	Selective schools admit some or all pupils based on academic tests (e.g., grammar schools). Non-selective schools admit without academic selection.
Whether the school has a religious character	Formal religious designation of the school	Indicates whether the school is officially associated with a religious denomination (e.g., Church of England, Roman Catholic, Jewish, Muslim, etc.).
School gender of entry (mixed or single gender)	Gender composition of pupils admitted	Mixed schools admit boys and girls. Single-gender schools admit only boys or only girls.

Appendix 4. SEM models

The models for the wider school information latent constructs are developed through a series of iterations. The Initial School Information Model (Figure 24) includes two factors:

- 'School Context' - the continuous variables of KS2_APS, FSM6, EAL and SEN which summarise characteristics of the school's student body and
- 'School Character' - those variables which are associated with what can be described as the character of the school rather than its cohort of students i.e. Selective, Religious_Char and Single_Gender

For each iteration, a variable with a low factor loading is removed and the refined model fit measures compared with its predecessor to ensure that the change had improved the model fit as well as parsimony.

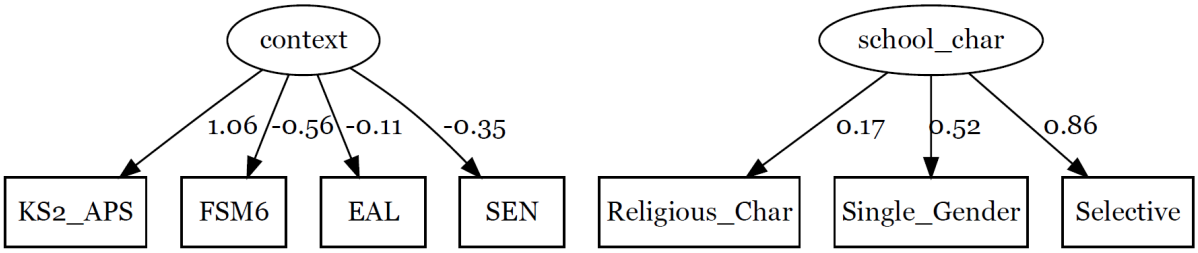


Figure 24. Model diagram for Initial School Information Model.

Table 32. Model diagram for Initial School Information Model.

GFI	NNFI	CFI	RMSEA
0.928	0.749	0.845	0.149

In this first refinement to the model, the variable with the lowest factor loading, EAL, is to removed. All four fit measures for the model are improved by this change (Table 30).

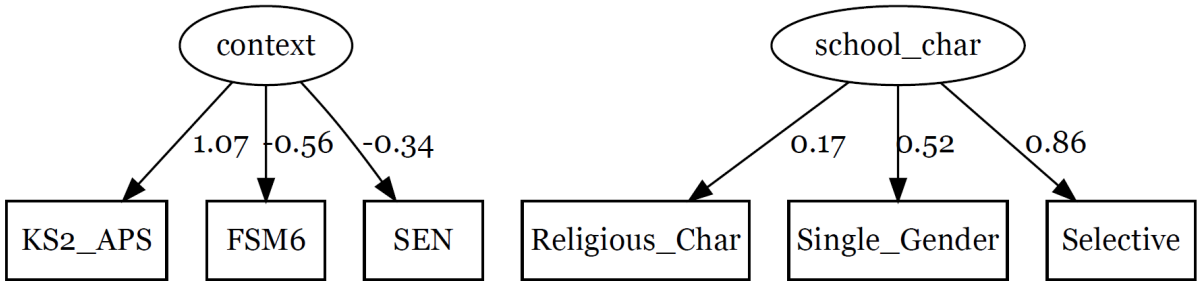


Figure 25. Model diagram for first Refined School Information Model.

Table 33. Model fit indices for first Refined School Information Model.

GFI	NNFI	CFI	RMSEA
0.986	0.951	0.974	0.072

The SEN variable is removed in the second refinement stage. Again this improves the model fit (Table 31).

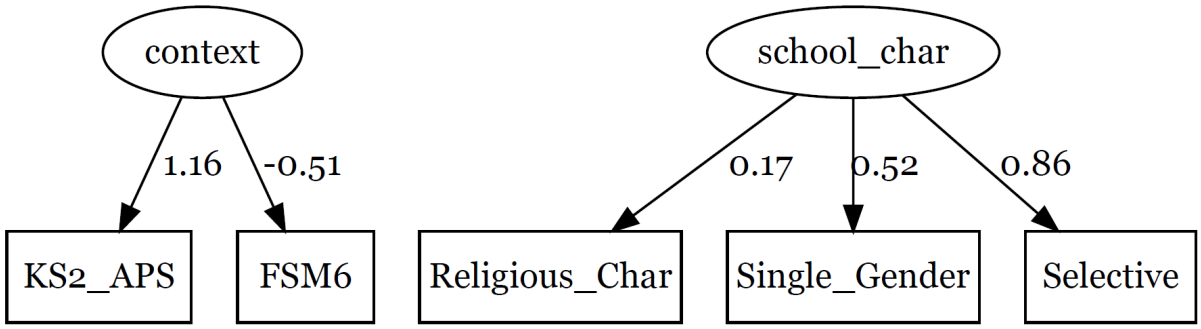


Figure 26. Model diagram for second Refined School Information Model.

Table 34. Model fit indices for second Refined School Information Model.

GFI	NNFI	CFI	RMSEA
0.998	0.992	0.997	0.033

Finally, the last variable with a low factor loading (Religious_Char) is removed, leaving the Final School Information Model (Figure 27) with very high values for all four fit measures (Table 32).

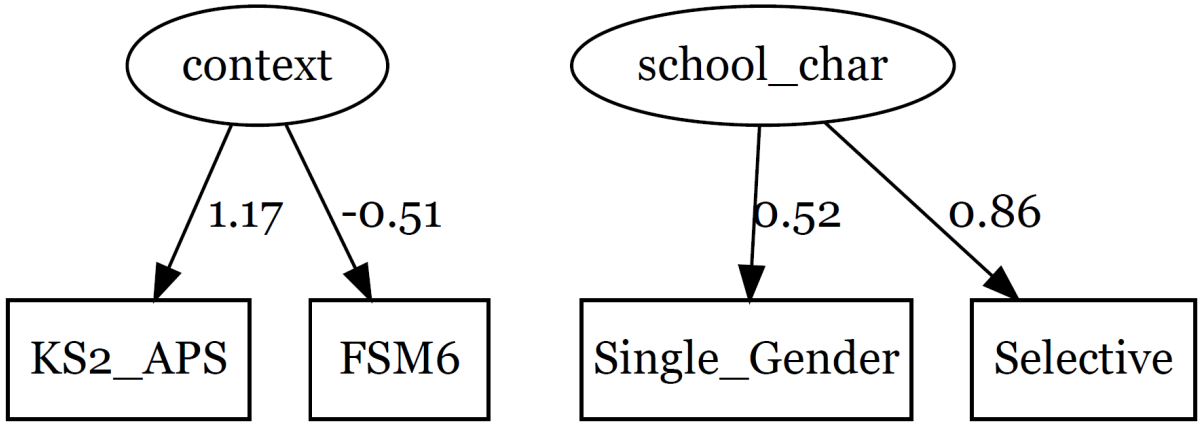


Figure 27. Model diagram for Final School Information Model.

Table 35. Model fit indices for Final School Information Model.

GFI	NNFI	CFI	RMSEA
0.999	0.995	0.999	0.034

Appendix 5. Details of School Clusters

The four clusters identified using agglomerative hierarchical clustering analysis of the six wider school information variables (with the exception of 'selective') for the 3125 mainstream state secondary schools in England have distinct characteristics:

Cluster 1 (n=1415)

The schools in this cluster, in general, have:

students with lower than national average prior attainment (mean KS2 APS is 102.15 compared to the national mean by school of 103.79)

higher than average percentages of students from disadvantaged backgrounds (mean FSM6 is almost 8% higher than the national mean of 26.5%)

slightly lower than average percentages of students with English as an additional language (mean EAL for schools in this cluster is 13.3% compared to 16.7% nationally)

higher than average percentages of students with special educational needs (mean SEN is over 18%, 3.7% higher than the national figure)

Generally, they:

are **non-selective** (no schools in cluster 1 are selective in their admissions policy), **very slightly less likely to have religious character** (51% of schools in cluster 1 compared to 54% nationally) and **much less likely to be single gender schools** (only 2% of schools are single gender, compared to a national figure of 11%).

With regard to headline school performance measures, they are generally **below average for all 6 metrics**.

Schools in this cluster have a mean **Attainment 8 value which is 5.63 below the national mean score**. This equates to **students, on average, attaining a quarter of a grade below students nationally with similar prior attainments** across the 8 subjects included in these measures (mean Progress 8 is -0.25).

The mean **percentage of students entered for the EBacc is nearly 10% less than the national average** of 39%.

The **mean Ebacc Average Points Score is the equivalent of over half a grade below national** (EBacc APS of 3.73 compared to the national figure of 4.33) and, on average, the **percentage of students gaining grade 5 or above in English and Maths is 11% lower than the national value** of 50%.

The **percentage of pupils continuing to a sustained education, apprenticeship or employment destination for this cluster is 2% lower than the national mean** of 93.9%.

Cluster 2 (n = 1292)

The schools in this cluster, in general, have:

students with above national average prior attainment (mean KS2 APS is 104.78 compared to the national mean by school of 103.79)

below average percentages of students from disadvantaged backgrounds (mean FSM6 is almost 9% lower than the national mean of 26.5%)

lower than average percentages of students with English as an additional language (mean EAL for schools in this cluster is 10.8% compared to 16.7% nationally)

below average percentages of students with special educational needs (mean SEN is over 11.7%, 2.8% lower than the national figure)

Generally, they:

are **non-selective** (only 2 schools in cluster 2 are selective in their admissions policy), **as likely to have religious character and to be single gender as all schools nationally.**

With regard to headline school performance measures, they are generally **above average for all 6 metrics**.

Schools in this cluster have a mean **Attainment 8 value which is 3.17 above the national mean score**.

The mean Progress 8 score for cluster 2 schools is 0.13, meaning that their students**, on average attain just over one eighth of a grade higher than their peers with similar prior attainment in each of the subjects included in the Progress 8 calculation**.

The mean **percentage of students entered for the EBacc is slightly higher than the national average** of 39% (cluster 2 % Ebacc Entry = 42.3%).

The **mean Ebacc Average Points Score is the equivalent of just less than a third of a grade above national** (EBacc APS of 4.64 compared to the national figure of 4.33) and, on average, the **percentage of students gaining grade 5 or above in English and Maths is 6.6% higher than the national value** of 50%.

The **percentage of pupils continuing to a sustained education, apprenticeship or employment destination for this cluster is 1.7% higher than the national mean** of 93.9%.

Cluster 3 (n = 255)

The schools in this cluster, in general, have:

students with lower than national average prior attainment (mean KS2 APS is 102.14 compared to the national mean by school of 103.79)

above average percentages of students from disadvantaged backgrounds (mean FSM6 is almost 15.4% higher than the national mean of 26.5%)

significantly higher than average percentages of students with English as an additional language (mean EAL for schools in this cluster is 66.3% compared to 16.7% nationally)

slightly below average percentages of students with special educational needs (mean SEN is 13.5%, 0.94% lower than the national figure)

Generally, they:

are **non-selective** (no schools in cluster 3 are selective in their admissions policy), **slightly less likely to have religious character** (49.4% of schools have religious character compared to 54.2% nationally) **and more likely to be single gender** (20% of schools are single gender compared to 11.3% of all schools nationally).

With regard to headline school performance measures, they present a **more complex picture, being generally above average for 4 of the metrics and below average for the remaining 2**.

Schools in this cluster have a mean **Attainment 8 value which is slightly below the national mean score** (48.9 compared to 49.5 nationally). This equates to **students, on average, attaining a quarter of a grade higher than students nationally with similar prior attainment** across the 8 subjects included in these measures (mean Progress 8 is 0.26).

The mean **percentage of students entered for the EBacc is significantly higher than the national average** of 39% (% Ebacc Entry for cluster 3 is 50.1%).

The **mean Ebacc Average Points Score is very close to the national average** (EBacc APS of 4.38 compared to the national figure of 4.33) and, on average, the **percentage of students gaining grade 5 or above in English and Maths is also very close to the national value** of 50%.

The **percentage of pupils continuing to a sustained education, apprenticeship or employment destination for this cluster is 1.4% lower than the national mean** of 93.9%.

Cluster 4 (n = 163)

The schools in this cluster, in general, have:

students with significantly higher than national average prior attainment (mean KS2 APS is 112.75 compared to the national mean by school of 103.79)

significantly below average percentages of students from disadvantaged backgrounds (mean FSM6 is almost 20% lower than the national mean of 26.5%, with only 7% of students classed as disadvantaged).

around average percentages of students with English as an additional language (mean EAL for schools in this cluster is 16.4% compared to 16.7% nationally)

significantly below average percentages of students with special educational needs (mean SEN is 6.2%, 8.3% lower than the national figure)

Generally, they:

are **selective** (161 of the 163 schools in cluster 4 are selective in their admissions policy), **significantly more likely to have religious character** (83.4% of schools have religious character compared to 54.2% nationally) **and more likely to be single gender** (71% of schools are single gender compared to 11.3% of all schools nationally).

With regard to headline school performance measures, they are generally **significantly above average for all six metrics**.

Schools in this cluster have a mean **Attainment 8 value which is well above the national mean score** (74.2 compared to 49.5 nationally). This equates to **students, on average, attaining over half of a grade higher than students nationally with similar prior attainment** in each of the 8 subjects included in these measures (mean Progress 8 is 0.57).

The mean **percentage of students entered for the EBacc is double the national average** of 39% (cluster 4 % Ebacc Entry = 78%).

The **mean Ebacc Average Points Score is significantly above national average** (EBacc APS of 7.00 compared to the national figure of 4.33).

On average, the **percentage of students gaining grade 5 or above in English and Maths** (94%) **is also very much higher than the national value** of 50%.

The **percentage of pupils continuing to a sustained education, apprenticeship or employment destination for this cluster is 99.1%, which is higher than the national mean** of 93.9%.

Appendix 6. Focus Group 3 information (alternative approaches to SPMs)

White Rose Social Sciences DTP

Economic and Social Research Council

Secondary School Performance Measures in England;
their effectiveness,
their effects,
and an exploration of alternative approaches

An ideas document for

Research Focus Group discussions

June 2025

University of Sheffield | Information School

A member of the iSchools network

Introduction

This document is a brief summary of the ideas which emerged from research focus group discussions conducted online between August 6, 2024 and September 17 (1st round) 2024 and September 20, 2024 and October 29, 2024 (2nd round). These groups included researchers, teachers, school leaders, governors, union officers and parents and formed part of a PhD project exploring secondary school performance measures (SPMs) in England; their effectiveness, their effects and an exploration of alternative approaches.

The ideas here are presented to inform discussion, to prompt consideration of the refinement of these concepts and to stimulate debate around potential new ideas related to alternative approaches to SPMs.

The 1st round of focus group discussions explored the purposes of SPMs, the effectiveness of SPMs, their wider effects. Semi- structured discussions were conducted online, with all discussions recorded. The transcripts were then thematically analysed, with the aim of identifying the key ideas and themes which emerged from the discussion. A brief summary of these themes and subthemes can be found in table 1.

Table 1. Themes and subthemes from analysis of Focus group round 1 discussions

The purpose of SPMs
Accountability (general)
Democratic good (e.g. transparency)
Inform parents & choice
School improvement
Enacting government policy
The effectiveness of SPMs
Issues with interpretation (related to effectiveness)
Different audiences have different needs and uses of the measures
Not focusing on the right things
Out of date
Oversimplified
Too complex
Zero sum game
Issues with the metrics (related to effectiveness)
Unfair or biased
The wider effects of SPMs
Impact of generating SPMs
Gaming behaviours
Loss of autonomy (or agency)
School segregation
Too much focus on measures
Operation of SPMs
Link with regulatory framework and legislation

	Part of the culture of schools
	Publication
	Work pressure and High Stakes

The 2nd round of group sessions built on the 1st round by considering a wide range of alternative approaches to SPMs. The discussions were rich and covered a broad range of subjects, from the potential underlying philosophies of SPMs and their related theory of change to examples of alternative approaches from other jurisdictions. For the purpose of this document, I am focusing in particular on the ideas for alternative approaches to SPMs which came from the group discussions.

These are summarised in table 2.

Table 2. Themes and subthemes from analysis of Focus group round 2 discussions

New approaches (retaining the same/similar metrics)	
	Change of locus of accountability
	Compliance (eg with min standard)
	How diverse data can be linked together to priorities eg national, regional or local
	How well the school fits what a parent is looking for
	Information focusing on what causes the differences between school outcomes
	Inspection as verification of self-evaluation
	Many diverse measures and randomly choose which to focus on (to address 'gaming')
	SPMs based around national priorities
	SPMs based around school priorities
New measures/metrics	
	Academic performance
	Budget management
	Contextualised data
	Evaluation of leadership
	How schools support their local community (including other schools)
	Inclusivity
	Safeguarding
	Staff development (wellbeing, recruitment)
	Student attitudes survey

Student development and wellbeing
New approaches to the communication of SPMs
Addressing the misinterpretation of SPMs
The communication of SPMs (e.g. data viz etc)
Whether all data is made public
Different Audiences for School Data

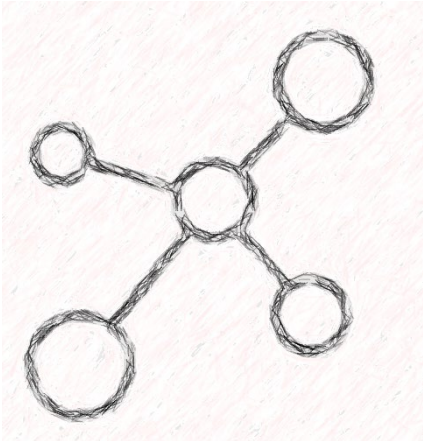
What follows is a brief outline of each of these alternative approaches. The information provided should not, in any way, be considered to be either comprehensive or complete, instead it is intended to act as a stimulus for further discussion.

New approaches (retaining the same/similar metrics)

Section contents

New approaches (retaining the same/similar metrics)
Change of locus of accountability
Compliance indicators (eg with min standard)
How diverse data can be linked together to priorities eg national, regional or local
How well the school fits what a parent is looking for
Information focusing on what causes the differences between school outcomes
Inspection as verification of self-evaluation
Many diverse measures and randomly choose which to focus on (to address 'gaming')
SPMs based around national priorities
SPMs based around school priorities

The ideas here present various alternative approaches to SPMs. The intention is to consider how **existing** metrics might be used in **new and innovative ways** in an attempt to address some of the issues previously identified.



New approach (retaining the same/similar metrics)

Change of locus of accountability

In Practice:

Shift the accountability focus from individual schools to groups of schools (e.g., trusts, local clusters, or area-based collectives). These groups might then be responsible for ensuring quality and equitable outcomes across the collective, rather than at single institutions.

What it Might Look Like:

Joint performance metrics for a group of schools.

Some shared improvement plans and resource allocation.

Accountability frameworks assessing group-level effectiveness, equity and collaboration.

Examples:

Greater Manchester Challenge (2008–2011): area-wide improvement strategy.

Collective impact models in the U.S. (e.g., StriveTogether network).

Some Multi-Academy Trusts already operate collective accountability mechanisms.

Education Policy Institute's Effective School Groups Data Tool (https://epi.org.uk/effective-school-groups-data-tool/). An online tool allowing users to benchmark performance between different school groups, such as Multi Academy Trusts, using data up to 2023.

Barriers:

Legal framework and funding structures are school-based.

Risk of diffusion of responsibility.

Need for strong collaboration infrastructure.

New approach (retaining the same/similar metrics)

Compliance indicators (e.g. with minimum standard)

In Practice:

Rather than ranking all schools, ensure each meets minimum thresholds (e.g., safeguarding, curriculum breadth, inclusion), like regulatory compliance checks.

What it Might Look Like:

'Meets standard' or 'Does not meet standard' labels.

Public reporting on whether schools pass a checklist of basic educational expectations.

Examples:

New Zealand's Education Review Office includes a focus on compliance.

Ofsted's minimum 'floor standard' used to play this role to some extent.

Barriers:

Risk of complacency if targets are too low.

Political pressure to differentiate between schools.

Possible neglect of excellence and innovation.

New approach (retaining the same/similar metrics)

Linking diverse data together to national, regional or local priorities

In Practice:

Combine data sources (e.g., school performance, health, wellbeing, attendance, neighbourhood deprivation) to align with strategic priorities at various governance levels.

What it Might Look Like:

Interactive dashboards showing local context and contributions to wider goals (e.g., climate literacy, child wellbeing).

Customised accountability reports for different regions.

Examples:

Scotland's National Improvement Framework includes local authority data use.

UN Sustainable Development Goals have inspired education indicators in some nations.

Barriers:

Data interoperability and privacy concerns.

Complexity in aligning different datasets.

Difficulty in comparing schools across diverse contexts.

New approach (retaining the same/similar metrics)

How well the school fits what a parent is looking for

In Practice:

Shift from a 'best school' narrative to 'best fit' based on parental values and preferences (e.g., arts, academic, inclusion, discipline, ethos).

What it Might Look Like:

School profiles with qualitative descriptions.

Surveys or compatibility tools (like dating app-style matching).

Use of case studies, parent feedback and student voice.

Examples:

School guides in Finland provide detailed, narrative information.

Some U.S. cities use 'GreatSchools' profiles that allow filtering by values and offerings.

Barriers:

Hard to standardise subjective information.

Risk of reinforcing segregation or social stratification.

Parental decisions may be influenced by social reputation, not values fit.

New approach (retaining the same/similar metrics)

Focus on what causes differences in school outcomes

In Practice:

Evaluate contextual drivers of outcomes rather than just the outcomes themselves—e.g., intake characteristics, resources, teacher turnover, parental engagement.

What it Might Look Like:

Diagnostic-style reporting that identifies contributing factors to performance.

Visual causal maps or explanatory narratives in performance tables.

Examples:

Education Endowment Foundation (EEF) Toolkit focuses on causes and impacts.

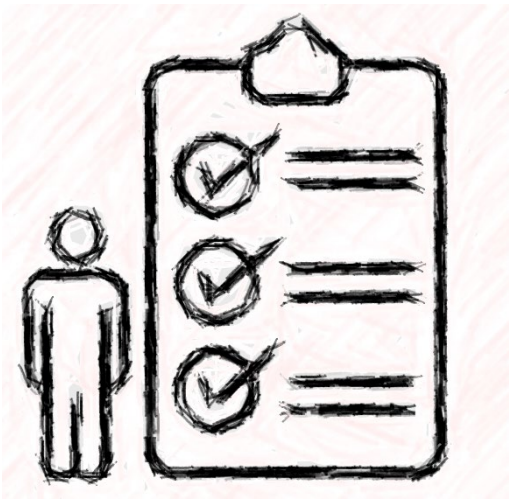
OECD's PISA contextual questionnaires explore background factors.

Barriers:

Complexity in communication.

May be seen as making excuses.

Requires high data literacy among users.



New approach (retaining the same/similar metrics)

Inspection as verification of self-evaluation

In Practice:

Schools conduct deep self-assessments; external inspections validate, challenge, or endorse them rather than inspect from scratch.

What it Might Look Like:

Published self-evaluation reports.

Ofsted-style inspectors act more like auditors or mentors.

Greater emphasis on internal reflective practice.

Examples:

Scotland's inspection model emphasises school-led self-evaluation.

Some international baccalaureate schools use similar models.

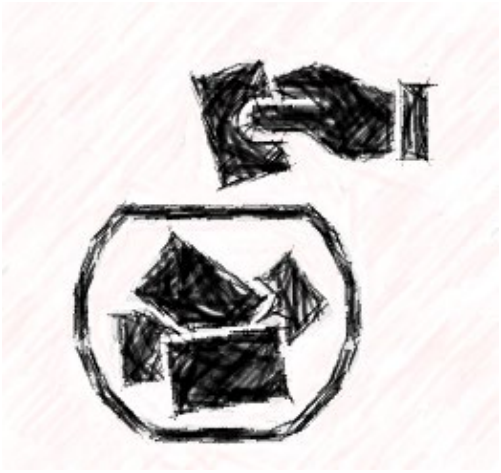
Barriers:

Trust and credibility—public may distrust self-evaluation.

Risk of gaming or inflated self-reports.

Time and capacity needed in schools for honest evaluation.

New approach (retaining the same/similar metrics)



Many diverse measures - randomly choose focus

In Practice:

Use a broader basket of performance indicators and only a rotating/randomised subset is used to determine formal judgments or funding. This reduces the incentive to game specific metrics.

What it Might Look Like:

'Assessment roulette' where only a sample of metrics are activated each year.

Wider accountability portfolio than used for judgment.

Examples:

Business world: some anti-gaming techniques in balanced scorecard systems.

Education: elements of this approach have been suggested, not widely implemented.

Barriers:

Seen as unfair or unpredictable.

Political opposition due to reduced transparency.

Requires very robust design to avoid anomalies.

New approach (retaining the same/similar metrics)

SPMs based around national priorities

In Practice:

Align performance measures with government-set strategic goals—e.g., improving literacy, reducing exclusions, environmental education.

What it Might Look Like:

Targets reflect national educational ambitions and social priorities.

Schools scored on their contribution to these goals, possibly using mixed methods.

Examples:

Singapore's national education goals guide assessments.

Wales' Curriculum for Wales and its focus on well-rounded citizens.

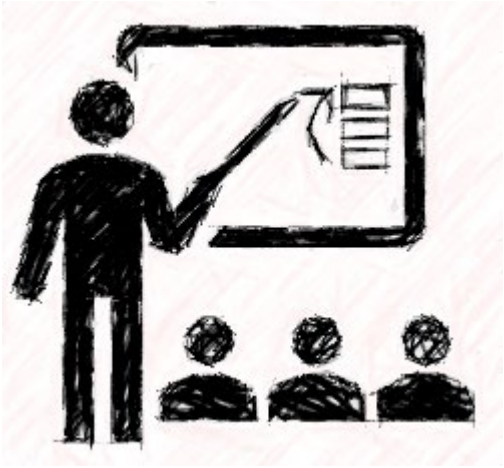
Barriers:

Risk of politicisation of priorities.

National goals may not suit local needs.

May lead to short-termism or narrow focus depending on political cycle.

New approach (retaining the same/similar metrics)



SPMs based around school priorities

In Practice:

Allow schools to develop and report on self-defined priorities aligned to their context, values and community aspirations.

What it Might Look Like:

Each school submits a 'public contract' or 'mission dashboard.'

Performance judged on progress toward their own declared objectives.

Examples:

Some international school accreditation models (e.g., CIS, IB).

Early models of co-operative schools in the UK emphasized local goals.

Barriers:

Comparison between schools becomes difficult.

Risk of setting vague or unambitious targets.

Requires a cultural shift in trust and accountability.

New measures/metrics

Section contents

New measures/metrics
Academic performance
Budget management
Contextualised data
Evaluation of leadership
How schools support their local community (including other schools)
Inclusivity
Safeguarding
Staff development (wellbeing, recruitment)
Student attitudes survey
Student development and wellbeing

The ideas here present various concepts for **new metrics/measures** within SPMs. The intention is to consider how new measures/metrics might be used within SPMs to address any issues identified and to bring about improvement.

New measures/metrics

Academic performance

In Practice:

Still central, but could be redefined—e.g., focusing on progress, subject breadth or deeper learning rather than raw attainment.

What it Might Look Like:

Measures of growth (like Progress 8) over time.

Depth of understanding assessments, e.g. project-based or portfolio-based.

Examples:

Progress 8 in England is already a contextualised progress measure.

International Baccalaureate (IB) assesses more holistic academic achievement.

New Zealand's NCEA includes diverse academic pathways.

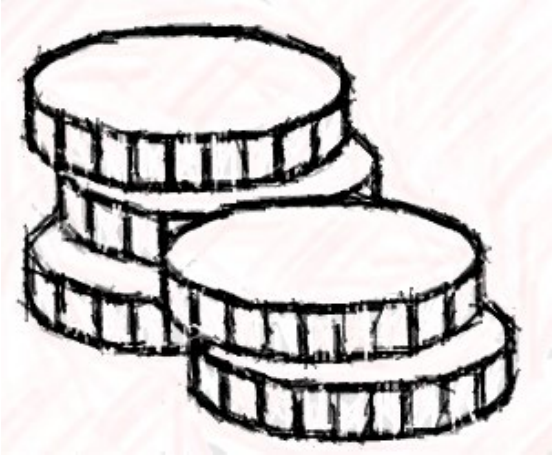
Barriers:

Pressure to reduce to simplistic rankings.

Test-driven systems can still distort learning.

Challenges in implementing more complex assessments at scale.

New measures/metrics



Budget management

In Practice:

Monitor how efficiently and effectively schools use public funds—linking financial responsibility to educational outcomes and long-term sustainability.

What it Might Look Like:

Transparent annual financial statements.

Value-for-money audits.

Metrics showing per-pupil spend linked to outcomes.

Examples:

UK academies already publish financial reports.

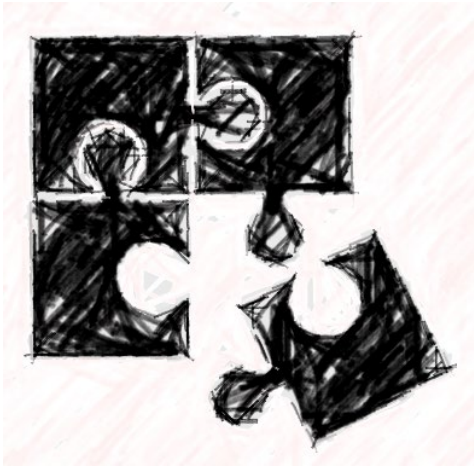
Ofsted reviews sometimes include commentary on budget use.

Barriers:

Financial literacy varies among school leaders.

Risk of cost-cutting becoming more valued than quality.

Context affects spending needs (e.g. SEND provision).



New measures/metrics

Contextualised data

In Practice:

Adjust performance data to reflect contextual factors (e.g., deprivation, prior attainment, %EAL, %SEN), acknowledging challenges and reducing bias.

What it Might Look Like:

Dashboards showing adjusted and unadjusted outcomes.

'Context scorecards' for fair comparisons.

Peer school benchmarking based on intake characteristics.

Examples:

FFT Aspire tools in the UK.

Scotland's Insight tool.

Progress 8 is a basic contextual measure, though controversial.

Barriers:

Difficult to design fair adjustment models.

Public may distrust adjusted scores.

Risk of excusing underperformance rather than addressing need.

New measures/metrics

Evaluation of leadership

In Practice:

Assess leadership effectiveness based on vision, school culture, staff development, stakeholder feedback and results over time.

What it Might Look Like:

360° feedback from staff, students, parents.

External peer review of leadership practices.

Tracking impact of leadership turnover and decisions.

Examples:

Some MATs and international schools use leadership reviews.

NPQ frameworks focus on leadership development and evaluation.

Barriers:

Leadership impact is indirect and slow to emerge.

Hard to separate leader effects from wider school/team factors.

Sensitive to measure—risk of politicisation or defensiveness.

New measures/metrics

How schools support their local community

In Practice:

Measure a school's role as a civic actor: outreach, shared facilities, support for feeder schools or struggling neighbours.

What it Might Look Like:

KPIs on community partnerships.

Evidence of system leadership or community involvement.

Contribution logs or case studies.

Examples:

Teaching School Hubs in England support others.

Community school models in the U.S. and Finland.

Co-operative schools include community impact in governance.

Barriers:

Not easy to quantify.

Time-consuming and not always aligned with performance incentives.

Risk of tokenism without structural support.

New measures/metrics

Inclusivity

In Practice:

Track how well schools include and serve all students—e.g., SEND, EAL, FSM, LAC (Looked After Children), etc.

What it Might Look Like:

Indicators of diversity in student intake.

SEND outcomes and inclusion strategies.

Disproportionality in exclusions or access to enrichment.

Examples:

SEND Green Paper (UK) highlights inclusion metrics.

UNCRPD framework encourages inclusive education monitoring.

.

Barriers:

Inclusion varies by need—can be misunderstood.

High-needs schools may be unfairly penalised.

Need for rich qualitative data alongside metrics.

New measures/metrics

Safeguarding

In Practice:

Ensure schools are safe and nurturing—beyond minimum compliance, focus on culture, training and responsiveness to concerns.

What it Might Look Like:

Safeguarding audits.

Incident tracking and resolution rates.

Student and parent surveys on safety and trust.

Examples:

Ofsted already includes safeguarding judgments.

Wales' Estyn inspections often foreground safeguarding.

Barriers:

Sensitive and potentially legally complex.

Risk of schools hiding problems to avoid reputational harm.

Measurement can feel bureaucratic unless handled well.

New measures/metrics

Staff development (wellbeing, recruitment…)

In Practice:

Monitor investment in staff development, workload management, retention, morale and CPD engagement

What it Might Look Like:

Annual staff surveys.

CPD hours per staff member.

Retention and workload statistics.

Examples:

England's Wellbeing Charter for Education.

OECD TALIS survey assesses teacher conditions globally.

Barriers:

Time and budget constraints.

Risk of 'box-ticking' CPD.

Honest reporting requires a culture of trust.

New measures/metrics

Student attitudes

In Practice:

Systematic collection of student voice on learning, safety, belonging, aspirations and school culture.

What it Might Look Like:

Annual student feedback surveys.

Summary scores published or used in inspection.

Longitudinal tracking of engagement and perceptions.

Examples:

The PASS survey (Pupil Attitudes to Self and School) in the UK.

International surveys like PISA Wellbeing and My Voice.

School climate surveys in some U.S. districts.

Barriers:

Requires careful design and anonymity.

Risk of manipulation or pressure to give good responses.

May not influence high-stakes accountability.

New measures/metrics

Student development and wellbeing

In Practice:

Track holistic student development: mental health, emotional intelligence, personal growth, co-curricular involvement.

What it Might Look Like:

Wellbeing scores, resilience indicators.

Participation in enrichment and community activities.

Support service uptake and outcomes.

Examples:

Wales' new curriculum foregrounds wellbeing.

Australia's National Student Wellbeing Framework.

Some schools use the Leuven Scale for wellbeing.

Barriers:

Hard to define and quantify.

Requires a shift in what is valued as 'educational success.'

Tension with academic pressures and inspection regimes.

New approaches to the communication of SPMs

Section contents

New approaches to the communication of SPMs
Addressing the misinterpretation of SPMs
The communication of SPMs (e.g. data viz etc)
Whether all data is made public
Different Audiences for School Data

The various ideas here present new approaches for how SPMs are communicated. The intention is to consider approaches which might improve their use by all audiences, for example in making them more accessible and understandable and communicating what they do (and what they do not) tell us.

New approaches to the communication of SPMs

Addressing the Misinterpretation of SPMs

In Practice:

Provide clear explanations and contextual information alongside performance measures to reduce misinterpretation.

What it Might Look Like:

Supplement data with caveats, confidence intervals and explanations of what the measures do—and do not—indicate.

Examples:

Ofsted's blog posts occasionally clarify misunderstandings about inspection judgements or progress measures (e.g. Education Inspection Blog).

FFT Aspire provides contextualised dashboards for schools with explanation text for each metric, helping staff interpret results responsibly.

Education Datalab blog publishes accessible analysis that helps the public and policymakers understand the limitations of school performance metrics.

Barriers:

Limited data literacy among audiences.

Resistance from media or stakeholders who benefit from simplified rankings.

Time and expertise required to produce nuanced explanations.

The Communication of SPMs

(e.g. Data Visualisation)

In Practice:

Redesign performance reports using intuitive, accessible visuals and narratives tailored to non-specialist audiences.

What it Might Look Like:

Interactive dashboards, colour-coded risk indicators, trend lines and contextual overlays.

Examples:

Ofsted's Inspection Data Summary Report (IDSR) uses visual tools like scatter plots and bar charts to summarise complex performance information for inspectors.

SchoolDash provides interactive data maps and time series visualisations tailored for public understanding and journalistic use.

The Education Policy Institute (EPI) publishes visual reports showing regional and socio-economic disparities in attainment using accessible charts and graphs.

Barriers:

Technical capacity in schools and DfE.

Risk of oversimplification or new forms of bias in visual formats.

Cost and time of implementation.

New measures/metrics

Whether All Data Is Made Public

In Practice:

Review and refine what data is shared publicly versus internally, based on purpose and risk of misuse.

What it Might Look Like:

A two-tier system: some data for internal improvement (e.g. teacher turnover, contextual risks) and some for accountability or transparency.

Examples:

Scotland’s Insight tool is used internally by schools and local authorities for self-evaluation, with only high-level data released to the public.

Wales' move away from public school league tables—instead using the My Local School website with balanced information.

Department for Education's Performance Tables include suppression rules (e.g., 'SUPP') to avoid disclosure of small or sensitive datasets.

Barriers:

Tensions between transparency and the potential for reputational harm.

Legal and ethical questions around data privacy.

Who decides what is withheld or shared.

New measures/metrics

Different Audiences for School Data

In Practice:

Tailor information and presentation styles to meet the needs and priorities of different user groups.

What it Might Look Like:

Multiple versions of performance reports: operational detail for schools, strategic summaries for governors, clear comparisons for parents.

Examples:

GovernorHub Knowledge offers clear summaries of performance data tailored for school governors' strategic decision-making.

Parent View (Ofsted) gives simplified insights based on parent surveys, reflecting parental priorities rather than technical indicators.

FFT Aspire's differentiated reports for teachers, senior leaders and governors each have distinct framing and content based on user needs.

Barriers:

Resource demands to produce differentiated outputs.

Risk of inconsistency in interpretation across groups.

Potential for misalignment or conflicting narratives between versions.

Appendix 7. Advertisement for participants

University of Sheffield | Information School

A member of the TSchools network

School Accountability Study: Research Participants Needed

Would you like to contribute to research which aims to help shape the future of school accountability?

School performance measures and 'league tables' were first introduced 30 years ago. Along with Ofsted, they form a central element of the English school accountability system today. This project is investigating the use, effectiveness and effects of secondary school performance measures in England. It also aims to suggest new approaches to school accountability which are responsible, fair and more fit for purpose.

We are looking for volunteers to discuss issues related to secondary school performance measures in England. Volunteers will also be invited to take part in a series of sessions in which new, alternative approaches to secondary school performance measures will be developed.

We are looking for:

- Teachers
- Researchers
- Policy Makers
- School Leaders
- School Governors
- Parents/carers of children in years 5 or 6 of primary school
- Parents/carers of secondary age children

Each participant will be asked to be involved in:

- One initial group discussion session*
- Three separate development sessions* in which alternative approaches will be developed

Location:

- It is expected that sessions will be online

Contact:

If you are interested in taking part in this study, or would like more information, please contact

Ian Widdows

icwiddows1@sheffield.ac.uk

**It is anticipated that each session will last between 1-2 hours (tbc)*

Appendix 8. Ethics Approval Letter

The University Of Sheffield.

Downloaded: 13/09/2023
Approved: 14/06/2023

Ian Widdows
Registration number: 220257097
Information School
Programme: Information Studies (Social) (PhD/Info Studies (Social) FT) - INFR33

Dear Ian

PROJECT TITLE: Secondary school performance measures in England - their effectiveness, effects and an exploration of alternative approaches
APPLICATION: Reference Number 052314

On behalf of the University ethics reviewers who reviewed your project, I am pleased to inform you that on 14/06/2023 the above-named project was **approved** on ethics grounds, on the basis that you will adhere to the following documentation that you submitted for ethics review:

- University research ethics application form 052314 (form submission date: 19/05/2023); (expected project end date: 31/12/2025).
- Participant information sheet 1119575 version 3 (21/04/2023).
- Participant information sheet 1121829 version 1 (12/05/2023).
- Participant consent form 1119577 version 1 (27/03/2023).
- Participant consent form 1121830 version 1 (12/05/2023).

If during the course of the project you need to deviate significantly from the above-approved documentation please inform me since written approval will be required.

Your responsibilities in delivering this research project are set out at the end of this letter.

Yours sincerely

Peter Bath
Ethics Administrator
Information School

Please note the following responsibilities of the researcher in delivering the research project:

- The project must abide by the University's Research Ethics Policy: https://www.sheffield.ac.uk/research-services/ethics-integrity/policy
- The project must abide by the University's Good Research & Innovation Practices Policy: https://www.sheffield.ac.uk/polopoly_fs/1.671066!/file/GRIPPolicy.pdf
- The researcher must inform their supervisor (in the case of a student) or Ethics Administrator (in the case of a member of staff) of any significant changes to the project or the approved documentation.
- The researcher must comply with the requirements of the law and relevant guidelines relating to security and confidentiality of personal data.
- The researcher is responsible for effectively managing the data collected both during and after the end of the project in line with best practice, and any relevant legislative, regulatory or contractual requirements.