

Taylor, Michael Bryan William (2026) *The debatable lands: Exploring the relationship between education policy intention and assessment practice across England and Scotland.* PhD thesis.

https://theses.gla.ac.uk/86351/

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Enlighten: Theses
https://theses.gla.ac.uk/
research-enlighten@glasgow.ac.uk

The debatable lands: Exploring the relationship between education policy intention and assessment practice across England and Scotland

Michael Bryan William Taylor
B.Sc., MSc.

Submitted in fulfilment of the requirements for
the Degree of Doctor of Philosophy

School of Education,
College of Social Sciences
University of Glasgow
June 2026

Abstract

The negative consequences of accountability-driven assessment policies for teaching and learning in both England and Scotland are well documented across two decades of reform. These practices persist despite rhetorical commitments to evidence-informed practice. This thesis explores why such policies persist, investigating the stated purposes and enacted realities underlying assessment reform in both countries between 1998 and 2023.

Semi-structured interviews with policy influencers were combined with case studies of two secondary schools, one in Liverpool and one in Dundee. Thematic analysis drew on Ball's policy enactment framework, Le Grand's knights and knaves typology, and Looney et al.'s assessment identity as a sensitising concept. Together, these frameworks traced both the national construction of assessment policy and its effects on teachers' conceptions of their assessment roles.

Deliberate political choice, rather than a failure of knowledge translation, offers an explanation of identified misalignments between assessment research, policy and practice. National policy influencers demonstrated explicit awareness of the detrimental consequences of accountability-intensive reforms. Nevertheless, they proceeded when such reforms served political objectives, a pattern characterised here as unintended but not unknown. Teachers' conceptions of assessment, commonly shaped by their experiences in prior policy environments, create temporal constraints that may delay changes to assessment practice in response to policy reform. Divergent accountability structures in England and Scotland are nonetheless convergent on similarly negative effects from accountability processes.

This thesis provides new empirical evidence that the policy-practice gap may be politically influenced. In doing so, it bridges a persistent divide in the literature between national policy analysis and teachers' professional understanding of assessment.

Table of Contents

Abstract **iii**
List of tables **x**
List of figures **xi**
Acknowledgement **xii**
Author's declaration xiii

Definitions/Abbreviations **xiv**
1 Introduction **1**
1.1 Research gap and rationale 4
1.2 Significance and contribution 7
1.3 Researcher positioning 8
1.4 Research approach 9
1.5 Thesis structure 12
1.6 Chapter summary 13

2 Literature review **15**
2.1 Chapter introduction 15
2.2 The working definition of assessment in this thesis 16
2.2.1 Formative assessment as a developing concept 17
2.2.2 High-stakes assessment and its consequences 27
2.2.3 Research evidence and the policymaking process 35
2.3 Teachers and the enactment of assessment policy 36
2.3.1 The development of assessment literacy 37

2.3.2 Teacher beliefs and classroom assessment 39

2.3.3 Conceptions of teacher assessment identity 44

2.3.4 The enactment framework and teacher mediation 46

2.3.5 The temporal dimension of teacher formation 56

2.4 Accountability and educational change in England and Scotland 57

2.4.1 Competition and marketised school structures 57

2.4.2 International benchmarking and policy borrowing 62

2.4.3 Mimetic isomorphism in assessment policy 63

2.4.4 Inspection regimes and accountability mechanisms 64

2.4.5 Motivational assumptions in accountability design 67

2.5 Policy context in England and Scotland 1998 to 2023 72

2.5.1 The 1998 starting point and the rationale for comparison. 72

2.5.2 Assessment policy in England under New Labour, 1997 to 2010 74

2.5.3 Assessment policy in England under the coalition and Conservative governments 2010 to 2023 81

2.5.4 Assessment policy in Scotland following devolution 1998 to 2014 90

2.5.5 Assessment policy in Scotland from the OECD review to 2023 94

2.6 Chapter summary 101

3 Methodology 103

3.1 Introduction 103

3.2 Ontology and epistemology 104

3.2.1 Researcher positionality 106

3.3 Theoretical framework 107

3.3.1 Structuralism and discourse 110

3.3.2 Critical theory in educational policy research 113

3.4 Research design 115

3.4.1 Case study approach 117

3.4.2 Case selection 118

3.4.3 Policy influencers 119

3.4.4 Data collection 125

3.4.5 Pilot study 130

3.4.6 Data analysis 133

3.5 Ethical considerations 139

3.6 Research limitations 141

3.7 Chapter summary 144

4 Findings 146

4.1 Introduction 146

4.2 Common themes within this chapter 147

4.3 England 151

4.3.1 What are the diverse purposes of assessment processes in education in England? 151

4.3.2 Stakeholder perspectives on assessment processes and their impact in England 163

4.3.3 To what extent does assessment policy impact practice in England? 174

4.3.4 In what ways does political thinking impact assessment policymaking in England? 186

4.3.5 What has the impact of COVID-19 been on assessment processes in England? 190

4.4 Scotland 195

4.4.1 What are the diverse purposes of assessment processes in education in Scotland? 195

4.4.2 Perceptions of assessment processes in Scotland 203

4.4.3 Assessment policy enactment in Scotland and the gap between intentions and practice 206

4.4.4 Political pressures and assessment policymaking in Scotland 217

4.4.5 Pandemic disruption and systemic vulnerabilities in Scottish Assessment Policy 227

4.4.6 Assessment reform as contested policy space 232

4.5 Chapter summary 234

5 Discussion 236

5.1.1 Introduction 236

5.1.2 Policy actors deploy accountability mechanisms, with awareness of detrimental consequences, for political expediency 236

5.1.3 Divergent accountability cultures in England and Scotland shape differential trust in policy mediators 242

5.1.4 Politicians selectively appropriate international assessment data to legitimise ideologically driven policy reforms 247

5.1.5 Teacher professional agency is constrained by temporal policy effects and professional formation under previous policy frameworks 252

5.1.6 Chapter summary 260

6 Conclusion 264

6.1 Response to research questions 264

6.1.1 The purposes of assessment policy 264

6.1.2 Alignment between policy and research 265

6.1.3 Teacher, school leader and policy influencer perceptions of assessment policy 266

6.2 Original contributions to knowledge 267

6.2.1 The unintended but not unknown pattern 269

6.2.2 Temporal policy effects and professional identity 276

6.3 Implications and recommendations 282

6.3.1 Professional formation and reform enactment 282

6.3.2 Political expediency and pedagogical evidence 283

6.3.3 Divergent accountability and policy mediation 284

6.3.4 Conceptualising policy divergence 285

6.4 Research limitations and delimitations 287

6.5 Final reflections 287

Appendices **291**

Appendix A: Interview schedule for Policy Influencer 291

Appendix B: Interview schedule for educational professionals 294

Appendix C: Consent form for policy influencers 299

Appendix D: Consent form for educational professionals 300

Appendix E: Participant information sheet 301

Appendix F: Ethics approval form 302

List of References **304**

List of tables

Table 2.1 Policy actor types and their activities within policy enactment (adapted from Ball et al. (2011)) 49
Table 3.2: Rationale for categories of policy influencers. 120
Table 3.3: Rationale for categories of school professionals. 124
Table 3.4: The roles of classroom participants in assessment. Adapted from Wiliam and Thompson (2007). 129
Table 4.5: Definitions for the multiple purposes of assessment processes. 148
Table 6.6: The categories and descriptions of public actors, including empirical evidence. Categories and behaviours adapted from Le Grand (2006). 280

List of figures

Figure 1: Components of Teacher Assessment Identity. Adapted from Looney et al. (2018). 45
Figure 2: A visual conceptualisation of human agency and motivation taken from Le Grand (2006, p. 16). 69
Figure 3: A stimulus image of a classroom. Image sourced from: www.unsplash.com. Accessed: 21/6/2021 128
Figure 4: Photograph of students and their teacher in a conversation. Obtained with permission for use from Pixabay.com 132
Figure 5: Framework for conceptualising teacher assessment identity extracted from Looney et al. (2018). 137
Figure 6: Framework exploring the intersection of policy intentions and anticipated consequences. 272

Acknowledgement

Ten years is a long time to continue on a single piece of work. When I started this thesis, I could not have known what those years would bring, and I could not have finished it without the people and the organisations that carried me through.

I came to this as someone from a background where university was not an obvious path. My mum Diane is the reason I had the audacity to try. She filled my childhood with second-hand books and told me, again and again, that hard things were worth doing. Everything that follows began there.

The thesis itself belongs in part to the schools that welcomed me, and to the many policy influencers who gave their time and their expertise without asking anything in return. I am deeply grateful to them.

I want to thank Chris and Louise, who supervised this work for the whole of its ten years and never let me give up on it. I am grateful too to the wider University of Glasgow community, Adele, Caitlin, and to the friends and colleagues for all the proofreading and advice.

These were not easy years. I lost my eyesight for periods; I worked full time; I studied remotely and worked through the pandemic and a good deal of life besides. The journey of producing this thesis has outlived many people whom I would have loved to show the final product.

Thank you to Bayley, my best friend, who walked beside me through the worst of the thinking. Some of the knots in this thesis only came loose on those long walks.

Keiron lived through all of it with me and put up with a partner for whom the PhD was forever the reason we could not do the things we wanted to do. Thank you for everything.

Author's declaration

I declare that, except where explicit reference is made to the contribution of others, this dissertation is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any institution.

Signature:



Printed name: Michael Taylor

Definitions/Abbreviations

AiFL	Assessment is for Learning
APP	Assessing Pupil Progress
BERA	British Educational Research Association
BGE	Broad General Education
CAGs	Centre-Assessed Grades
CfE	Curriculum for Excellence
CPD	Continuing Professional Development
CVA	Contextual Value Added
DES	Department of Education and Science
DfE	Department for Education
DfEE	Department for Education and Employment
EBacc	English Baccalaureate
ECT	Early Career Teacher
EPI	Education Policy Institute
GCSE	General Certificate of Secondary Education
HMI	His Majesty's Inspectorate
HMIE	His Majesty's Inspectorate of Education
MATs	Multi-Academy Trusts
NCSL	National College for School Leadership
NQTs	Newly Qualified Teachers
OECD	Organisation for Economic Co-operation and Development
Ofqual	Office of Qualifications and Examinations Regulation
Ofsted	Office for Standards in Education, Children's Services and Skills
P8	Progress 8
PISA	Programme for International Student Assessment
PIRLS	Progress in International Reading Literacy Study
PRUs	Pupil Referral Units
QCA	Qualifications and Curriculum Authority
RICs	Regional Improvement Collaboratives
SATs	Standardised Assessment Tests
SCQF	Scottish Credit and Qualifications Framework
SEND	Special Educational Needs and Disabilities
SIMD	Scottish Index of Multiple Deprivation
SNP	Scottish National Party
SOEID	Scottish Office Education and Industry Department
SQA	Scottish Qualifications Authority
SQH	Scottish Qualification for Headship
TAGs	Teacher-Assessed Grades
TALIS	Teaching and Learning International Survey
TIMSS	Trends in International Mathematics and Science Study
UK	United Kingdom

1 Introduction

Educational assessment serves multiple, often competing purposes within education systems internationally (Newton, 2007; Baird *et al.*, 2017). It informs teaching and learning, certifies individual achievement, communicates what is valued in a curriculum and provides data through which governments can hold schools and teachers to account (Stobart, 2008). How these purposes are balanced carries significant implications for teaching, learning and educational equity (Lingard, Martino and Rezai-Rashti, 2013). Across diverse national contexts, research has consistently identified a tension between the formative potential of assessment and the summative and accountability functions that tend to dominate policy frameworks (Black and Wiliam, 1998a; Harlen, 2007).

Within parts of the United Kingdom, this tension takes a distinctive form. England and Scotland have developed divergent approaches to educational assessment, operating within separate legislative, curriculum and institutional structures (Arnott, 2005; Arnott and Menter, 2007; Hayward *et al.*, 2012). Scotland remains the only devolved nation to maintain an entirely separate education and examination system, reflecting a broader tradition of distinctive educational governance (Walford, 2005; Bryce and Humes, 2018b). Despite these structural differences, the evidence suggests that both jurisdictions have experienced persistent challenges aligning assessment policy with the research evidence on effective assessment practice (Hayward, Macbride and Spencer, 2016). Understanding how and why these challenges persist, and the mechanisms by which they shape classroom practice, remains insufficiently explored. This thesis addresses that gap.

Assessment, curriculum, and pedagogy operate as interconnected elements of classroom practice, forming what Hayward (2012) conceptualised as a learning triangle, with learners and learning positioned at its centre. Which of these elements exerts the greatest influence over teachers' work, and whether this influence aligns with policy intentions, has generated substantial theoretical debate and empirical investigation (O'Neill, 2013; Newton, 2017). In England, accountability operates through standardised performance metrics such as Progress 8 (P8) and the publication of school league tables, as well as policy tools that incentivise inter-school competition and direct pedagogical attention

towards examination outcomes (Wrigley, Lingard and Thomson, 2012; Bhattacharya, 2023). Scotland has, by contrast, favoured a more professionally mediated approach, with accountability embedded within the broad capacities of Curriculum for Excellence and operating with substantially less competition between schools (Hayward, Macbride and Spencer, 2016; Bhattacharya, 2023).

My experience as a teacher in England's secondary education system prompts this research. Over more than a decade in the classroom, I observed how pressures to enhance student performance in high-stakes examinations intensified as students progressed from lower to upper secondary education. Lessons that initially prioritised curiosity and exploration gradually gave way to preparation for examinations. Valuable curriculum time is routinely consumed by mock examinations, grade predictions, and intervention sessions intended to help students cross grade boundaries. Every aspect of the school system seemed focused on the overriding goal of improving grades for all students, yet there is rarely an opportunity to consider whether this emphasis serves genuine educational purposes.

These observations raise questions about how assessment policy operates more broadly, and whether the tensions I have encountered are particular to England or reflect wider patterns in how accountability-intensive systems function.

Both Scotland and England have, over time, recognised pressures within their assessment systems and acted to address them. In Scotland, the Hayward Review (Scottish Government, 2023) recommended a Scottish Diploma of Achievement to broaden recognition of learner accomplishments. The subsequent Education (Scotland) Act (Scottish Government, 2025) established Qualifications Scotland to replace the Scottish Qualifications Authority (SQA), with curriculum reform scheduled through to 2032. In England, the Curriculum and Assessment Review (Department for Education, 2025) has prompted the removal of the English Baccalaureate (EBacc) performance measure and proposed reforms to the P8 accountability mechanism. Both programmes echo longstanding research recommendations to reduce high-stakes examination pressure, recognise diverse achievements, and restore professional agency. Whether these commitments will translate into meaningful change for teachers and students remains an open

question. This thesis considers that question by examining what earlier reforms achieved and what they left unresolved.

In both England and Scotland, judgements of educational quality occur within highly contested political environments. This contestation makes confident judgement difficult, as competing political actors deploy assessment data to assert rival claims about standards, and these claims are amplified through newspaper coverage and combative political rhetoric directed at education professionals (Arnott and Menter, 2007). A body of myths and legends about the comparative quality of schooling accumulates within this dynamic. Several of the more reliable mechanisms for distinguishing fact from fiction have been withdrawn in both jurisdictions. Both have eliminated national assessment surveys, which research has identified as more contextually reliable (Hayward, Macbride and Spencer, 2016; McEnaney, 2021). They have instead moved towards international surveys such as PISA, which scholars have criticised for misuse in cross-jurisdictional comparisons and for evaluating individual policy impact (Sahlberg and Hargreaves, 2015; Jerrim, 2021). This thesis examines how accountability-driven assessment reform is enacted in schools, considering whether divergent policy structures in England and Scotland generate distinct pressures on classroom assessment practice. It explores how assessment processes inform ideas about improving educational quality across the two systems. Drawing on Cuban (1993), it considers the extent to which large, sweeping national policy changes may exert less influence on the “ocean floor” of the classroom than headline metrics might suggest. Assessment, high-stakes examinations, and qualifications appear in both contexts as primary mechanisms for gauging educational standards. Whether the instruments deployed are valid and fit to support such judgements remains a matter of scholarly and policy debate.

Within the UK Arnott and Menter (2007) argue that England and Scotland have developed divergent approaches to managing the relationship between assessment, curriculum, and pedagogy, with varying emphases on accountability mechanisms and professional trust. This divergence has been particularly pronounced following devolution, with the modernisation of the teaching profession taking markedly different forms on either side of the border. Kreijkes

and Johnson (2023) suggest that these post-devolution differences have continued to deepen, producing variation in how institutions and actors exert control over school governance and curriculum across the four nations. Specifically, Arnott and Menter (2007) describe how England has moved towards market-driven accountability rooted in New Public Management, whilst Scotland has maintained a more collegial, trust-based relationship between government and the profession. Exploring how assessment policy is enacted in these contrasting national contexts offers insight into how policy influencers' intentions translate, or fail to translate, into classroom practice. Less well examined, however, is a direct tracing of how these contrasting aspirations have taken shape in the day-to-day work of classroom assessment.

1.1 Research gap and rationale

Assessment policy research identifies more than one mechanism through which the gap between policy intention and classroom practice arises. Hayward *et al.* (2008) attribute this gap to accountability pressures. Marshall and Drummond (2006) describe how formative assessment strategies can become ends in themselves once detached from their underlying purpose, leaving teachers enacting the 'letter' rather than the 'spirit' of reform. Internationally, the OECD (2026, p. 175) observes that when reforming upper secondary external assessments, 'matters of credibility, relevance and fairness tend to dominate policy debates and stakeholder consultation'. Manageability, the workload implications for students, teachers and assessment systems, emerges as the "forgotten principle" and yet as the factor that makes or breaks certification reform. Conditions cited for enactment that honour the spirit of reform include professional development time, local agency, and alignment with teachers' beliefs (Harlen, 2010; Wallace and Priestley, 2011). What remains less clear is how these mechanisms operate together rather than in isolation, particularly across jurisdictions with contrasting accountability structures. Whether the conditions for enactment that scholars describe can be sustained where accountability is organised around comparable performance data remains an open question.

Ultimately, how policy intentions translate into classroom practice in England and Scotland, and the identity of teachers as assessors, working within these

contrasting accountability systems. This evidence offers insight into how these factors operate within different education systems.

Three research questions were posed:

1. Within the context of policy reform (1998-2023) in Scotland and England, what are the intended purposes of assessment?

2. To what extent are assessment policies in Scotland and England aligned with assessment research?

3. How do teachers, school leaders and policy influencers perceive educational assessment policy in the Scottish and English education systems?

These questions are of particular interest given current reform trajectories in both jurisdictions: in Scotland, the Hayward Review (Scottish Government, 2023) has proposed substantial changes to qualifications following extended post-pandemic consultation; In England, the Curriculum and Assessment Review (Department for Education, 2025) signals a possible departure from the accountability mechanisms that have shaped policy since 2010. Together, these reviews offer a rare opportunity to examine how two systems with shared legislative origins, but divergent post-devolution trajectories (see p. 1), approach the same dilemma at the same moment.

The period this thesis examines, 1998 to 2023, is bookended by events that shaped the assessment systems of both nations. In 1998, Black and Wiliam published their review of formative assessment, which invited educators to consider conceptualisations of assessment beyond summative purposes. The following year, the establishment of the Scottish Parliament initiated the post-devolution divergence in administrative structures and policy networks (Arnott, 2005). The intervening period closed with politically disruptive events, most notably the COVID-19 pandemic, which exposed competing public perceptions of examination systems (Richardson, 2022).

This research consists of two complementary strands. The first involves interviews with national-level policy influencers from both nations, constructing an understanding of the political and social contexts for assessment policy intentions and, where considered, plans for their enactment. The second strand involves case studies with teachers and school leaders in one secondary school in Liverpool (England) and one in Dundee (Scotland), illuminating how policies are interpreted and enacted at the classroom level. These sites were not selected as representative of their respective cities, nor as proxies for all schools within England or Scotland; rather, they function as contextual cases through which the enactment of national and local assessment policy can be explored in depth. As Yin (2018) argues, the strength of a multiple-case study design lies not in statistical representativeness but in its analytic potential for cross-case comparison. This allows this thesis to explore whether patterns identified in one context are replicated, contrasted or elaborated in the other.

A further rationale for selecting the two cities was their structural and socioeconomic comparability. Both cities share histories as major industrial, naval and trading ports, each of which subsequently experienced significant economic decline following the contraction of those industries. This parallel journey of deindustrialisation has produced comparable patterns of material deprivation. According to the English Indices of Deprivation 2025, Liverpool ranks 12th of 296 local authority districts in England, placing it within the most deprived 5 per cent nationally (Ministry of Housing, Communities and Local Government, 2025). In Scotland, the Scottish Index of Multiple Deprivation 2020 (SIMD 2020v2) identifies Dundee City as one of the most deprived local authorities in the country, with 38 per cent of children aged 0-15 residing in those most deprived zones (Scottish Government, 2020; Dundee City Council, 2022). The two cities differ considerably in absolute population size. Dundee's population is approximately 148,710 (Scottish Government, 2020), compared with the Liverpool City Region's 1.59 million (Office for National Statistics, 2024). This disparity is, however, broadly proportional to the relative scales of Scotland and England as nations. The socioeconomic profiles of the two sites are therefore comparable. This approach enables comparison of policy trajectories across both jurisdictions, critically examining the extent to which assessment policy achieves its intended outcomes.

1.2 Significance and contribution

This thesis is concerned with a question that sits at the intersection of policy, practice and professional experience: why do accountability-intensive assessment policies persist, even when their consequences for teachers and pupils are well documented? The opening chapters seek to establish the conditions under which this question becomes both necessary and urgent. Current understandings of the relationship between policy formation and classroom practice leave considerable ground unexplored, and it is within that space that this research is situated and contributes.

A central focus of what follows is the relationship between policymakers' awareness of assessment processes and the decisions that shape how those processes are enacted in schools. The thesis considers whether documented negative consequences associated with high-stakes assessment mechanisms inform subsequent policy directions, or whether other factors bear more significantly upon those decisions. In doing so, it invites a reconsideration of how policy enactment is understood, particularly the relationship between research evidence and political decision-making in educational assessment across two distinct national contexts.

Pursuing this question invites closer attention to the motivational assumptions that policy frameworks encode about those tasked with assessing and enacting policy. The thesis engages with Le Grand's (2006) conceptual distinction between 'knights', professionals assumed to act from public-spirited motivation, and 'knaves', actors assumed to be self-interested and requiring external constraint. This distinction provides a lens for examining how England's vertical accountability structures and Scotland's horizontal collaborative mechanisms may encode contrasting expectations of teachers. The study further considers how education practitioners interpret and mediate those expectations as policy initiatives are translated into classroom practice. The practical articulation of this position warrants engagement, and Barber (2015), formerly head of Tony Blair's Prime Minister's Delivery Unit, provides the most influential contemporary statement of it. He argues that visible performance data and clear targets are necessary to translate political intent into improved outcomes for citizens, and that without these mechanisms, public services drift towards inertia. This

perspective offers the contemporary defence of the structures and processes that this thesis examines, and it is engaged with again in the discussion of policy influencer accounts.

Teachers and school leaders, however, are those whose task it is to enact assessment policy in classrooms, and this thesis explores their experiences of working within these contrasting systems. A further area of investigation concerns the temporal effects shaping that enactment. The research explores how teachers' assessment identities, formed through professional socialisation under previous policy regimes, may shape their responses to current reforms. This temporal dimension raises questions about whether policy enactment can be understood without attending to the accumulated effects of earlier reform cycles on professional identity and practice.

This thesis explores how the enactment of assessment policy reform interacts with the prior policy frameworks that shape the work of policy influencers, school leaders, and teachers. It considers how school leaders mediate between national policy and classroom practice, treating the purposes of assessment as sites of negotiation rather than settled premises. As established in the previous section, the comparative design across Scotland and England provides a basis for examining how national policies are enacted within distinct professional and political contexts.

1.3 Researcher positioning

Understanding the significance of this research requires acknowledgement of the standpoint from which it has been conducted, as discussed at the outset of this chapter. My professional and epistemological background has shaped both the research questions and methods used to explore them. As a former secondary school teacher in England and subsequently a teacher educator, I bring a degree of professional familiarity with one of the two jurisdictions under examination that does not extend equally to the Scottish context. Alongside this, my initial academic training in the natural sciences required a deliberate epistemological reorientation when I moved to social science research. This is a process which Elliott (2023) notes is rarely straightforward, given that ontological and

epistemological assumptions formed through prior education and experience are deeply ingrained and do not shift easily.

This dual background produces a positionality that is neither straightforwardly insider nor outsider in relation to the research contexts. Trowler (2011) suggests that such positionality sits at a particular point along a continuum between insider and outsider research. Rather than treating this as a limitation, I consider the productive tension between professional proximity and critical distance to be a generative feature of this comparative design. The methodological implications of this positioning, and the reflexive strategies employed in response, are examined fully in the methodology chapter.

1.4 Research approach

The three research questions that guide this study require an approach that can examine policy formation and professional interpretation simultaneously across two distinct national contexts. Answering these questions demands access to the perspectives of those who construct assessment policy and those who enact it in professional practice. No single method would adequately address this range of experience. A comparative case study design was therefore adopted (Yin, 2018), enabling in-depth, contextually sensitive inquiry within and across the two jurisdictions of England and Scotland.

The first research question concerns the intended purposes of assessment within the period of policy reform examined in this study (1998-2023). To address it, the study required direct engagement with those involved in shaping national policy. Semi-structured interviews were conducted with 26 national policy influencers, equally distributed across England and Scotland, including academics, former senior civil servants, national policy leaders and local policy leaders. Semi-structured interviews were appropriate in this case because they provided sufficient structure to address the research questions while allowing participants to articulate the reasoning and priorities that shaped policy development (Cohen, Manion and Morrison, 2018b). This flexibility was essential for the multiple purposes within policy documents are rarely stated explicitly (Newton, 2007).

The second research question asks about the alignment between assessment policy and research. Addressing it required analysis of the discursive choices made within policy texts and by policy influencers themselves. Critical discourse analysis provided the tools to examine how language constructs particular versions of educational problems and solutions, and to assess whether the evidence base informing policy reflects what the research literature recommends (Fairclough, 2013). The combination of documentary analysis and interview data enabled triangulation between what policy texts claim and what national policy influencers report as their intentions.

The third research question addresses how teachers, school leaders and policy influencers perceive educational assessment policy within their respective systems. To explore this, focused case studies were conducted in Liverpool and Dundee, two city regions selected for their comparable socio-economic profiles and their location within contrasting national policy environments (see page 6). Twenty education practitioners across these schools participated in semi-structured interviews, representing a range of career stages and leadership positions. This design enabled systematic comparison of how national policy trajectories are interpreted and enacted at the institutional level (Ball *et al.*, 2011). By situating practitioner accounts within their policy contexts, the case study approach captured the conditions and constraints that shape professional interpretation and enactment of policy.

The two-tiered design, which combines national-level interviews with policy influencers and school-level case studies with teachers and school leaders, reflects the layered character of the research questions. It enables policy to be traced from its construction at the national level to its enactment in classrooms, a trajectory that single-site or single-cohort studies would struggle to evidence. Reflexive thematic analysis (Braun and Clarke, 2022) was used to identify patterns of meaning across participant groups and jurisdictions.

Several limitations follow from the design choices that shape this study. Focusing on two of the four nations of the United Kingdom enables sustained engagement with the distinct policy histories of England and Scotland and supports richer case-study analysis. It does, however, exclude Wales and Northern Ireland, whose educational systems might have offered further comparative diversity.

Selecting a single secondary school within each nation similarly precludes examination of inter-school and phase variation, local cultures and school funding arrangements. Methodological choices impose further constraints on what the study can claim. Relying solely on semi-structured interviews, without accompanying classroom observation, means the study draws on teachers' self-reported accounts of practice, which may not fully reflect their enacted approach. The original design included observation, but consent for school visits was not granted owing to COVID-19 restrictions.

Restrictions imposed during the COVID-19 pandemic presented significant challenges for both school recruitment and ethical access. Approaching potential participants required careful negotiation, with particular attention to minimising disruption to school life. This was especially complex when recruiting schools directly in England and when working through local authorities in Scotland. As a result, the research design required revision, replacing direct classroom observation with online interviews in which teachers and school leaders discussed their assessment practice. In recognition of the demands placed on participating schools, the research drew on principles of reciprocity, considering how its findings might serve teachers' and school leaders' professional needs (Taylor, 2023). Although this change restricted direct access to assessment enactment in the classroom, it afforded scope to examine how teachers and school leaders understood and articulated their approach to assessment. That both jurisdictions were experiencing ongoing revision to national assessment policy at the time of data collection adds a further layer of context to the scope of the claims this study is positioned to make.

Much existing research into assessment policy explores either the intentions of national policy or the experiences of those who enact policy in schools. Less attention has been paid to the relationship between these two levels of account. This thesis addresses that gap by combining interviews with national policy influencers in England and Scotland with case study evidence from teachers and school leaders in two secondary schools. By cross-referencing accounts from these two levels within contrasting policy environments, this work explores whether the gap between policy intention and classroom practice might be better understood as a structural feature of assessment reform rather than an

isolated local phenomenon. In doing so, it seeks to provide a foundation for analytical generalisation to jurisdictions where similar tensions between policy intent and professional practice may be present.

1.5 Thesis structure

The thesis is organised into six chapters. Chapter One, this introduction, has already established the research context and rationale. It identified the central problem the thesis addresses, sets out the three research questions, and outlines the researcher's position in relation to the two countries explored in this work.

Chapter Two maps the conceptual literature on which this thesis builds. It explores research literature across four interconnected domains: the purposes of assessment processes; teacher practice and identity; accountability and educational change; and the policy chronologies of England and Scotland from 1998 to 2023. Together, these domains provide the conceptual foundation for the empirical analysis.

The review draws on Looney et al.'s (2018) model of assessment identity as a sensitising concept to explore how teachers' beliefs, knowledge and professional formation shape their engagement with assessment policy. This groundwork examines the policy environments within which teachers operate and introduces analytical resources for understanding how professionals at different levels of the system mediate policy. Ball, Maguire and Braun's (2012) enactment framework and Ball et al.'s (2011) typology of policy actors provide vocabularies for examining school-level mediation, whilst the review also considers how accountability structures position teachers differentially across the two jurisdictions.

Chapter Three establishes the methodological framework for the study. It justifies a qualitative, interpretivist approach before presenting the rationale for a comparative case study design across two jurisdictions. The chapter addresses the researcher's positionality as a former English secondary school teacher and identifies the reflexive strategies used to manage insider and outsider perspectives throughout the study. It concludes by framing the design's inherent

limitations as deliberate methodological choices that prioritise contextual depth over breadth.

Chapter Four presents the empirical findings drawn from thematic analysis of qualitative interviews with national policy influencers and education practitioners across England and Scotland. It reveals relationships between policy construction at the national level and its enactment in the classroom. The chapter addresses all three research questions, examining the intended purposes of assessment within the period of policy reform under study, exploring the alignment between policy and the research evidence base, and illuminating practitioners' experiences of working within divergent policy environments.

Chapter Five develops the analytical discussion by exploring why accountability-intensive assessment reforms persist despite documented detrimental consequences. It demonstrates that national policy influencers proceeded with reform enactment despite explicit awareness of its negative impacts. The chapter analyses the selective appropriation of international comparative data to legitimise ideologically driven reforms and examines the temporal constraints that teachers' accumulated professional identities place upon reform enactment. It argues that the policy-practice gap reflects complex political, temporal and cultural dynamics rather than insufficient evidence or awareness.

Chapter six draws together the empirical findings and analytical threads developed throughout the thesis. It provides direct responses to each of the three research questions and articulates the study's original contributions to understanding educational assessment policy across two national contexts. The chapter concludes by presenting the study's limitations and offering recommendations for applying its findings.

1.6 Chapter summary

This chapter has established the context, rationale and structure for a comparative examination of assessment policy enactment across England and Scotland between 1998 and 2023. It identified a persistent tension within both jurisdictions between the formative potential of assessment and the accountability-intensive functions that have come to dominate policy

frameworks. This tension has not diminished despite two decades of reform activity in both nations.

The chapter introduced three research questions concerned with the intended purposes of assessment policy, the alignment between that policy and assessment research, and the perceptions of teachers, school leaders and policy influencers working within contrasting accountability systems. These questions were situated within the context of current reform trajectories in both jurisdictions, where recent developments in Scotland and England signal potential shifts in the dominant policy logic without yet resolving the underlying tensions this thesis examines.

A two-tiered comparative case study design was justified as the most appropriate methodological approach. This design combines semi-structured interviews with 26 national policy influencers across England and Scotland. It runs in parallel with case studies of education practitioners in secondary schools in Liverpool and Dundee. The rationale for selecting these sites was discussed, with reference to their comparable socioeconomic profiles and their locations within contrasting national policy environments.

Addressing the researcher's positionality, this chapter also considered the stance shaped by professional experience as a secondary school teacher and teacher educator in England. That stance is neither wholly insider nor outsider in relation to the two research contexts. The methodological implications are further explored in the methodology chapter.

The chapter concluded with an overview of the thesis structure, outlining the contribution of each subsequent chapter to the overall argument. The Literature Review that follows examines three interconnected themes: the purposes of assessment and the rise of accountability-driven reform, teachers' assessment identities and the influence of prior policy regimes, and the accountability structures that have developed in England and Scotland since devolution. Together, these themes establish the scholarly context within which the empirical inquiry is situated.

2 Literature review

2.1 Chapter introduction

This chapter reviews the research literature in four sections that together frame the territory from which this thesis builds. It takes 1998 as its starting point, the year of Black and Wiliam's review of classroom assessment and the year before the establishment of the Scottish Parliament.

Section 1 explores the multiple purposes of assessment processes, distinguishing between formative and summative purposes. It also considers the potential tensions that may occur when assessment processes are used for multiple purposes. Finally, the extent to which assessment policymaking engages with existing research evidence is considered.

Section 2 turns from the purposes of assessment processes to the professionals who enact them. It reviews research on assessment literacy, teacher beliefs and assessment identity, and draws on Ball, Maguire and Braun, (2012) policy enactment framework to examine how teachers mediate assessment policy. Finally, the section also examines how the literature has approached the temporal dimension of teacher formation under successive policy regimes.

Section 3 explores the role of accountability and educational change in England and Scotland. It considers how competition, international benchmarking, institutional isomorphism, inspection and motivational assumptions about public-sector professionals shape the design of accountability mechanisms in each jurisdiction.

Section 4 outlines the context of changes to assessment policy in England and Scotland from 1998 to 2023. It traces the chronological development of assessment policy in each country, drawing on official documentation and contemporary media reporting to examine how successive governments have shaped the assessment policy systems that teachers work within.

Together, the four sections move outwards from the purposes of assessment to the professional mediation of policy, the accountability processes of educational

change, and the policy chronologies of England and Scotland. This progression frames the territory in which the empirical analysis later situates itself.

2.2 The working definition of assessment in this thesis

Assessment plays a key role in education, providing insight into learning within educational institutions. The term assessment derives from the Latin assidere, meaning to sit beside (Oxford Learner's Dictionaries, 2020). However, the term assessment, as a process for eliciting learning, is often used interchangeably with the tools by which it is achieved, such as exams, and is therefore used without precision. Further, the processes that underlie what is meant by those using the term assessment are often poorly understood and, therefore, oversimplified (Richardson, 2022). Stobart (2008a) highlights the dilemma at the heart of the problem, arguing that the term 'assessment' is often used interchangeably with 'test' and 'examination', especially in the United Kingdom. Stobart (2008a, p. 8) defines assessment as an activity which is 'encompassing a range of approaches to evidence gathering'. This posits the role of assessment as a process of inferring an individual's mastery of a set of skills or knowledge. However, it doesn't attempt to address what these inferences might be used for, which is an important aspect of defining an assessment process. Baird et al. (2017, p. 317) further build on this: 'Assessments are used to investigate what people know and can do and to make decisions regarding whether they have learned what was expected'.

While both definitions above present assessment as a means to appraise performance, they exemplify what some authors assert is a lack of discussion about what action should follow assessment (Pellegrino, Chudowsky, and Glaser, 2001). Popham (1999) advances the idea that we might put students in certain situations and observe their responses he managerial role of assessments, but it is Harlen (2013) who first discusses the prospect of using this judgement. Baird et al. (2017) then advance this definition by suggesting that the assessment can be reviewed to determine whether students have learned what the teacher intended them to learn. It is possible, however, for assessment to have a diversity of purposes and some of these definitions suggest that a judgement is necessary in all assessment processes, but this may not necessarily be cumulative to produce a grade or a level. When supporting learning, it might be

the process or steps along the way to an eventual judgement that a teacher uses to infer about learning; there is often nothing enlightening about the grade or level in itself (Newton, 2007).

To conclude, within this work, assessment will be defined as a variety of activities that gather evidence of student learning, which is then used to make a judgement or improve learning. Evidence gathering can serve as a basis for further action; however, that evidence can be used for a variety of purposes. The purposes of assessment are tied to how we define the different forms of assessment mentioned in the literature, which will be explored in the next section. Having established this working definition, this section now examines how assessment processes serve different purposes, and the tensions that can arise when a single assessment is expected to serve more than one.

2.2.1 Formative assessment as a developing concept

Black (1997, p. 155) identified three assessment purposes, of which the first, assessment to inform learning, captured the attention of researchers and policymakers. A year later, Black and Wiliam (1998a, p. 140) developed this purpose further, arguing that assessment should improve learning as often and as fully as possible. They held that teachers 'need to know about their pupils' progress and difficulties with learning so that they can adapt their own work to meet pupils' needs'. An accompanying pamphlet (Black and Wiliam, 1998b, p. 141) reported effect sizes of between 0.4 and 0.7, claimed to be 'larger than most of those found for educational interventions'. The same pamphlet warned, however, of a 'poverty of practice' in the deployment of formative assessment, organised into three concerns covering: effective learning, negative impact, and the managerial role of assessments.

The first challenge to the effective use of formative assessment raised by Black and Wiliam (1998a) is that its potential impact might be impeded by a lack of clarity about what counts as valuable learning. If the classroom focus is primarily on assessing rote and superficial learning, there will be few opportunities for students to demonstrate a deeper understanding. In some assessments, there can be a disparity when focusing on the agreed criteria for success, where shallow judgements, such as the presentation of work, are evaluated instead.

The second issue is that giving students judgemental feedback, such as grades, can be demotivating. Kluger and DeNisi (1996) found that such feedback can distract pupils from the task at hand, redirecting their attention towards their self-image rather than towards improving their work. In further support of this, Black and Wiliam (1998a) argue that students who receive feedback that they are underperforming may give up or attempt to continue in ignorance of what could be improved. The final issue concerns the ways in which assessment evidence is used without adequate consideration of a student's learning needs. The authors suggest that teachers commonly do not fully consider the student's role in reflecting on the learning process. For example, students are not involved in setting the learning intentions for lessons or in deciding how learning will progress after the assessment. The latter produced booklet (Black and Wiliam, 1998b) appealed to policymakers wanting to find methods to improve education as it provided evidence of the impact of formative assessment practice, for example, 'These effect sizes are larger than most of those found for educational interventions' (p. 141). Secondly, phrases claiming an improvement in equality of achievement, such as 'improved formative assessment helps the low achievers more than other students and so reduces the range of achievement while raising achievement overall' (p. 141), give justification for academics and policymakers to propose introducing these particular practices nationally.

In proposing improvements to teaching and learning practice, Black and Wiliam were hesitant to offer prescribed methods or practices that should be deployed in policy or practice. Bennett (2011, p. 19) in his critique, stated that 'The term, 'formative assessment', does not yet represent a well-defined set of artefacts of practices. A meaningful definition requires a theory of action and one or more concrete instantiations, which in his view, is not provided by the work of Black and Wiliam reviewed above. This definition may also be needed to adequately differentiate between formative and summative assessment. Black and Wiliam (2018, p. 553) responded to this critique and demand for clarification by reasserting that formative assessment practice is at heart, 'a distinction in the kinds of inferences being drawn from assessment outcomes' rather than a defined set of tools or practices. Teacher learning would involve live examples of the proposed concepts, but not direct prescriptions for the tools to be used.

Whilst much of the literature discussed so far describes the purposes of assessment in England, similar projects were undertaken worldwide. Also of interest here is the journey undertaken within Scotland. Simpson and Hayward (1998) describe the developing role of formative assessment in Scotland, emerging as a fundamental by-product of the innovation in national assessment. Through periods of school-based initiatives used to disseminate research findings, the role of assessment as an 'integral component of learning and teaching was permeated throughout the educational community' (Simpson and Hayward, 1998, p. 445). Whilst the authors identified areas of change, they highlighted the need for further work to ensure that the changes were not superficial in school practice and demonstrated a 'redistribution of knowledge and power between teachers and pupils' (Simpson and Hayward, 1998, p. 446).

Discussions within the literature attempting to distinguish between the formative and summative purposes of assessment have continued from Black and Wiliam's paper in 1998 to the present day. Harlen (2005b) highlighted some areas of practice where there appeared to be tensions between the use of assessment for formative and summative purposes. They concluded that the formative purposes of assessment must clearly inform learning and teaching, whereas summative purposes may serve several different judgemental purposes. This debate, however, did not provide a lens into the complexity of how assessment could be used and how these purposes may interact.

Newton (2007) sought to move beyond this binary framing by disaggregating assessment purpose into three distinct levels: judgement level, the reason behind the design of the assessment; decision level, the desired use of assessment results; and impact level, the impact that may occur as a result of the assessment taking place (p. 150). Newton proposed a revised list of eighteen uses for assessment (p. 163-165) and cautioned against conflating purposes intentionally or unintentionally (p. 154). Finally, the paper warned that the purposes must be carefully considered before an assessment takes place and that grand overarching categories of purpose cannot be created, as this may cause assessment policy to utilise assessments for multiple purposes.

The assessment processes at the centre of this thesis illustrate precisely this risk of purpose conflation. The GCSE in England and the Scottish National 5 were

designed to certify individual learning at the point of leaving school. A further purpose of these qualifications for certification has been to serve as a driver of school improvement, holding schools accountable for the number of students who achieve top exam scores. They also show the various ways in which assessments might be used and, as Stobart (2008a) illustrated, abused. These assessments can have many effects, such as shaping what is taught in classrooms and how it is taught, often called washback (Messick, 1996). The results of these tests and the accountability for teacher pay and school league tables have been attributed to a narrowing of the curriculum, restricting what is taught to what is tested (Amrein and Berliner, 2003; Spielman, 2017). The inappropriate use of learning time in schools to maximise scores on summative tests has been called “gaming” by politicians and school inspectors, but is also sometimes seen as a necessary way for teachers to bring better futures for their students (Popham, 2001; Johnson and Crisp, 2009).

Some teachers may change aspects of their teaching and even the formative assessment used in the classroom in order to 'prepare' students for terminal assessments. Boyle and Bragg (2006) found that between 1997 and 2004, Key Stage 2 (ages 7-11) timetabled teaching time rose for English (+3.7%) and Maths (+2.6%), whilst falling for Science (-1.6%) and History (+1.6 %), changes attributed to central policy requirements. Stobart (2008a) argues that assessment is commonly used for accountability, where students' assessments are also used to quantify the quality of teaching within institutions. Assessments such as these are also used to measure the quality of schools and teachers (Baird et al., 2017). Accountability resulting from educational assessment can even reach policymakers and governments, as in the case of the Programme for International Student Assessment (PISA) (Stobart and Eggen, 2012). It has been argued that the use of assessment for accountability purposes is a recent phenomenon and that it is a tool neoliberal policymakers use to drive competition between educational institutions and professionals (Santori, 2020). However, Stobart (2008a) demonstrates that this has, in fact, been taking place since 1862. Not only has assessment affected what is taught, but it has clearly had a long-term impact on how teaching and learning are viewed. Many argue for greater alignment between assessment and learning so that the assessment process becomes more relevant to the learning it aims to measure (Baird et al.,

2017). This view has been built progressively from emerging works that question the definition and academic understanding of assessment processes that aim to serve a formative process.

As discussed above (p. 18), the absence of an agreed definition of formative assessment makes it difficult to measure its effectiveness or to distinguish formative practice from other assessment activity. This definitional ambiguity has practical consequences for each professional group involved. For policymakers, academic clarity about what formative assessment involves is a prerequisite for any policy intended to increase its incidence in the classroom. For school leaders, a lack of an agreed definition may prevent meaningful evaluation of how effectively formative practice has been enacted within their school. For classroom practitioners, assessment activities risk being adopted without a clear understanding of their formative intent (Marshall and Drummond, 2006).

Two facets of formative assessment in practice merit attention. Wiliam and Thompson (2007) propose that an assessment can serve either formative or summative purposes, with the use made of the information determining which. They also set out a framework of strategies for conceptualising formative assessment in the classroom. The first claim is not without difficulty, as earlier discussion noted, because formative purpose cannot be attributed to a fixed set of activities. The assessment activity itself is not formative, although the subsequent use of the information may be. Their framework directs teachers to consider where pupils aim to reach, where they currently stand in their learning, and what they need to do to progress. This thinking has informed a range of tools and strategies promoted through policy processes in England and Scotland, and now features within teacher training.

This definition can be problematic, as previously discussed, as it is difficult to attribute it to a set of activities that are either formative in purpose or a completely different type of use. The assessment process or activity itself is not formative, but the use of the information may be. Wiliam and Thompson (2007) developed a useful framework of strategies that practitioners might use to conceptualise how formative assessment can be used in their classrooms. This framework allows teachers to focus first on where their students aim to get to,

where they currently are in their learning, and what students need to do to improve further. This framework has led to a multitude of tools and strategies being promoted via policy processes to classrooms across England and Scotland and incorporated within teacher training.

In response to the desire for concrete examples that teachers can use in the classroom, Wiliam (2011) published Embedded Formative Assessment, a guide for teachers, complete with pages of practical techniques. Wiliam's book helped clarify and link the intentions behind using assessment for formative purposes and how this might be enacted in the classroom. Examples of formative techniques include using traffic lights to allow students to express their confidence in the learning, and focusing on the lesson's learning intentions before commencing an assessed task. The formative assessment techniques advocated by Wiliam were enthusiastically promoted by educational policy and influential middle-tier stakeholders (Black and Wiliam, 2003; Hayward, Priestley and Young, 2004). However, its adoption and integration into everyday teaching and learning activities were not as pervasive or uniform across the two school systems (Hayward, Priestley and Young, 2004; Kirton et al., 2007). In addition, some workers have raised concerns that they do not reflect the procedural rigour or reliability of assessments (Bennett, 2011). Some assessment techniques, such as traffic light self-assessment, (Wiliam, 2011) might involve them inflating their self-perceived performance, mistaking excellence in learning for simple presentation or not understanding the domain-specific knowledge that is being assessed (Coffey et al., 2011). Work on formative assessment needed to be aligned with the learning that was intended and perhaps even the subject-specific learning that a teacher to take place.

To achieve alignment between assessment and learning, Bennett (2011, p. 16) suggested that formative assessment might be 'more profitably conceptualised and instantiated within specific domains'. This need for formative assessment to make more sense for different domains and theories of learning has continued to be a theme for research (Baird *et al.*, 2017).

Black (1999) linked the ways in which assessment can be deployed that correspond to learning theory. He started this discussion by pointing out that many assume knowledge can be taught through chunking into smaller pieces of

information that are repeated in isolation. From this conclusion, Black warns that it is tempting to assume that with repetition, information can be recalled and considered as learned. With this example, behaviourist ideas can be linked to the rehearsal of correct responses. James (2008) argues that the assessment method and the learning theory need to be linked for valid conclusions about the learning to be drawn. However, this is an aspect of assessment that has not always been well considered, including in formative assessment.

Formative assessment, as framed in policy guidance and by middle-tier advocates, has tended to be reduced to a set of strategies and techniques. Coffey et al. (2011) suggest that this narrow framing leaves teachers without the domain-specific knowledge needed to interpret what students reveal about their thinking, and without a broader conversation about how classroom assessment connects to the intended learning and the subject at hand. Bennett (2011) adds that teachers often lack the assessment expertise required to draw valid inferences from classroom evidence. Without these capacities, formative assessment may appear in the classroom as a procedural gesture rather than a pedagogical practice. For formative assessment to serve learning, it needs grounding in learning theory and in the substantial body of evidence already gathered about assessment practice (Baird *et al.*, 2017).

This section has explored how formative assessment is critiqued for lacking a unifying theory of what it means to assess formatively and for being detached from the learning it intends to evaluate. Using assessment processes for formative purposes cannot be affected by assessments used for other purposes, hoping that it will improve scores just through practise. In this way, school-leaving examinations might be used as sample assessments, with teachers allowing students to practise their performance and, at the same time, to glean information about student learning. In this way, formative assessment in some schools has been affected by the pressure faced by national external examinations. These assessments can tempt teachers to use tasks that may not align with the learning taking place, but hope to improve student performance by providing an opportunity to practise, under pressure to perform well in the examination.

In England and Scotland, high-stakes assessment has been documented as one challenge to the effective enactment of formative assessment practice (Wiliam et al., 2004). Formative assessment can be seen as a method to practise for summative assessments and as a way of collecting evidence for formative purposes (Stobart, 2008). The focus on grades or scores that the assessments produce can detract from the learning and lead to more superficial learning in the short term, and reduce learning overall (James, 2008). Teachers may want to improve their students' learning but feel forced to abandon a true focus on the substance of learning in favour of developing examination techniques for high-stakes assessment.

A focus on learning and unifying the method of assessment and the intended learning is an important consideration for valid assessment (Harlen and James, 1997; Hayward et al., 2008; Baird et al., 2017). The professional development opportunities necessary for teachers to construct assessment opportunities and monitor the information produced by these assessments are often not available (Black and Wiliam, 2018). With sufficient time to allow teachers to reflect and improve on their formative practice, there is an increase in learning, but too often, not enough time is allocated to allow this to occur (Leahy and Wiliam, 2012).

Limited professional time also constrains the development of robust summative assessment. The validity and reliability of assessment tools, along with how information from them is used, are central to whether assessment can serve summative purposes. Marking criteria, standards, and inconsistencies in teacher conduct can all lead to variation in overall judgement, making moderation essential. Hayward et al. (2012) argue that bringing teachers together to discuss required standards produces greater consistency in allocating judgements. They present moderation, sometimes termed standardisation, as an opportunity for participants to scrutinise and validate assessment standards using student examples. This process helps prevent misinterpretation of criteria and supports accurate application of mark schemes. They also position moderation as a professional development practice through which teachers internalise expected performance levels and carry those standards into both instructional practice and formative assessment.

One of the largest barriers to educators meeting to discuss standards in assessment is the time and costs involved in the process (Bloxham, Hughes and Adie, 2016). Methods for limiting this cost include submitting samples to moderators, having moderators visit schools, and administering a common written test to estimate the mean and range for an assessed group. In Scotland, the introduction of the 35-hour contractual statement on workload already causes tension between what must be done and what ought to be done by teachers as committed professionals (Mcphee and Patrick, 2009). The implications are that if further moderation or standardisation were added to the workload, something would need to be removed. The desire and need for teachers to improve assessment practices are not always matched by the time and funding necessary to do so, and recent fiscal pressures in both jurisdictions suggest this gap may be widening.

Professional development in assessment is shaped by the financial and temporal constraints of contemporary schooling in England and Scotland. Sustained reductions in school budgets have decreased the availability of teacher development (Mason, 2022), and school leaders may hesitate to allocate scarce funding to assessment training that does not directly influence performance indicators. A Teacher Development Trust report found that 21,000 teachers had received near-zero CPD of any kind, and that only 0.7 per cent of school budgets across the sector were allocated to teacher CPD (Weston, 2018). Time presents a further constraint. The process of transforming practice has been estimated to take at least 2 terms, more usually a year (or longer) (Teacher Development Trust, 2014). Within this constrained environment, Harlen (2010) sets out several criteria for effective professional development in assessment. Teachers require time to meet and discuss new ideas, drawing on input that helps them reflect on the underlying purposes of assessment practice. This time enables teachers to create and adapt strategies that respond to changing class needs. Teachers also require the professional freedom to experiment with classroom innovations, supported by encouragement and shared goals with school leaders.

There is a clear risk that assessment development will be ill-planned, with top-down initiatives confining teacher practice. This risk can be removed, however, if teachers are given agency and collective accountability through reflection, in

which the impact of developmental input and trials can be reviewed in ways that promote teacher professionalism (Gardner et al., 2008). Wallace and Priestley (2011) praise the improvements made by teachers through the Assessment is for Learning programme of reforms in Scotland. They highlight the Highland Council's 'empowerment' (p. 377) of classroom teachers, allowing them to tailor the policy reform to the needs of their own students and classrooms. For policy enactment, however, there remain challenges to ensure successful reform.

Policy reform involving assessment can be extremely wide-ranging, such as the Assessment is For Learning initiative in Scotland, which considered all assessment activities, including formative, summative and any activities that monitor student learning (Spencer and Hayward, 2016). The degree to which this policy was successful has been called into question many times, with some suggestions that much had been left unchanged in classroom practice (Hayward, Priestley and Young, 2004; Gardner et al., 2008).

Scotland's policy trajectory illustrates how strong uptake in particular areas does not guarantee coherent system-wide integration. Hayward and Hedge (2005) drew attention to a persistent lack of awareness or concern for the ways in which different parts of an assessment system shape one another. In August 2010, Scotland launched Curriculum for Excellence, within which the Assessment is for Learning programme was subsumed. The enactment of these policies, and the outcomes that followed, is examined in subsequent sections. Hayward (2015) suggests that embedding Assessment is for Learning into practice encountered substantial obstacles, attributable to declining teacher salaries, rising workload demands, and other systemic pressures within the education system.

In England, a parallel trajectory unfolded. The Assessing Pupils' Progress initiative and the National Curriculum levels, intended to support valid teacher assessment, were progressively co-opted for accountability purposes until levels were abandoned in 2015 (see further discussion on page 53). The challenges to enactment in both countries share a common pattern. Policies designed to enhance teacher assessment were repurposed as accountability mechanisms, resulting in increased workload, superficial compliance, and calls for further reform.

This section has explored many challenges faced by teachers and policymakers in enacting assessment policy. There is a trend in policy enactment where the proposed reforms can claim to offer freedom and strip back perceived constraints in bureaucracy, but are blamed for holding back student learning. This is combined with accountability measures rather than support in the form of professional development and time. Professionals then often develop superficial practices that have not had time to be informed by research evidence and reflect on their impact on student learning. When much of the work focuses on meeting these accountability criteria, meeting the policy can become hard work, and the professional workload becomes unmanageable. Once this occurs, calls for the next policy reform are made.

2.2.2 High-stakes assessment and its consequences

When used for the purpose of selection, assessment can determine students' admittance to courses and professions, and even their future status as successful learners. The degree of success students achieve in achieving positive examination results can have several long-term consequences for their futures, creating a disparity in their life chances (Baird *et al.*, 2017). In the discussion of high-stakes assessment, these consequences define the stakes, where successful performance brings valuable rewards denied to those who may be said to have failed. As described in previous sections, these high-stakes consequences can create negative implications for teaching and curriculum.

In his definition of high-stakes assessment, Madaus (1988) encompasses any assessment that has the consequence of separating students, teachers, or schools into successful or unsuccessful groups, with implications for both, which can lead to progression in education, career, or reputation. Berlak (2000, p. 189) argued that in the UK and the US, high-stakes assessments were becoming increasingly common, and the test scores could play a role in deciding ‘where and how individuals are educated and [whether they] ultimately gain or fail to gain access to particular schools, jobs and professions.’ In these environments with such consequences, it is clear why the results of such assessments can be considered highly consequential.

Stobart and Eggen (2012, p. 1) define an assessment as high-stakes when its results carry serious consequences for any one of its stakeholders, and identify three principal functions: 'selection and placement, raising standards and accountability'. This framing broadens the field of impact beyond the student, since teachers, school leaders and policymakers may also be affected by the outcomes of such assessments. Examining the discourse around these functions and the consequences they generate helps to surface the unintended effects that may follow from accountability-driven testing.

In high-stakes assessment, validity, fairness and reliability are important factors used to evaluate their effectiveness. These three terms represent fundamental principles that are considered in the design of assessments, especially where they are said to be high-stakes judgements of performance (Dunbar, Koretz and Hoover, 1991). For a given assessment item, these principles cannot be considered in isolation; a test that is truly valid will be so for many learners undertaking it. In this section, each of the above principles is considered, exploring why they are necessary factors and what the possible implications may be if they are not well thought out.

Newton (2012, p. 25) defines validity as '… the difference between having, or not having, a solid technical justification for using a given assessment procedure to make a certain kind of decision…'. As discussed previously, high-stakes assessments can determine whether a student can advance to a particular course or career. A valid assessment must also determine accurately when a candidate lacks the qualities required for entry. Without this, the refusal of entry cannot be justified based on the assessment itself. If these conditions are met, the assessment process can be said to possess validity for its intended purpose.

Related to validity is the need for high-stakes assessments to enable fair choices about students and their futures. Gardner and Cowan (2007) raise concerns about the Transfer Procedure Test, also known as the 11-plus, which determines progression within a selective schooling system. Students whose results are deemed unsatisfactory can be prevented from entering their school of choice. On reviewing performance, the authors (p. 159) found that pupils scoring as high as 70% could be awarded a grade 'D' and deemed to have failed. The example illustrates how successful learners may be labelled as unsuccessful. Judgements

with such consequences require consistent application that reflects each student's relative attainment, which remains a persistent concern for education policy and those who enact assessment (Black and Wiliam, 2012).

In scientific experiments, measurements taken are said to be reliable if they reveal the same result each time. In assessing the reliability of assessments, the accuracy and the precision of the tasks to measure student responses (Baird, Greatorex and Bell, 2004). An assessment tool is a representation of learning and can only serve as a proxy for a student's full capabilities. An ideal assessment process would yield similar indications of a student's abilities even when multiple assessment tools are used to make a judgement.

Black and Wiliam (2012) discuss the ideal of consistency between multiple assessments as reliability, breaking the concept down further into three sources of what they term 'construct-irrelevance variance' (p. 4). They identify various aspects that factors that may contribute to such variance, including disparities in marking across raters, differences in students' performance on certain days, and the influence of the type of question on the student's comprehension and knowledge being sampled. These three contributors to variance in student outcomes, combined with the previously discussed negative impacts of high-stakes assessment, raise many questions about how we best distinguish between students' admittance into academic tracks based on their performance in an assessment process. These issues relate to concepts of fairness in assessment, whereby students can expect to be reliably judged against the same standards.

Fairness in assessment can vary, not just in whether a test item is fair, but also in the process and outcomes of the assessment practice. Isaacs *et al.* (2013, p. 57) propose that for an assessment to be considered fair, 'the content, its context and the expectation of the assessment's performance should provide all candidates with an equal opportunity to demonstrate their ability' (p. 57). This is especially important in high-stakes assessments, where the extent to which a student demonstrates their ability determines the consequences of the assessment results. Educational institutions rely on fair assessments to evaluate prospective students. These institutions need accurate and equitable measures of student performance to make informed admissions and placement decisions. Fairness in assessment ensures that all stakeholders can trust the results as valid

indicators of student capabilities. Further to fairness and to help students perform their best in assessments, this can cause a number of effects in the classroom, both intended and unintended.

High-stakes assessments can operate as a method of governmental control. In England, central control over curriculum and pedagogy is maintained through the requirements of high-stakes assessment (Wrigley, Lingard and Thomson, 2012). Assessment and accountability systems are intertwined in modern education, with high-stakes examinations directing teachers' attention towards examined content. Such assessments may also inform school inspection outcomes, school choice and teacher pay. Teachers in Scotland face comparable pressures, given the need to prepare students for examined qualifications. Hayward et al. (2008) examine how the pressures of high-stakes testing during exam season can deprioritise effective pedagogical techniques such as formative assessment. They observe that teachers may describe a 'risk' of diversifying their approaches in the face of high-stakes assessment (p. 1). This sense of risk, and the ways teachers seek to mitigate it for themselves and their students, can alter classroom teaching.

Alderson and Wall (1993) critically examine how summative assessments influence teaching practice, specifically leading teachers to focus instruction on content and skills frequently featured in the assessments. They term this feedback from assessment to pedagogy ‘washback’ or ‘backwash’ (p. 115). As previously mentioned, there remain calls for greater alignment between assessment and learning (Baird *et al.*, 2017). This is particularly important, as in the absence of this alignment, some studies have shown that high-stakes assessments can narrow the curriculum being taught (Au, 2007). This can leave the curriculum and student progress to be defined by summative assessments, as students and teachers work to frame learning under their requirements (Daly *et al.*, 2012).

Refining the curriculum and pedagogy with reference to summative assessment can have both positive and negative consequences. Koretz, McCaffrey and Hamilton (2001) highlighted several positive influences on teachers from high-stakes assessment: increased motivation, increased prioritisation, and alignment with standards. All three of these influences are important considerations that

require further discussion, especially for interpreting national policy and school practice.

Using high-stakes assessment processes as a motivational tool in education have achieved varying degrees of success. Stobart (2008) would agree with many national policies that aim to use the standards set by summative assessment, especially where they provide a means to raise expectations and set standards for attainment. However, he does refine the effective responses to 'adding more days to the school year; instituting remedial classes outside normal school hours; or devoting more of the school year to actual teaching' (p. 122). These changes will provide more teaching time, which might be more conducive to student learning or provide time for increased bureaucratic activities. The allocation of extra resources and time may not be equal, and there is evidence that they may be directed towards schools or groups of students that best improve the relative performance of those who allocate them (West, 2010). In opposition to this, however, it should be questioned why these initiatives cannot be instigated without the need for high-stakes examinations and the negative effects they cause.

The rewards and sanctions associated with high-stakes testing can lead students to become less intrinsically motivated and have a detrimental effect on their motivation to learn (Amrein and Berliner, 2003). In addition, the need for learning to take place with the aim of passing an assessment can result in much shallower learning, aimed at achieving a certain pass mark in the assessment (Ames and Archer, 1988). Harlen (2006) adds that the focus is often on performance in the assessment rather than on real intrinsic motivation for learning. With minimum standards defined by assessment tools, students may limit their practise to meeting these basic requirements, and teachers may focus on meeting these expectations.

The focus of a specific examination's performance criteria can lead to selective prioritisation of curriculum aspects, thereby privileging individual skills and knowledge. In England, the National Numeracy Strategy and National Literacy Strategy, alongside the end of Primary Standardised Assessment Tests (SATs) have not only affected what is taught but also how teachers view particular imposed teaching approaches (Webb and Vulliamy, 2007). The literacy and

numeracy hours still have a lasting legacy, which has been strengthened since the cessation of the Science assessment (Stobart, 2008). This phenomenon is not unique to England. In Scotland, the Scottish Office Education and Industry Department (SOEID) collect assessment data on English Language and mathematics, which increased the stakes for testing in these subjects (Hutchinson and Hayward, 2005). The result of this response to high-stakes assessment driven by policy reform is ‘a primary curriculum dominated by teaching time allocated to English and mathematics’ (Boyle and Bragg, 2005, p. 435). Some arguments in support of this prioritisation of the curriculum might include the claim that it enables all students to master key skills such as mathematics and English, which are important for accessing other aspects of the curriculum. However, as Boyle and Bragg (2005) describe, this can lead to the marginalisation of other subjects that are not assessed and deny students the opportunity to access other parts of the curriculum. High-stakes assessments can incentivise this curricular constriction, with teachers trying to improve outcomes at any cost due to potential accountability consequences (Polesel, Rice and Dulfer, 2014).

These consequences of high-stakes assessment reflect a pattern that Stobart (2008a) captures through his application of Goodhart’s Law to education. Originally derived from economics, this principle holds that when a narrow indicator is selected as a target, it distorts the activity it was intended to measure. In education, this suggests that once examination scores become the primary measure by which schools are judged, the behaviours adopted to improve those scores can erode their validity as indicators of educational quality in accountability processes.

Many international comparisons claim to have insights that may be of use to policymakers in the jurisdictions that they survey. The reports and country rankings are taken as quick evidence of a country’s educational performance; the cautions about statistical significance offered by the reports are quickly brushed aside and replaced with sweeping statements of glory or shame (Jerrim, 2021). Such cautions, however, for using national measures to compare nations that may be very different culturally often go unheard by policymakers (Atkin and Black, 1997). International assessment tools produce information that can

be used for comparison, but how this information is interpreted and collected can be adjusted for many purposes.

Politicians with potentially short tenures will wish to demonstrate improvement in education standards, and national or international assessment data can be an effective means of collecting that information. Whilst the OECD reports the UK as a single country, Scotland has elected to be reported as a sub-identity (Jerrim, 2021). In 2018, Scotland received PISA assessment survey results that were reported as hugely disappointing in some media outlets and mixed in others (McEnaney, 2021). A large number of students (9.3%) were termed ineligible for this assessment survey event, which is thought to be due to the change of date for the assessments taking place in 2021 from March and May to October and December (Jerrim, 2021). Despite discrepancies raised about the data, some claim that conclusive evidence about the state of schooling in Scotland still exists.

In 2009, Mike Russell, the Scottish Secretary for Education and Lifelong Learning, portrayed the levelling of performance in PISA as ‘cast-iron’ evidence for improvement (Buie, 2010, paras 3–4). While comparative performance did not decline, it is questionable whether a stagnation in results constitutes improvement in standards. Mike Russell did, at the same time, however, note that the Scottish results were the ‘same as in England and Northern Ireland and better than Wales’ (paras 3-4). There may be differences of opinion about which data provide crucial information; however, this might not be the only desirable outcome of Scotland's participation in the PISA assessments. By virtue of participating as a separate nation in the PISA assessments, Scotland gained a level of independent political status in the world education ranking (Grek, 2009; Hayward, Macbride and Spencer, 2016).

Labelling a whole country’s education system as requiring improvement somewhat understandably generates a lot of internal reflection for politicians. The performance of a country on PISA assessment can cause international celebration, such as in the perceived success of Finland. At the same time, it can also provide challenging internal reflection and policy changes in reaction to poor performance (Sahlberg and Hargreaves, 2015). The PISA assessment results can serve as a convenient justification for politicians’ priorities and for

borrowing policies from nations that have performed well and have convenient education policies. Criticism has been made of countries trying to create ‘hasty and desperate’ strategies to move up the PISA rankings at any cost (Sahlberg and Hargreaves, 2015).

It has been cynically noted by some in the teaching profession that the PISA programme may serve no better purpose than serving to provide evidence for whichever point is favoured by the politician selecting the data (McEnaney, 2021). Hayward, Macbride and Spencer (2016) have carefully explored the justifications presented by politicians who use policy priorities of the day to select PISA data to support their ambitions. One example of this is the use by the English Secretary of State for Education, Michael Gove, to justify the redundancy of the National Curriculum in 2011. Politicians rarely acknowledge the cultural contexts of the data produced by PISA surveys, and in this case, it was also not acknowledged that there might have been pre-existing animosity towards the existing curriculum (Hayward, Macbride and Spencer, 2016). In this example, it may, therefore, be seen that politicians may utilise perceived trends in the data as justification to make policy changes that they may be motivated to develop anyway.

Criticisms of the PISA measures in Scotland include the extent to which they only measure what can easily be measured, such as knowledge recall from standardised tests (McEnaney, 2021) and properties of excellent schooling, such as pastoral care, personal development and equity (Sahlberg and Hargreaves, 2015). It has also been questioned whether the student samples are representative, as in countries such as England and Scotland, where so many schools asked to participate fail to do so. The UK is remarkable in the extent to which it has such a low overall participation rate (Jerrim and Sims, 2021).

When the scores from the PISA measurements might look to evidence a positive impact from government reform, and for policies they have enacted, it would be convenient for them to celebrate these and use them as justification of the policy changes. Trends that do not comprehensively support or show improvements because of education policy changes might simply not be referenced. In this way, although international assessments can be said to reflect a country’s educational quality, the data can be evaluated in a number

of ways to describe educational success or to evidence a need for reform. For control, PISA results can be used to support a wide range of educational arguments.

The selective deployment of evidence by political actors is not confined to educational assessment. Butler (2025, p. 5) examines how government decision-makers in UK contexts process policy-relevant information, arguing that political actors frequently engage in what he terms ‘motivated reasoning’. Rather than appraising evidence to reach valid conclusions about likely consequences, decision-makers assess information to confirm existing policy preferences. Evidence inconsistent with those preferences is dismissed or reframed; evidence supporting them is amplified. Butler demonstrates this pattern across three high-profile cases of UK governmental decision-making, finding that access to warning information did not prevent potentially damaging policies from being enacted.

2.2.3 Research evidence and the policymaking process

Policymaking is a process in which evidence may inform the developing policy, but it is not correct to say that the process occurs solely as a result of evidence (Smith and Haux, 2017). Ideas form part of the process, whereby evidence that is concordant with these ideas is selected or is more likely to be included as part of the rationale for the policy, whilst evidence that is less agreeable to the current ideology is less likely to be used or is not commissioned. Previous work has attempted to understand the processes of policymaking and how ideas and evidence may be incorporated into these processes (Smith and Katikireddi, 2013).

Parkhurst (2017a) develops this argument further by treating the politicisation of evidence as a structural feature of contemporary policy systems, rather than a localised failing within any single jurisdiction. He distinguishes between technical bias, in the production and selection of evidence, and issue bias, in which actors prefer evidence over other legitimate influences on policy. His framework of the good governance of evidence offers a route for examining how research findings enter and exit policy processes, and is returned to in the discussion of policy influencer accounts. Reckhow, Tompkins-Stange and Galey-Horn (2021) provided empirical support for this view in their analysis of teacher

evaluation policy in United States federal discourse. They examine how the provenance of evidence, including who funds and promotes it, weighs on policy outcomes. They argue that the political economy of knowledge production often outweighs the methodological quality of research is revisited in the analysis of policy influencer accounts.

Smith and Katikireddi (2013) developed a glossary of theories to aid in the understanding of the policymaking processes. They elaborate on the idea that policy might refer to ‘a context or broader direction (e.g. ‘free-market policies’) or a process, involving multiple stages (including implementation)’ (p. 198). They then classify policy change and the mechanisms by which policy can be thought to be enacted. These include phases of policy resistance and the agents that act within these processes.

This section has explored the multiple purposes of assessment, distinguished between formative and summative uses, and explored how the tensions between these purposes play out in policy and practice. It has also considered the role of high-stakes assessment and the challenges of international benchmarking. The relationship between research evidence and policymaking has emerged as a recurring theme, with evidence frequently selected or reframed to support pre-existing policy preferences. The following section turns from these systemic questions to the professionals who work within them, examining how teachers’ assessment literacy, beliefs, and professional identity shape their responses to assessment policy.

2.3 Teachers and the enactment of assessment policy

The preceding section explored the purposes of assessment and the tensions that arise when assessment processes serve multiple functions within a policy environment shaped by accountability. This section examines the professionals who work through these tensions. It considers how teachers' understanding of assessment, often framed in the literature as assessment literacy, intersects with their beliefs and professional identity to shape how they respond to assessment policy. It then draws on Ball, Maguire and Braun's (2012) enactment framework to explore how teachers mediate policy within their particular institutional contexts.

2.3.1 The development of assessment literacy

Stiggins (1991) defines assessment literacy as an individual's understanding of what it takes to create accurate and meaningful data about learning. This definition, encompassing the design, use and interpretation of devices such as tests to measure learning, has been reiterated in many other works (Fulcher, 2012; Looney *et al.*, 2018). Assessment identity is a fundamental aspect of practice for all teachers, as their knowledge of what assessment means and how to ensure its quality will inform what they choose to deploy and exclude from their classroom practices. A teacher's level of assessment literacy, their understanding of assessment principles and practices, can help predict how they are likely to approach classroom assessment activities.

Teachers who may be considered assessment-literate demonstrate a range of distinctive attributes that distinguish them from educators who lack knowledge of assessment principles and practices. DeLuca and Klinger (2010) make use of teachers' abilities to draw accurate inferences about learning taking place. If teachers can analyse assessment information more reliably, then the purpose for which the assessment has been deployed will likely be more successful. For example, if teachers can more accurately identify what their students do not know, they can more effectively teach to address these learning gaps. In addition, teachers can design targeted assessments that yield specific data to inform their pedagogical decision-making. Stiggins (1991) asserts that to be assessment-literate, decision-makers must be able to recognise the differences between robust assessment data and data that may have less meaning. When teachers develop assessment literacy, they are better positioned to create robust and valid assessment tools. If teachers are equipped with the knowledge of what makes good assessment processes, the chances are their assessment tools will be of quality (Popham, 2009). The underpinning pedagogical rationale for teachers' selection and enactment of assessment instruments is often obscured. While commercially produced curriculum packages frequently include pre-designed assessment tools, there is a risk that uncritical adoption of these resources may lead to inappropriate formative and summative judgements.

The relationship between a teacher's choice of assessment tool or navigation of assessment processes may not always be clear. Many commercially available

curriculum resources include assessment tools that teachers can use to make summative or formative inferences with little thought. Braun and Marion (2022) identified how teachers may attempt to improve their students' performance on accountability tests by purchasing commercially available assessment tools in a similar format rather than devising their own systems. In addition, past summative assessment materials sourced from published external examinations are widely used by teachers and are considered high-quality. Alderson and Wall (1993) found that teachers' use of past examination materials increases in frequency as examinations approach. Assessment-literate teachers may still use such materials. The concern is that reliance on pre-published resources may undervalue teachers' capacity as discerning developers and consumers of assessment data (Stiggins, 1991). The wider adoption of standardised assessment tools, particularly those derived from previous high-stakes examinations, can serve as an indirect mechanism for influencing pedagogical practice in classrooms (Popham, 2001). Teachers may operate within institutional contexts that constrain their professional autonomy regarding the deployment of assessment tools. These constraints may manifest in two ways: either through mandated assessment protocols that become the dominant model of practice or through imposed assessment requirements that conflict with teachers' professional judgements about effective assessment strategies.

Recent empirical work on teacher assessment literacy by Pastore and Mincu (2024, p. 14) has shown the development of what the authors term a 'measurement tradition.' Through interviews with Italian teachers, the authors found that teachers often view assessment processes as tools to measure and track learning, prioritising the creation of assessments with as much validity and reliability as possible and constructing a cumulative record of data that describes students' current achievement. Although the authors found that their interview participants could describe the contribution assessment could make to improve teaching and learning, there were no mentions of assessment processes used for formative purposes. The authors were not able to fully account for the 'crude conceptions of assessment' (p. 14) found, but suggest this may be attributable to a need for policy change and teacher professional standards in Italy.

If policymakers wish for teachers to make more use of assessment tools to improve classroom learning, teachers will need to be trained in crafting and deploying them effectively. While assessment literacy is increasingly recognised as a core competency for professional teachers (DeLuca and Johnson, 2017), questions remain about the extent to which pre-service training programmes prepare teachers effectively to design and enact robust assessment practices (DeLuca and Klinger, 2010; Xu and Brown, 2016). An exploration of practising teachers' assessment needs and experiences might provide valuable insights into the specific components of assessment literacy required for effective classroom practice.

The way teachers use their knowledge may differ as they enact these policies and the needs of their students beyond what is captured by assessment literacy. Looney *et al*. (2018) argue for considering teacher assessment within a broader framework, beyond assessment literacy alone. This can help us appreciate that teachers' beliefs about what makes effective assessment processes and how they should be used to serve their needs and those of their students can also impact how they choose to employ them in their classroom.

2.3.2 Teacher beliefs and classroom assessment

Teachers' beliefs can shape their practice and professional identity. In such a social profession, it is important to consider that teachers have beliefs that affect their responses to their school environment, as well as to their intellectual and physical reactions (Mockler, 2011). The history of teacher beliefs or perceptions and their impact on classroom practice has been described as an important theoretical basis for understanding how to drive change (Gardner and Galanouli, 2016). Teachers' beliefs may be shaped through a complex interplay of their personal experience, cultural norms and knowledge (Pajares, 1992; Richardson, 1996). Although teachers may hold beliefs about many aspects of their practice, assessment is among the most contentious areas (Hayward, 2012). It is, therefore, important to consider how a teacher's conceptions and beliefs may shape and impact their practice, and how these terms may be defined.

Teacher beliefs are a central aspect of their thoughts and actions, serving as personal truths that guide behaviour and decision-making (Pajares, 1992; Richardson, 1996). They can be compared to perceptions, which are the ways individuals interpret and make sense of their experiences and engage with new ideas (Gardner and Galanouli, 2016). The background and interplay between these ideas are useful for understanding how teachers might understand and interpret their practice.

This section explores how teachers' beliefs and perceptions function as a lens, shaped by their professional context and experiences. Bliem and Davinroy (1997, p. 1) explain how such a lens can:

> *'facilitate or hinder teachers' efforts as they set about altering their actions in the classroom, depending on whether and the extent to which their existing beliefs overlap with the philosophical underpinnings of proposed changes to their practices'.*

Previous sections of this thesis have discussed what assessment practice means and how policy may seek to influence it. This section explores how teachers' beliefs and perceptions function as a lens, shaped by their professional context and experiences.

Previous research on the success of policy reform has sought to understand the extent to which teacher actions and the decisions they make as part of their practice might be informed by their beliefs (Biesta, Priestley and Robinson, 2015). Understanding these sometimes-underlying beliefs can help us to analyse the behaviours of teachers who may follow the policy to the letter, whilst others might attempt to mediate them with their own belief systems. Through these mediations, multiple teachers may exhibit various practices within the same policy system, and their beliefs will affect their practice in different ways.

Not all forms of teacher resistance involve outright rebellion against policy demands. Ball et al. (2011) describe a negotiator identity in which teachers work within institutional expectations whilst attempting to preserve their own sense of what constitutes sound professional practice. This navigation can involve selective enactment, where practitioners comply with those aspects of policy they find defensible whilst paying lip service to requirements they consider less

important or educationally harmful. Eisenbach (2012) illustrates this pattern by describing teachers who publicly displayed policy-compliant work whilst quietly maintaining their habitual classroom approaches. The negotiator position, however, carries a psychological cost. Ball et al. (2011) note that such teachers may experience continuous turmoil, seeking to preserve their professional integrity whilst outwardly conforming to expectations that conflict with their values. Eisenbach (2012, p. 154) characterises this as an accommodator stance, in which a teacher's observable actions are no longer congruent with their personal beliefs. Where this dissonance persists without resolution, the teacher must either reframe their values to align with institutional demands or risk an unsustainable relationship with the profession.

The preceding discussion has explored how teacher beliefs are not fixed and can therefore be influenced at various points in their lives. Fang (1996) identifies multiple factors that influence teachers' beliefs in their practice, such as expectations of student ability and beliefs about a subject or theories of learning. These beliefs will affect practice and the choices that teachers make in their classroom. Understanding the impact of these factors and how they shape teacher beliefs might allow us to question what has influenced teacher practice.

Fullan (2017) describes the importance of acknowledging and incorporating teacher beliefs into future education policies. He goes on to explain that if policies are to be adopted by practitioners, then they need to be designed following engagement with teacher beliefs (Fullan, 2017, pp. 35–36). In agreement with this, Priestley *et al.* (2015) discuss how past experiences can impact beliefs and then play a role in teachers achieving agency in their practice. Working within a policy framework, teachers may find that their belief systems enable them to function successfully and achieve their professional objectives or serve their students well (Priestley *et al.*, 2015). Problems may arise when policy frameworks conflict with their belief systems.

One area of policy where teacher beliefs and policy may intersect is the use of assessment in classroom teaching. It may be asserted that teachers draw on past experiences to evaluate the use of assessment in certain aspects of their teaching. For example, Shulman (1986) uses the example of teachers managing

behaviour in a classroom to influence what we see as appropriate behaviour. He goes on to link this to how we judge good practice and how these practice models influence what we perceive as successful examples we might wish to emulate. In this way, teachers might use assessment practices that worked well for them, while also believing that those that did not work for them should be avoided.

Brown (2004) has argued that any policy, local or national, should attempt to take account of teachers' ideas about assessment. Brown argues that these perceptions have a 'complex structure' and require acknowledgement, warning that without such engagement, professional development alone may produce little change (2004, p. 314). This complexity is evident in the 525 questionnaires, which revealed multiple, sometimes conflicting perceptions held by teachers. Brown's findings raise the question of whether reform initiatives that fail to engage with teachers' existing perceptions of their practice can produce meaningful change.

Teachers' perceptions may also be resistant to change, as they may not be aware of how their perceptions act as a filter through which adjustments to practice may or may not be accepted and made (Bliem and Davinroy, 1997). A more recent synthesis of the challenges to enacting changes in assessment practice has also highlighted the need to consider the culture and the necessary changes in how teachers perceive their role as enactors of assessment and how assessment is becoming part of teaching and learning (Oo *et al.*, 2024). The influence is further supported by Correia and Harrison (2020) who found that science teachers who believed in enquiry-based learning adopted more flexible assessment opportunities for students.

Schmidt et al. (2017) describe clear perceptions among teachers regarding assessment, particularly formative assessment. For the teachers in this study, formative assessment provided vital information about how they might more effectively instruct their students and teach better lessons. These descriptions indicate that teachers value formative assessment for the information it provides about their practice. Not all assessments described, however, elicited the same positive response from teachers, especially when they served the purpose of accountability.

It may also be the case that teachers believe they are enacting formative assessment practices, but upon observation, gaps may emerge between their perceptions and their practice. Dayal (2021) completed a case study of the formative assessment practices of two secondary maths teachers and compared their observed practice with their beliefs on the purpose of assessment. Whilst this was a limited number of practitioners, the authors suggested it was possible to compare teacher perceptions with their practice.

A more significant study by Marshall and Drummond (2006) reviewed 27 lesson recordings of teachers and observed their formative assessment practice, comparing their practice with that described in interviews. They describe the difference between teachers following the spirit of formative assessment, promoting pupil autonomy, and the letter of formative assessment, where they merely follow the actions of formative assessment. Concluding the study, they found that there were relationships between the beliefs teachers held and their practice, such as teachers who embodied the spirit of formative assessment and a strong sense of agency for their students and an idea of growth for their learning. Their work supports Ball *et al.'s* (2011) description of policy recipients who comply with policy messages to the letter to please transactors who may hold them accountable for performing the actions detailed in the policy. On the other hand, Ball et al. describe policy critics who either rebel against or mediate the enactment of policy, for example, based on their perceptions of how it may negatively impact teacher workload.

This section has examined how teachers' beliefs may shape practice. However, these perceptions are not developed in the short term, and even dramatic policy changes may not have the power to stir changes in practice and perceptions (Hayward, Priestley and Young, 2004). Teachers may believe that they have made adjustments to their practice to adopt formative practices but make little connection between the practices and the meaning behind them (Hayward and Spencer, 2010). This disconnect can be especially problematic when accountability overshadows a teacher’s ability to receive meaningful support in developing assessment practices that accommodate their beliefs and experience (Hayward and Spencer, 2010).

2.3.3 Conceptions of teacher assessment identity

Teachers' beliefs are effective indicators of their day-to-day decisions (Dewey, 1997; Bandura, 2002). They can use their belief systems as a guide to build upon their theory of assessment and how it might be deployed (Xu and Brown, 2016). These beliefs are often so strong that they persist beyond training and can even persist when they differ from the purposes proposed in the training (Xu and Brown, 2016). Teachers may have the training to complete the correct assessments, but their practice may be affected by the belief that these assessments will not work (Looney *et al.*, 2018). Further work has shown that other teachers may follow a policy line but do not always understand or believe the spirit of the document. This is evidenced in work by Marshall and Drummond (2006), who term this as following the 'letter' of the policy but not the 'spirit', especially in terms of following the Assessment for Learning policy. Marshall and Drummond's (2006) line of argument suggests there may be many nuanced and external pressures that can affect what we see overall when observing teacher practice.

Looney *et al.* (2018) outline how a focus on assessment concepts can yield a psychological and sociological perspective on how they may manifest in observable classroom practices. Looney *et al.* (2018, p. 14) propose a new concept of assessment identity, which encapsulates the 'dynamic and interactive assessment identity constituted by beliefs, feelings, knowledge, and skills'. Figure 1 below illustrates the author's proposed model, which is a useful instrument for further analysis of teacher effectiveness. This model expands our understanding of the various facets that inform how teachers navigate assessment processes. It incorporates and expands upon the previous understanding of teachers as assessors.



Figure 1: Components of Teacher Assessment Identity. Adapted from Looney et al. (2018).

The framework's principal contribution lies in broadening conceptions of what it means for teachers to understand assessment processes. Where earlier constructs, such as assessment literacy (Stiggins, 1991; DeLuca and Klinger, 2010) focused primarily on the cognitive and technical dimensions of teachers' assessment knowledge, Looney et al.'s (2018) model integrates affective and experiential dimensions, beliefs about what constitutes effective assessment, feelings of confidence or discomfort when enacting particular processes, and perceptions of professional role into a single, integrated framework. This integration explains why teachers may relate their ideas about assessment to multiple factors in their lived experiences, rather than solely to their formal training. In this sense, the framework moves the analytical focus from what teachers know about assessment to how they relate to assessment as a dimension of their professional identity. This broader lens proves particularly valuable when exploring how teachers respond to policy change, as it suggests that resistance or adaptation may be driven not only by knowledge deficits but by deeply held beliefs and affective responses shaped over the course of a career.

This section has established assessment identity as a multidimensional construct that extends understanding of teachers' relationships with assessment beyond the narrower frame of assessment literacy. Looney *et al.*'s (2018) framework serves as a sensitising concept within this study, providing a conceptual vocabulary to contribute towards the analysis of interviews with teacher participants.

2.3.4 The enactment framework and teacher mediation

Policy creation acts through multiple processes whereby various agents incorporate ideas and evidence (Smith and Katikireddi, 2013). Incorporating and synthesising these ideas can occur at various layers of policy creation and interpretation as the policy intentions are transmitted from ideas to enactment. Conceptualising the various layers of actors in policy in education is useful, as it allows for an understanding of how policy might be mediated as it permeates or is influenced by these layers throughout enactment. Work to develop a framework for policy creation and enactment has been completed in the area of curriculum policy, so the contributions of various actors in enacting these policies might be understood.

Priestley *et al.* (2021, p. 13) frame sites of curriculum-making into five levels or sites of activity for the enactment process: 'Supra, Macro, Meso, Micro and Nano'. While the framework outlines a hierarchy with Supra incorporating international policy actors and Nano, including teachers and students, it isn't intended to suggest that policy enactment works downward from the top. Organising the various actors in a hierarchy such as this, however, is a useful framework to visualise how policy may be shaped by the actions of actors at various levels. Actors at the Nano level, such as teachers with particular subject interests or expertise, may work to influence national policy at the Macro level by participating in focus groups or being chosen as special advisors. Teachers will, of course, also act at the Nano level by mediating the enactment of policy in their own classrooms.

When exploring the role of policy actors in broader aspects of education beyond curriculum, the hierarchy of various policy actors discussed is useful. It helps to find the interconnectedness of policy influencers during the process of enactment as a social practice and classify the actions of various actors involved. This can then build to an appreciation of how these contributions work to amplify or mediate the success of policy enactment through national, local and even classroom-level activities. However, the nature of activities that shape the education aspect, such as curriculum and assessment, may differ and involve different organisations and individuals.

One of the ways in which discourse can be said to be shaped is through the creation of education policy, where policy is responsible for the creation of a state discourse (Codd, 1988). Consecutive governments in both Scotland and England have attempted to use education policy to reform education, and this has created an environment of perpetual policy change since the 1980s (Forrester and Garratt, 2016). The discourses shaped by policy processes can reveal the priorities and interests of those influencing policy and should be explored to reveal whose interests might be served by them (Gale, 1994).

The term "implementation" carries a number of embedded assumptions about how policy operates, assumptions that will be critiqued here. Ball *et al. (*2012) argue that the meaning of policy itself is frequently taken for granted or defined superficially as an attempt to solve a problem. This normative approach to policy analysis treats policy as a closed preserve of the formal government policymaking toolkit (Ozga, 2000). The language of implementation, therefore, assumes a linear, top-down trajectory in which policy is conceived at the centre and delivered, unchanged, to practitioners at the periphery. As Ball *et al.* (2012, p. 2) contend, this framing erases the contested, creative and mundane social interactions through which policy is negotiated, and renders school-based actors simply as actors who "implement". Similarly, Spillane (2004) observes that conventional accounts of policy tend to foreground rational-choice theory and personal interest, producing an understanding in which policy is either put into practice or resisted based on what practitioners decide suits their interests or those of their students. Such accounts of policy processes remain set within a linear, top-down idea of policy work in schools that views all policies, all schools and all teachers the same way in what Ball *et al.* (2012) describe as an institutionally and socially 'thin' account of policy processes. It is for these reasons that Ball *et al.* (2012) advocate for the concept of 'enactment', which recognises policy as diversely and repeatedly contested and subject to different interpretations as it is enacted, rather than implemented, in original and creative ways within institutions and classrooms.

As discussed in the preceding sections, policy texts are products of compromise and negotiation, open to interpretation and reinterpretation by those who read and act upon them. Yet Ball (1993) argues that policy simultaneously operates as

discourse, drawing on Foucault to contend that policy constrains what can legitimately be thought and said within educational practice. Policy texts and discourses are not alternative descriptions of the same phenomenon; rather, they operate concurrently, with texts providing the material for interpretation whilst discourses delimit the boundaries within which that interpretation occurs. This dual operation is central to the enactment perspective adopted in this thesis, as it positions education practitioners not as passive recipients of policy instructions but as active interpreters working within discursively constrained professional spaces.

Ball, Maguire and Braun (2012) develop this foundation into an analytical framework. They argue that policies do not arrive at schools ready to be adopted, but are enacted through processes shaped by professional memory, resource availability and institutional cultures. The framework therefore sets itself against accounts that treat all schools, all teachers and all policies as interchangeable. The enactment framework instead foregrounds the creative and constrained agency of professionals as they negotiate policy meanings within their particular institutional contexts.

Policy enactment in schools involves various actors within the social community, each taking on roles in mediating policy enactment. The way these actors adopt, mediate, or avoid policy processes can be influenced by their positions within the school and their various professional roles. Ball *et al.* (2011) define these roles within schools and how they operate within the policy enactment processes. These actor roles may have particular relationships with certain school professional roles, but the activities of these actors are not exclusive to those roles. The table below, adapted from Ball et al., outlines their suggested actor roles and the activities they might perform in the policy enactment process.

Policy Actor	Activities performed by these actors
Narrators	Take external policy communications, interpret and prioritise their messages. They are then often charged with ensuring compliance with these messages in school.
Entrepreneurs	Often involved in the process of launching policy initiatives, creatively designing ways of adopting the policy within the school.
Outsiders	Operating externally to the school, often providing advice of how policy might look, or be integrated most successfully and helping school leadership teams to evaluate policy enactment.
Transactors	Taking the role of ensuring policy is seen to be 'implemented' and collecting evidence to demonstrate that this is the case, especially in the case of collecting evidence for inspection.
Enthusiasts	Demonstrate policy success, perhaps through showing good practice and working to encourage school colleagues to adopt practices advocated by policy entrepreneurs.
Critics	Play a crucial role in mediating the impacts of policy, often with a concern for teacher workload and potential negative impacts of policy enactment processes. It can also involve the production of discourses in opposition to the processes advocated in the name of policy.
Receivers	Act gratefully to receive guidance on what is expected to perform successfully in their role as school professionals. They often work to the letter of policy messages in the hope of appeasing transactors in the leadership team who may evaluate their work in performance management processes.

Table 2.1 Policy actor types and their activities within policy enactment (adapted from Ball et al. (2011))

Through the work of these various actors, classrooms may be affected differently depending on the composition of classroom teachers, department leaders, and the school leadership team. Ball *et al.* (2011, pp. 636-637) point out that 'the school is not always sensible as the unit of analysis for policy research', and therefore, discussions of policy change on teaching practice should consider the roles of individual teachers and their roles as actors of policy enactment.

To understand the intended policy discourses that may have reached schools, the key policy texts for England and Scotland during the period of focus (1998-2023) will be explored in section 2.5. As described previously, this period was chosen as a key period for the creation of education assessment policy and for divergence between England and Scotland. The ways in which assessment processes are proposed in these documents will be explored to identify trends and patterns in their intended use.

In England and Scotland, high-stakes assessments have coincided with attempts to implement formative assessment policy. Formative assessment has been used as a method of practising for summative assessments, as well as a way of

collecting evidence for formative purposes (Stobart, 2008). These uses distract from or hinder the intended learning because the assessment is underway. Furthermore, the focus on grades or scores that the assessments produce can detract from learning, making it more shallow and superficial in the short term (James, 2008). Teachers may want to improve the learning of their students, but may feel forced to abandon a true focus on the substance of their subjects for the practice of high-stakes assessment.

A focus on learning and unifying the method of assessment and the intended learning is an important consideration for valid assessment (Harlen and James, 1997; Hayward *et al.*, 2008; Baird *et al.*, 2017). The professional development opportunities necessary for teachers to construct assessment opportunities and monitor the information produced by these assessments are often not available (Hopfenbeck, 2018). With an appropriate provision of time to allow teachers to reflect and improve on their formative practice, there is an increase in learning, but too often, not enough time is allocated to allow this to take place (Leahy and Wiliam, 2012).

The lack of professional time is also a limitation for developing effective summative assessment processes. Key facets for assessment to serve summative purposes, as mentioned previously, are validity and reliability. For summative assessment, the marking criteria and standards, as well as errors in teacher practice, can all create inconsistencies in the overall judgement that is made (Gipps, 2011). To avoid this, the moderation of assessment judgements needs to take place. Allowing teachers who are marking assessments to gather and discuss the standards required leads to greater consistency in deciding the allocation of assessment judgements (Harlen, 2007). Moderation, or standardisation as it is sometimes known, can allow all members the chance to question standards and test them against real-life examples. Without this process, there is likely to be misinterpretation of the criteria and mistakes in the application of the mark schemes (Harlen, 2007). Moderation can also serve as a useful form of professional development, whereby teachers can learn the assessed requirements and apply them to their teaching and formative assessment (Sadler, 1989).

One of the biggest barriers to teachers meeting to discuss assessment standards is the time and cost involved in the process. Methods that can be used to limit this cost include submitting samples to moderators, having moderators visit schools, and administering a common written test to estimate a mean and range for an assessed group. When moderation takes place, the calculated mean is assumed to be representative of the marking for the whole group. This, of course, is not always the case, and so some students may be marked correctly but have their grades brought down alongside their classmates through the moderation process. These processes do not really work to improve teacher assessment, but attempt to discover when a correction in marking is needed. In Scotland, the introduction of the 35-hour contractual statement on workload already causes tension between what must be done and what ought to be done by teachers as committed professionals (Mcphee and Patrick, 2009). If moderation processes were added to the workload, something would have to be removed. The desire and need for teachers to improve assessment practices are not always matched by the time and funding necessary to do so.

School budgets have been continually restricted within England and Scotland in recent years, which has led to a decrease in available training for teachers (Weston, 2018). School leaders may be hesitant to allocate funding to improvements in assessment practice, such as teacher assessment training if it will not directly influence performance indicators of exam results. A report by the Teacher Development Trust (2014) found that 21,000 teachers have received near-zero expenditure on CPD of any type, and across the sector, only 0.7% of school budgets were allocated to professional development (Teacher Development Trust, 2014). Another impact on the cost of professional development is the time required to embed the development process. The time required for professional development to embed into teacher practice and ‘transform’ practice has been cited as ‘at least 2 terms, more usually a year (or longer)’ (Teacher Development Trust, 2014, p. 12). To successfully enact professional development, school leaders must find a way to build skills in an educational landscape where time and cost are at a premium.

To allow professional development to take place despite time and cost constraints, there can be a rush to show teachers what to do and to expect

compliance. Harlen (2010) proposes several criteria for effective professional development in assessment. The first requirement is the time needed for teachers to meet and discuss new ideas, following input that helps teachers to reflect on the spirit behind assessment practice. By doing this, teachers will understand how to create and adapt strategies to suit their changing class needs (Harlen, 2010). In addition, teachers need the professional freedom to experiment and trial innovations in their classroom practice. This should be achieved through encouragement and shared clarity of goals between teachers and school leaders, with effective feedback during practice observations. To summarise, there is a clear risk that assessment development will be ill-planned, with top-down initiatives confining teacher practice. This risk can, however, be mitigated if teachers are given agency and collective accountability through reflection, in which the impact of developmental input and trials can be reviewed in ways that promote teacher professionalism (Gardner *et al.*, 2008). Wallace and Priestley (2011) praise the practice of allowing teachers to participate in the reform of assessment is for learning practice, rather than simply producing the activities requested. They highlight the Highland Council's 'empowerment' (Wallace and Priestley, 2011, p. 377) of classroom teachers, which enables them to tailor the policy reform to the needs of their own students and classrooms. For Policy enactment, however, there remain challenges to ensure successful reform.

Policy reform involving assessment can be extremely wide-ranging. For example, the *Assessment is for Learning* initiative in Scotland considered all assessment activities, including formative and summative assessments, and any activities that monitor student learning (Hutchinson and Hayward, 2005). The degree to which this policy was successful has been called into question many times, with some suggestions that much had been left unchanged in classroom practice (Hayward, Priestley and Young, 2004). Whilst some areas of strong policy uptake showed promise, there remained a lack of awareness or concern about how different parts of an assessment system interacted (Hayward and Hedge, 2005). In August 2010, Scotland's *Curriculum for Excellence* was launched, and *Assessment is for Learning* policies were subsumed within it. How these policies developed will be discussed later in section 2.5, but for many reasons, coinciding

with teacher salary decline and workload inflation, its enactment has not been seen as successful (Hayward, 2015).

In England, the *National Strategies* were published as methods to monitor and enhance children's progress. The Key Stage 3 project 'Assessing Pupils' Progress Key Stage 3' outlined assessment focuses which aimed to help teachers assess students at the national level descriptors for each of the *National Curriculum* levels with each subject (Qualifications and Curriculum Authority, 2010). In addition, handbooks were created with model examples to drive a consistent approach to the awarding of levels to student work. These levels were aimed to be a helpful guide for teachers to make a valid assessment of students as they progressed through the levels (Harlen, 2010). Despite the intention to inform assessment and create reliable teacher judgements, the levels began to be used as a method for accountability, where teachers and students were judged by their ability to move through the descriptors with demonstrations of performance (McIntosh, 2015). The pace at which students were expected to move through the levels increased, with teachers being held accountable for not moving students through levels fast enough. Eventually, teachers were pressured to move students quickly through the levels to demonstrate progress, which meant the criteria by which students could be judged to reach a particular level were 'widely misunderstood' (Oates, 2011, p. 128). This evaluation helped reform the 2014 *National Curriculum* in England, and from September 2015, *National Curriculum* levels were no longer compulsory in English schools (Gibb, 2015a). This marked a new period of education in England, in which teachers have to balance the professional freedom to devise their own summative assessments with the constraints of high-stakes, standardised national assessments. Teachers in England have been challenged to be creative with their progress measures, knowing ultimately that they will be held accountable for the results of external assessment at the end of schooling.

This section identifies many challenges that teachers and policymakers face in enacting assessment policy. There is a trend in policy enactment where the proposed reforms can claim to offer freedom and strip back perceived constraints in bureaucracy, and which are blamed for holding back student learning. This is combined with accountability measures rather than support in

the form of professional development and time. Consequently, some professionals may develop superficial practices that have not had time to be grounded in research evidence and to be reflected on its impact on student learning. When much of the work focuses on meeting these accountability criteria, meeting the policy can become hard work, and the professional workload may become unmanageable. Once this occurs, there may be calls for the next policy reform.

This section has explored the multiple purposes of assessment, the tensions between formative and summative uses, and the ways in which high-stakes accountability can distort classroom practice. Research evidence suggests these tensions are well-documented, yet policymakers have continued to rely on assessment as a primary accountability mechanism. The following section turns from these systemic questions to the professionals who work through them, examining how teachers' assessment literacy, beliefs and professional identity shape their responses to assessment policy.

When assessment processes are used for summative purposes, they should not influence the learning taking place. This section emphasises that when this terminal assessment is high-stakes, there is a temptation for it to significantly affect what is taught and how it is learned (Baird *et al.*, 2017). This can lead to shallow learning and an effort to cover the curriculum as quickly as possible, ultimately resulting in a loss of learning opportunities (Alderson and Wall, 1993). Richardson (2022) describes these effects as ‘assessment dysmorphia’ capturing the ways in which the pressure from assessment outcomes may drive a preoccupation with the results of high-stakes assessments. As discussed, this does not lead to better learning for students, and any perceived superiority in attainment between schools may be more attributable to students' socio-economic backgrounds than to school effectiveness (Muijs and Chapman, 2009).

The use of international assessment results is another form of high-stakes testing. Countries have used these assessments to benchmark their nation’s performance and appraise current education policies (Jerrim, 2021). This section describes how this can lead to policy borrowing between countries and the effects this may have. One example is policy borrowing without due attention to

subtle cultural differences, which may heighten the effects of policy in one country and dampen it in another (Ball, 2021).

In some countries, their positioning in international comparison assessments, such as PISA, has also led to increased accountability and further national deployment of high-stakes assessments. This section outlines that while international assessments provide policymakers with comparative data on their education systems, these assessments often have an indirect influence on classroom practices through policy changes rather than directly improving student learning. In fact, due to the washback effect of high-stakes assessments, this practice may have some unexpected consequences.

Lastly, this section on assessment academic literature considered the challenges to policymakers in enacting assessment policy in Scotland and England. The Scottish Assessment is for Learning (AiFL), and English Assessing Pupil Progress (APP) policies are aimed at increasing the information available for teachers to improve learning (Hutchinson and Hayward, 2005; Ofsted, 2011). Whilst these policies aimed to be ultimately formative, the use of APP to track students' progress meant that it was more commonly deployed as a series of summative assessments with large, unwieldy spreadsheets of data. In Scotland, the AiFL programme has been subsumed into the *Curriculum for Excellence*, with descriptors of sections of the curriculum that may be assessed. Whether the descriptors produced for CfE would follow a similar trajectory to England's national levels was, at the time, an open question, one that subsequent policy developments addressed.

In summary, the academic literature highlights that effective assessment policy changes require consideration of relevant research theories and authentic classroom enactment. However, the success of these changes may be limited by teachers' understanding and acceptance of the assessment processes. The following section explores how teachers interpret and enact assessment policies, drawing upon their professional knowledge to shape policy in their particular contexts.

2.3.5 The temporal dimension of teacher formation

The assessment identity framework, however, is not without theoretical limitations. As a conceptual model, it foregrounds individual-level factors, such as the beliefs, feelings, knowledge and skills of the individual teacher, yet gives comparatively little attention to the structural and institutional conditions within which assessment identity is formed and expressed. The policy environment, accountability pressures, school leadership expectations and the broader political context in which teachers operate are not systematically integrated into the model's architecture. This individual-level orientation raises questions about the framework's capacity to account for how teachers' assessment identities may be constrained, compressed or reshaped by external forces beyond their control. Whilst Looney et al. (2018) describe assessment identity as 'dynamic and interactive', the model does not fully theorise the temporal mechanisms through which identity changes over a career, nor does it address how successive policy regimes might produce layered or contradictory identity forms within the same practitioner. These are dimensions that require supplementary theoretical resources, resources which Le Grand's (2006) motivational typology, discussed below and in the following section, are better positioned to provide.

The interaction between assessment identity and policy context raises a further question: how are teachers positioned by policymakers within the systems they inhabit? Le Grand's (2006) conceptual framework, discussed in detail in Section 3 of this review, distinguishes between motivational assumptions that position public sector professionals as altruistic "knights" deserving professional trust and assumptions that position them as self-interested "knaves" requiring external oversight. Where policymakers construct accountability structures premised on distrust, the structural conditions within which assessment identity is expressed may be significantly constrained, and teachers may subordinate their assessment identities to compliance-driven behaviours shaped by surveillance and performativity (Ball, 2003). In other policy environments, teachers may be positioned as trusted professionals, and there may be greater space for the full range of assessment identity components to inform practice. This suggests that the relationship between assessment identity and policy context is not solely additive, but rather one in which teachers operate within

accountability environments that may shape which aspects of their assessment identity are expressed, suppressed, or perhaps reshaped over time.

Teachers operate within complex educational systems that shape their professional identities. This section has discussed how policy frameworks, institutional cultures and assessment regimes collectively influence how teachers perceive their role and enact professional responsibilities. Teachers' professional identities are not framed in isolation (Ball, 2015a), but through dynamic interactions between policy discourses, institutional structures and lived experience.

2.4 Accountability and educational change in England and Scotland

The preceding sections have examined the purposes of assessment and how teachers mediate policy through their professional knowledge and identity. This section turns to the accountability structures within which both policy design and teacher practice operate. It examines how competition, international benchmarking, institutional isomorphism, inspection regimes, and motivational assumptions about public-sector professionals shape the design and consequences of accountability mechanisms in England and Scotland.

2.4.1 Competition and marketised school structures

With the building of a national curriculum for England in 1988, the conservative government ensured that schools could be compared and thus, be more easily held accountable through school examinations (Leat, Livingston and Priestley, 2013). National tests enabled simple comparisons between schools and created a culture of performance to ensure curricular adherence (Leat, Livingston and Priestley, 2013). School examination results were first publicly published in England in 1992, allowing the public to view average secondary school results. Media outlets used these results to compile school league tables (West and Pennell, 2000). This kind of public accountability can lead to changes in behaviour among teachers and schools, as it is a crucial aspect of the management system for schools in England (Wiggins and Tymms, 2002). The

proposed consequences, both positive and negative, of this public league table system include enhanced competition and increased parent choice.

In England, a school comparison website (Find and compare schools in England (www.gov.uk) allows parents to quickly compare Primary and Secondary Schools based on their exam results and Ofsted results. The hope is that they will then be able to make choices about the school their child attends, rewarding the highest-achieving schools in these areas with increased student numbers and resources. One example is how school choice is facilitated through a formal request procedure every year, offered to all parents. Schools in England display key performance indicators such as exam results, pupil progress ratings and Ofsted grades on their website and at school gates. Ofsted ratings and performance indicators are extreme examples of a marketised education system in England, where collaboration might seem absent (Armstrong, Brown and Chapman, 2021). This use of school performance to compare schools is different in nuanced ways from how exam results are used for school comparison in Scotland.

Although Scottish schools face pressure to improve, league tables are less widely used as a mechanism for driving that improvement than in England (Wiggins and Tymms, 2002). Scottish schools are not encouraged to compete for pupil enrolment, and parents do not commonly seek places outside their local school (Wiggins and Tymms, 2002). Where parental preference is exercised, it operates through a formal request process rather than as a routine feature of the system, and nearly 90% of Scottish parents do not make such a request (Teelken, 2000).

This is not to say that school results and other individual differences do not make parents concerned about which school their child goes to, it is just that these concerns or requests are often more to do with the social demographic of particular schools, rather than their position in exam result league tables or ratings from external assessments (Teelken, 2000). In Scottish education, the school system's overall performance compared to other nations appears to be of great concern. Indeed, the then First Minister, Sturgeon, made clear that she considered school data and the publication of school performance essential to improving Scottish education.

There appear to be differences between the two nations in how competition is encouraged or discouraged and with whom. Where competition is encouraged, Governments expect it to drive school improvement, pushing schools to raise their grades at almost any cost. Whilst central government may encourage competition to drive school improvement, local collaborative structures are also increasingly used by the government to improve school performance (McEnaney, 2021).

England has undergone a dramatic diversification of its schooling structure. The New Labour government introduced Academy schools to replace schools that had been considered as failing by multiple Ofsted inspections. By becoming ‘sponsored’ academies, these schools were part-funded by businesses, individuals, churches, and voluntary bodies, with the main source of school income provided by the central government. In 2010, the Academies Act allowed all schools to become voluntary academies, removing the Local Authority's control and associations and granting the Headteacher and school Governors greater strategic and financial autonomy. In England, school governors have a role in school leadership, setting teacher pay and holding school leaders accountable. In Academy schools, school governors are referred to as local advisory boards, whose function is primarily to establish and monitor how the Academy Trust policies are being enacted within the school.

England's structural arrangements appear to deliver the autonomy that policymakers have prioritised, yet several mechanisms complicate that picture. The absence of a local authority middle tier leaves school leaders with considerable discretion over how their schools are run (Kreijkes and Johnson, 2023). Sibieta and Jerrim (2021) note, however, that around three-quarters of academy schools belong to a multi-academy trust. These trusts can function as governing bodies, take decisions across school operations, and even inspect their constituent schools. They further observe that Regional School Commissioners oversee academy conversions, authorise free schools, and monitor performance. Ofsted, meanwhile, retains oversight of curriculum breadth across all state-funded schools. Taken together, these arrangements raise questions about how far autonomy has genuinely been devolved to school leaders in England.

In Scotland, a simpler system of school structures exists. Most schools are publicly funded, with the Local Authority responsible for funding, admissions, and governance. There are also a small number of grant-aided schools, which are usually special schools and are independently governed. Schools in Scotland appoint a Parent Council, which is used to inform school improvement by providing representation of parent views to school leaders (Education Scotland, 2023). Compared with England, Parent Councils are more advisory and provide the parent voice to school leadership rather than serving an accountability function.

Much of this national and local-level policy change can affect school structures and identity, but the overall effect on classrooms may not be as evident. Policy enactment may be said to occur more generally within individual classrooms and lessons. It is, therefore, important to consider how teachers might respond in different ways to enact, mediate or resist educational change.

Making changes to policy from the central government to enact revolutions in classroom practice is a challenge and great expectation. As detailed in the previous section on policy change, there can be quite dramatic reforms whilst very little change is noticed in classrooms. Despite this, school effectiveness and school improvement strategies continue to aim to change classroom practices to drive school change.

The assumption that changes to school policy enacted under pressure from national policy will be sufficient to alter classroom practice has long been challenged in the literature on school improvement. Ainscow, Barrs and Martin (1998) explore some of the difficulties in implementing classroom change through staff development initiatives, particularly in a secondary school. The larger staff numbers and the difficulty in establishing a shared language of practice make it hard for all staff to interpret the expected changes or understand the philosophy behind them. The idea of teachers having difficulty with accommodating changes to practice in a permanent way has been explored previously, with the section on assessment and the ways in which formative assessment has been difficult to establish securely in some Scottish classrooms (Hayward, Priestley and Young, 2004; Hayward and Hedge, 2005).

Chapman (2011) argues that there has been a recent shift of focus from advances in teacher and classroom practice to the influence of policymakers for changes in the whole system. With this shift, he also identifies an emerging gap between the fields of research on educational change and the priorities of schools and teachers. If policymakers are hoping to make school changes to ensure that they improve, it is vital that the changes can coincide and not inadvertently contradict the efforts of schools. It may be evident that when policies are enacted that do in fact contradict teachers' priorities and their students' needs, they are more likely to be ignored or only enacted to the letter, not the spirit of the policy. Fullan (1993) explains the difficulty of policy 'mandating what matters' (p. 22), as teachers are professional workers with experience and pedagogical training rather than technicians accustomed to following scripted actions. In addition, many of the changes required of teachers take place in challenging contexts, amid the variables that characterise classrooms filled with children.

The literature suggests that schools serving challenging communities face particular difficulties in sustaining classroom-level change. Thomson (2002) examines evidence of policy being both accepted and rejected in such contexts, noting that accountability-driven competition for scarce resources can reduce cooperation between neighbouring schools (p.132). Senior leaders in these settings occupy what Thomson terms a 'policy mediators' role (p.137), selectively engaging with policy elements that serve their communities whilst resisting those perceived as detrimental. This pattern of selective enactment, conditioned by context and resource constraints, is significant for understanding how accountability policies encounter differential responses across challenging school environments.

Beyond headteachers, teachers, parents and students may also resist policy control, exercising democratic rights through trade unions, subject associations and pressure groups (Wrigley, 2006). Where personal philosophies conflict with the spirit of policy documents, teachers may operate in ways that run counter to stated expectations, even under the scrutiny of accountability mechanisms such as Ofsted (Wrigley, 2006, p. 9). Teachers therefore act as agents within the policy environment, bringing their own belief systems to bear on practice.

Understanding what takes place in classrooms consequently requires an understanding of the teacher, which the following section explores.

2.4.2 International benchmarking and policy borrowing

Both Scotland and England participate in many international assessments to benchmark the quality of their education programmes against those of other nations. One of the most rigorous of these programmes is the Programme for International Student Assessment (PISA) conducted by the Organisation for Economic Co-operation and Development (OECD). Section 4 of this chapter traces how Scotland and England have used PISA results to justify policy reforms. This section examines the broader patterns of international benchmarking and policy borrowing that underpin those uses. to justify policy reforms.

In Scotland, there is a drive to improve the quality and equity of schools, which is a common use of international comparisons (Schleicher, 2009). In addition to the PISA tests, Scotland has invited the OECD to conduct reviews of its education system, particularly in comparison with other OECD countries. In this way, Scotland uses international assessments and the OECD as a basis for international comparisons and benchmarks, taking great stock of the recommendations. It is important to note that the OECD was created to raise economic growth and employment, and therefore sees education as a tool for that purpose (Knodel, Martens and Niemann, 2013). The OECD provides a form of Governance to Scotland and, in Knodel, Martens and Niemann's (2013, p. 423) terms, acts as a ‘policy entrepreneur’.

England has also used the results of PISA and the Trends in International Mathematics and Science Study (TIMSS) to benchmark its educational performance. As discussed in the previous section on education policy, these have not been used to make major shifts in English education policy, but rather to substantiate points to make England more competitive. One example of this is the speech in 2015 by the School’s Minister Nick Gibb (Gibb, 2015b). In this speech, the OECD is praised as ‘being at the forefront of producing high-quality international indicators’(para. 3) and then used to justify the market model of schools, namely, academisation. The justification is that with ‘increased autonomy’, higher results and raised standards will follow. This speech gives

examples of how the academies programme is influenced by handpicked countries, such as Sweden and the USA, which have used local school deregulation to drive improvements. Handpicking policies and applying them without consideration of unique cultural and regional differences has been termed policy borrowing (Ball, 2018).

Using selective policy borrowing and international comparisons, both Scotland and England seek to draw on ideas from abroad to support their own education policies. This is, of course, a top-down control method, consistent with the school effectiveness research mentioned previously. Other methods of eliciting education change include monitoring and inspection visits through the Office for Standards in Education, Children's Services and Skills (Ofsted) in England and in His Majesty's Inspectorate of Education in Scotland (HMIE).

2.4.3 Mimetic isomorphism in assessment policy

The concept of institutional isomorphism (DiMaggio and Powell, 1983) provides a framework for examining how organisations adopt similar practices under different forms of pressure. In education, coercive isomorphism operates through statutory requirements and inspection regimes that compel schools to adopt prescribed practices. Mimetic isomorphism emerges when expectations are ambiguous, and schools look to perceived exemplars for guidance. Normative isomorphism arises through professional bodies and networks that define standards of good practice. All three forms are visible in the assessment policy landscapes of England and Scotland, though their relative prominence differs between jurisdictions. In England, the removal of national curriculum levels in 2014 created conditions for mimetic isomorphism, as schools sought commercial or externally validated frameworks to replace government-prescribed assessment systems (Poet *et al.*, 2018; Staufenberg, 2018). In Scotland, the broad expectations of Curriculum for Excellence produced comparable uncertainty, with teachers reporting insufficient clarity around standards and expectations (Hayward and Hutchinson, 2013; Priestley and Minty, 2013). The following subsection examines how inspection regimes in each jurisdiction operate as mechanisms of coercive isomorphism.

Working in environments with multiple influences from isomorphic factors, schools are challenged to compete or collaborate. When ‘best practice’ is ill-defined, some institutions aim to collaborate and establish for themselves, whilst others may pioneer this by themselves in an aim to out-compete others and gain an advantage. These various policy environments can create aspects of practice unique to each nation of the United Kingdom by fostering different isomorphic environments. From their various points of difference, governments look for educational improvements by benchmarking themselves against other countries they may choose to be more closely associated with.

2.4.4 Inspection regimes and accountability mechanisms

Inspection in both Scotland and England has followed a line of distinctiveness that predated devolution. In Scotland, the HMIE operated to provide direction for the education system and was a ‘major force’ in the landscape of policy decisions and improvement (Ozga and Lawn, 2014, p. 15). In England, the inspectorate has always maintained its independence from the government, with HMIs appointed by order of the Privy Council. Unlike in Scotland, England’s HMI was reformed into Ofsted in the early 1990s to help drive output-based performance indicators along with wider public sector reform. Ofsted now assigns private sector providers to carry out inspections of state institutions rather than HMI education leaders, as educationalists were seen by the government as ‘elitist, more focused on influencing government than on schools, and vulnerable to producer capture’ (Ozga and Lawn, 2014, p. 14).

Ofsted conducts school inspections through conversations with staff, students, governors, and parents, as well as classroom observations. Following these inspections, a report is produced that grades the school's different aspects as Outstanding, Good, Requires Improvement, or Unsatisfactory. The inspections create significant pressure on teachers to ensure that their observed lessons are judged Good or Outstanding (Leat, Livingston and Priestley, 2013) school judgements are publicly available on the Ofsted website. Whilst Ofsted exerts significant pressure on schools to improve, this is not backed by support or extra resources; in short, the effects on school leaders and teachers can be hostile and critical (Chapman, 2000). Perhaps it can be considered more of a summative assessment of a school, creating a strong driving force for improvement.

It has been argued that the premise of a school's autonomy to develop their own routes to improvement is not possible under the Ofsted framework (Thompson, Lingard and Ball, 2020). Schools that do not follow the guidelines for schooling published in the Ofsted inspection handbooks can face enforced academy conversion and a reduction in autonomy (Mills, 2015), have their headteacher dismissed or even closed (Courtney, 2016). Under these conditions, the only schools that might consider themselves free of the pressures of 'pleasing' Ofsted are those that have been rated Outstanding and are currently exempt from regular inspection. These schools may feel freer to develop practices and cultures without periodic judgements from Ofsted; however, even if they know that their exemption does not mean Ofsted will never visit, this can still happen for a variety of reasons. These schools will be reaping the rewards of being perceived as at an advantage in a competitive school market, attracting parents and students with Ofsted Outstanding ratings on their websites and even at school gates. They will also be anxious to keep this at all costs or else lose the competitive advantage it bestows.

With the Scotland Act in 1998, the inspectorate body became an agency of the Scottish Government in 2001. In 2011, HMIE merged with Learning and Teaching Scotland, becoming the agency now known as Education Scotland. This formation of a government body to support the enactment of curriculum and build capacity for school improvement occurred at a time when similar bodies in local education authorities in England were being dismantled. Education Scotland does not just exert pressure on schools to improve, as Ofsted in England does; it also aims to build support and capacity for improvement.

Scotland's HMIE inspectors form part of the government's mechanism to evaluate the output of policy, including curriculum development. In this way, the inspectorate is used to monitor the progress schools are making towards achieving the changes set out in policy. This is one way in which both England and Scotland use inspection tools to ensure compliance with policy demands and to secure what the government sees as school changes that will lead to improvement through greater accountability (Leat, Livingston and Priestley, 2013; Thompson, Lingard and Ball, 2020).

Accountability is a means of evaluating teachers and schools based on student results (West and Pennell, 2000), inspection judgements (Leat, Livingston and Priestley, 2013; Thompson, Lingard and Ball, 2020) and other devices to summarise school effectiveness. The mix of these mechanisms for judgement and how they are used to drive accountability systems has varied and has changed for both Scotland and England since 1998. England might be seen to have continuously pursued outcomes-based accountability mechanisms using standardised national assessments since the Education Reform Act in 1988 (Boyle and Bragg, 2006). Conversely, Scotland has often been described as operating in two modes, with accountability initially framed through professional conversations of performance and improvement (Arnott and Menter, 2007), with others noting concerns of influence from external assessment results and competition among schools (Hutchinson and Young, 2011; Hayward and Hutchinson, 2013; Hayward, Macbride and Spencer, 2016). Whilst there may be similarities in the way assessment and accountability mechanisms have operated in Scotland and England (Leat, Livingston and Priestley, 2013), the mechanisms behind these operate differently.

In England, school improvement is reduced to easily measurable parameters, with great emphasis on improving examination results, in a style that reduces teachers' roles to technicians pulling the right levers to generate adequate test scores (Wrigley, Lingard and Thomson, 2012). Through the marketisation of school choice and the allocation of funding based on pupil numbers, there is a strong driving force in England to raise grades at all costs, with no excuses for failure (Gray, 2010). Excellent schools are seen to perform well, as measured by examination results, and are held to account by parents choosing the best school for their children, which is seen as the best route to school improvement. This accountability has necessitated strong agency for headteachers in England, where powers over admissions, staff pay, and wider school budgets are now in their control. This drive for school autonomy and competition has seen schools move from local authority control to different structures of free schools, multi-academy trusts and remaining local authority schools in a complex policy environment, discussed in further detail in section 4 of this chapter.

Compared to England's diverse structures of school types, many of Scotland's schools remain in local authority groups. There is, however, diversity of opinion on how autonomy and accountability balance for schools in Scotland. Some have described a landscape of relative school agency gifted through the flexibility of Curriculum for Excellence (Leat, Livingston and Priestley, 2013), leading to possible confusion over the understanding of learning standards (Hayward and Hutchinson, 2013). The 2000 Standards in Scotland's Schools Act awarded control of school improvement to school boards (later replaced by Parent Councils with the Scottish Schools (Parental Involvement) Act 2006); however, since then, the influence of school governance and improvement has been in the hands of the teaching profession and local authorities (Arnott and Menter, 2007). Other researchers have questioned the extent to which schools have autonomy in Scotland, where many national agencies, such as Education Scotland, formed in 2011, are responsible for curriculum policy, school inspection and driving school improvement (Kreijkes and Johnson, 2023).

Whilst accountability mechanisms attempt to drive educational change, as the preceding discussion suggests, both intended and unintended consequences are debated. In both England and Scotland, they work through structures and hierarchies of institutions that need to be understood to better explain how educational change might be achieved. Central to those structures are the assumptions policymakers make about how professionals are motivated and how much autonomy they should exercise, assumptions that shape the very design of accountability itself.

2.4.5 Motivational assumptions in accountability design

Le Grand's (2006) conceptual framework examining motivation and agency in public policy provides a valuable analytical lens for understanding how policymakers construct professional roles within education systems. Developed initially to analyse welfare state transformations in Britain and internationally, Le Grand's framework identifies two fundamental dimensions along which policymakers position public sector actors. Figure 2 illustrates two fundamental dimensions along which policymakers position public sector actors: the motivational axis (ranging from 'Knights' to 'Knaves') and the agency axis (ranging from 'pawns' to 'queens'). This framework reveals how implicit

assumptions about professional motivation and capacity for autonomous action fundamentally shape policy design, accountability structures, and enactment strategies.

The motivational dimension distinguishes between conceptions of public service workers as altruistic ‘knights’, professionals primarily motivated to serve others’ interests, and self-interested ‘knaves’, actors principally concerned with advancing their personal gain. Le Grand argues that these assumptions, whether explicit or implicit, profoundly influence the design of policy instruments. Policies constructed on knightly assumptions typically emphasise professional autonomy, trust-based accountability, and collaborative governance structures. Conversely, policies premised on knavish assumptions deploy external oversight mechanisms, performance metrics, and competitive frameworks designed to align self-interest with policy objectives through market-like incentives (Le Grand, 2006).

The agency dimension addresses assumptions about individuals' capacity for independent action. Positioning actors as 'pawns' suggests viewing them as passive recipients of policy directives, victims of structural constraints with limited autonomous capacity. Treating actors as 'queens' acknowledges their status as active agents possessing substantial power to shape, resist, or transform policy initiatives. These assumptions about agency influence whether policies attempt to direct behaviour through prescription or enable autonomous professional judgement within broader frameworks.

Le Grand's framework emerged from analysis of ideological transformations within British welfare state policy, particularly the transition from social democratic to neoliberal governance approaches during the 1980s and 1990s. Social democratic traditions, dominant in the post-war welfare state architecture, rested on assumptions that positioned public sector professionals as knights serving passive beneficiaries (pawns). Professional autonomy was valued, with policy design reflecting confidence in practitioners' public-spirited motivations and expert judgement. Neoliberal ideological shifts challenged these assumptions, arguing that self-interest constitutes the more reliable motivational foundation. This perspective advocated replacing trust-based professional discretion with market-based accountability structures designed to harness self-interest through competition (Le Grand, 2006).

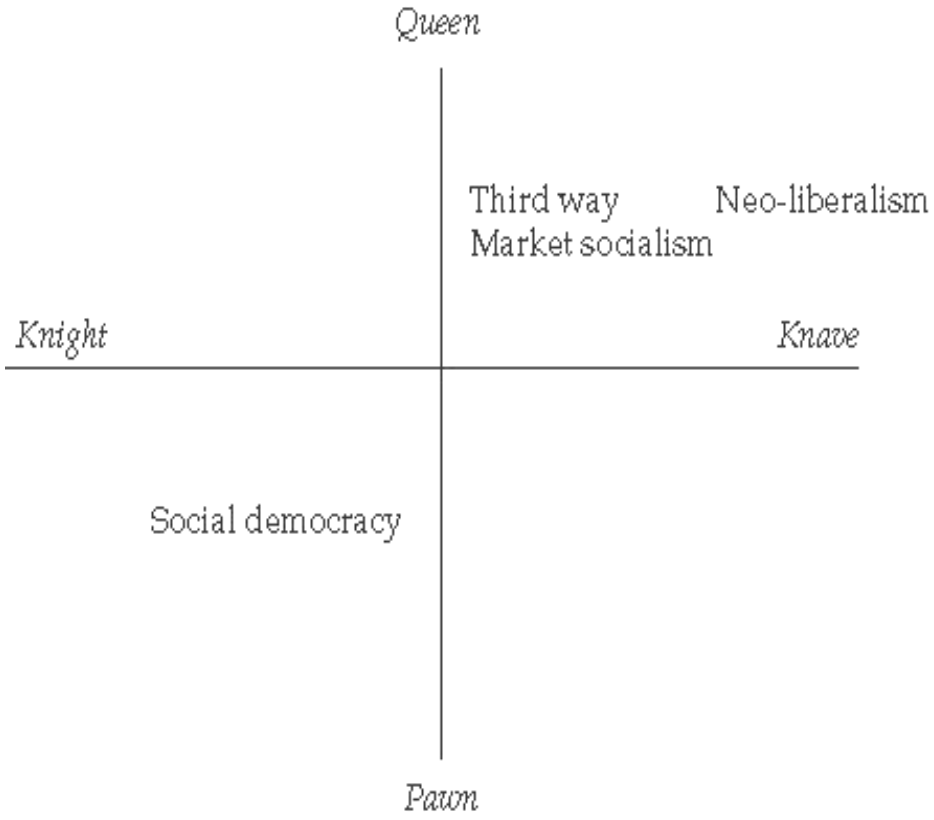


Figure 2: A visual conceptualisation of human agency and motivation taken from Le Grand (2006, p. 16).

The quasi-market revolution in English education exemplifies this ideological transformation. The 1988 Education Reform Act fundamentally restructured school governance by introducing formula funding tied to pupil numbers,

parental choice mechanisms, and the publication of performance data. Le Grand (2006) argues these reforms reflected a shift in assumptions about teacher motivation, from knights deserving professional trust to knaves requiring external constraint through competitive pressures. Simultaneously, parents transitioned from pawns with limited influence to queens exercising market power through school choice.

This section has examined how accountability mechanisms in England and Scotland operate through different structures of competition, inspection, and institutional isomorphism, whilst sharing an increasing reliance on assessment data to evaluate school performance. The following section traces how these accountability pressures have shaped the chronological development of assessment policy in each jurisdiction from 1998 to 2023.

The mechanisms of policy must move beyond intentions documented in procedures that schools must adhere to in order to be successful. Research reviewed in this section suggests that schools in challenging contexts must mediate oversimplification or a severe lack of understanding of the community challenges schools operate within. Resistance from these communities and the teachers who serve them shows that unabridged policy borrowing is ineffective for bringing about meaningful educational change. Too often, it leads to opposition and even rebellion. In England, this opposition sought to build an alliance of teachers and parents against the system of external testing and league tables in Primary schools (Coughlan, 2003). The enactment of policy across a nation remains a challenge for education reform, especially when it is insensitive to cultural needs and values, as seen in trends between England and Scotland since 1998.

In England, changes to how accountability operates have shifted accountability from inspectors who hold schools to account to local authorities and advisors who enforce improvement. This has shifted to academy systems, with schools competing for students, promising better funding and well-qualified staff. When students become routes to funding, attracting them becomes a marketing strategy, as seen in school advertising campaigns in local press, open events, and websites proudly displaying a school’s favourable Ofsted rating. As discussed above, schools may be forced to join multi-academy trusts to share resources

and return to a state of semi-cooperation with those in the same ‘club’ or Multi-Academy Trust, as they are more formally known. The few schools that remain outside these Trusts may lack the economies-of-scale advantages and face a more challenging funding situation on their own. Some Multi-Academy Trusts have even secured access to Newly Qualified Teachers (NQTs) and further funding routes through incorporating teacher training within their Trusts in alliance with universities.

In Scotland, the pressures of school choice are present but much less overt than in England. Local authorities and HMIE hold schools to account with set key performance indicators, including external examination results. In a similar way to England, then, Scotland attaches importance to academic output in schools, with performance in these exams seen as a key indicator of a successful education. The similarities between England and Scotland also include the use of regional improvement collaboratives to drive educational change. Although this form of collaboration does not involve changes in legal status for schools or the removal of local funding, as it does in England, it does allow schools access to similar benefits of scale, such as pooled CPD resources. Though these collaborations do not go as far as changing the local structures of education in terms of local authority control over schools, this may be a further development in the future.

It can be quite rightly said that much of this educational change takes place beyond the classroom and may not actually affect what happens there. How these mediate from school policy to classroom practice can only be mediated through the actions of the teachers operating there. Teacher development stemming from policy changes must factor in a teacher's ability to adopt these changes, taking into account professional knowledge, beliefs, experiences, and roles within the school.

This section has examined how accountability mechanisms in England and Scotland operate through different structures of competition, inspection, and institutional isomorphism, whilst sharing an increasing reliance on assessment data to evaluate school performance. Le Grand's (2006) framework has provided a vocabulary for examining the motivational assumptions embedded in these accountability designs, distinguishing between policies that position teachers as

trusted professionals and those that treat them as requiring external oversight. The following section sets out the chronological development of assessment policy in each jurisdiction from 1998 to 2023, tracing how these accountability pressures have shaped the policy landscape that teachers navigate.

2.5 Policy context in England and Scotland 1998 to 2023

The preceding sections have examined the purposes of assessment, the professionals who enact it, and the accountability structures within which both operate. This section traces the chronological development of assessment policy in England and Scotland from 1998 to 2023, drawing on official documentation and contemporary reporting to examine how successive governments have shaped the assessment environment that teachers respond to.

2.5.1 The 1998 starting point and the rationale for comparison.

In this study, education policy encompasses both the written text and the actions taken in its enactment. Ball (1993) argues that policy is always in a transient state, its effects persisting long after they were intended to expire. He further observes that policy is often enacted to resolve a problem for those who design it, while posing distinct problems for those subject to it (1993). For this reason, the study examines the policy effects of the two decades under scrutiny here. How education professionals operate within their practice following policy change may indicate significant shifts in approach, as well as limited improvement in student learning outcomes. Cuban (Cuban, 1993, p. 2) explored the degree to which national policy change could shape classroom practice. He compared the shifting policy landscape to a hurricane that generates waves of change on the surface, whilst leaving the ocean floor of classroom practice in serene calm.

Educational change does not always occur immediately following policy changes. It is useful to understand which conditions are conducive to change through policy processes and which may impede them. With a better understanding, we might create more effective policy environments and drive improvements in education.

Attempting to understand the multiple intentions behind policy texts is also essential to understanding how these intentions might be enacted in school classrooms. Another area of interest in considering policy design and its impact on teacher practice is the difference between national policy and single-institution localised policy. Ball (2017) has defined this difference as policy with a big ‘P’ for national policy and a small ‘p’ for local policy. Policy changes enacted may operate at various scales and seem to follow themes termed periodisation, especially in the context of teacher development (Menter, Mutton and Burn, 2019).

In discussions of policy, it can be defined as a principle or guide used in public to inform decisions and actions within an organisation. Policies can often be written into texts, documents, or laws to provide guidance and clarification on specific circumstances or issues. Whilst policy is often simplified as a textual tool to solve certain problems, this work will later explore other mechanisms by which policy can act and exert influence, rather than merely exist as written documents (Braun, Maguire and Ball, 2010). In this section, the policy texts produced between 1998 and 2023 will be explored.

The policies developed since 1998 have been chosen throughout this work because of the significance of 1998, as explained in the introduction. This time brought devolved education powers in Scotland, allowing further differentiation from England in education policy. In this section, these times are explored through key assessment policy texts and the purposes for which they describe assessment. Key themes in these documents will be explored, especially their descriptions of how assessment processes should be used in classrooms and how they should not. In particular, the justifications for these processes in terms of why these processes are necessary and how they might operate will be explored to identify any differences between England and Scotland, should they exist.

The first period of policy texts described will be England, beginning with the New Labour government. Key policy developments aimed at affecting assessment at the national and classroom levels will be traced through 2023, after which this section will discuss the same period for Scotland.

2.5.2 Assessment policy in England under New Labour, 1997 to 2010

The New Labour government was elected in 1997 with a policy manifesto focused largely on the mantra of ‘education, education, education’ (Walford, 2005). It has been argued that this mantra of rhetoric and the prioritisation of confronting social inequalities led to a clear and distinctive new agenda (Arnott and Menter, 2007; Gunter and Chapman, 2009). The extent to which this is the case, however, should be explored critically following an account of some of the key New Labour education policies developed at this time. This analysis is necessary because national policy may not only produce its intended educational changes but also generate conditions that give rise to unintended consequences. Understanding the impacts of past policy enactment can help develop an understanding of the contemporary education landscape and offer caution for future planned policy development if heeded (Smith and Katikireddi, 2013). The focus of New Labour policies for the purpose of this work will be limited to assessment and how it is intended to be proposed.

In an attempt to improve public services, the New Labour Government declared that it would follow a more evidence-based approach to olicymaking (Blunkett, 2000). However, research on the extent to which this was more than political rhetoric has been challenged by a range of sources (Tomlinson, 2005; Wells, 2007; Heath *et al.*, 2013). The role of ideas in policy formation has been recognised as an important ingredient in the olicymaking process. It has been argued that within health, for example, ideas often influence policy just as much, if not more, than evidence (Smith, 2014). Ideas and values have limited the extent to which evidence is included in policy creation and how it is presented to government ministers, considering ideology and what might be considered favourable (Smith, 2010). It would seem that ideas can play a role in shaping the extent to which policy can be influenced by evidence, and this isn’t just in health policy.

In educational policy, school choice and the idea of parental rights to choose their children’s schools have also played a role at the national level. In England, New Labour was keen to stress parents' rights to choose the school they believed was right for their children (Bhattacharya, 2021; Brighouse and Waters, 2022).

This concept of parental choice from New Labour has influenced neoliberal policies for inspection and school funding, as discussed earlier. It also positioned schools in direct competition for students, with each vying to increase the number of parents choosing their school as a marketable product. Although during this period, New Labour asserted ideas for reducing exclusion rates in schools, it was noted that issues around the exclusionary effects on pupils who may not be able to conform to and perform in their local schools (Ozga *et al.*, 2023). This was catalysed by the production of league tables in 1993 and led to such an increase in exclusions that the government had to establish networks of Pupil Referral Units (PRUs) to provide temporary education (Ball, 2021). In England, marketisation extends to the point of school selection itself, where every family is required to submit an ordered list of preferred schools. By contrast, 87% of Scottish families do not make a formal placing request (Bhattacharya, 2021). This pattern reflects a wider divergence in how market principles have been embedded within each system. English schools operate within a competitive framework in which exclusion rates and parental choice mechanisms reinforce market-driven behaviours, whereas Scottish schools, shaped by a stronger discourse of inclusion, exhibit markedly lower rates of both (Ball, Bowe and Gewirtz, 1996). The interplay of ideas, values, and evidence within each policy environment shapes what is considered effective practice and, in turn, the framing of policy research (Smith, 2014).

The New Labour Government quickly introduced changes to the national curriculum and a renewed focus on literacy and numeracy, especially in the Primary phase (Heath *et al.*, 2013). This refocusing on so-called core subjects marked a change in the curriculum that preceded the publication of a new national curriculum the following year. The National Curriculum document produced by New Labour (Department for Education and Employment, 1999) introduces the use of level descriptors to make judgements about pupil performance as they move through key stages of England's schooling. Although there is a caution about using the descriptors to measure individual pieces of work, this warning foreshadows some of the ways in which level descriptors were used. It has been noted that 'two years into the Blair government, the quest for 'education, education, education' was alleged by some to be 'results, results, results' or 'targets, targets, targets' (Brighouse and Waters, 2022, p. 47). As the

New Labour government increased its commitment to advancing education, assessment became an important tool for measuring performance and driving accountability in schools.

In 2001, Schools Achieving Success (Department for Education and Skills, 2001) was published. In this document, New Labour describes how assessment would play a key role in measuring success. The key purposes for assessment outlined in this policy document are system monitoring to evaluate national education in England and institution monitoring for individual schools. The document explicitly states that there was a 5% increase in the number of Secondary schools achieving five or more GCSEs since 1996 (page 8), demonstrating that educational standards were improving. This illustrates how it positions a system-monitoring (Newton, 2007) purpose for GCSE assessment that, in the authors' view, offers a reasonable use of the examinations taken in England at age 16. Assessment for accountability purposes permeates the document, and formative purposes are rarely mentioned.

The thread of using external assessments to evidence school improvement runs through the document without acknowledging the possible negative consequences for teaching and learning. In the previous section of this literature review (page 31) it was clear that a focus on high-stakes accountability in external assessments can result in shallow learning and the loss of learning opportunities as students are taught to the test. This is not acknowledged within the document, and any increase in school performance is taken as a proxy for better teaching and learning standards within that institution.

Chapters two and three of Schools Achieving Success (Department for Education and Skills, 2001) detail support to be provided to improve teaching and learning based on performance in summative assessments and for schools identified as not achieving the required standards. Assessment is not discussed as serving formative purposes beyond presenting to teachers whether their students are on track for certain levels of achievement. The document positions target setting and accountability as drivers of school improvement, although the mechanisms through which improvement would be realised are not specified. ‘We must also drive up standards at those schools which, while not failing, perform well below the national average’ (Department for Education and Skills, 2001, p. 6). This

statement seeks to convey the urgency of rapid improvement driven by measurable outcomes, with a focus on the 'national average' to increase competition amongst schools. The lack of learning from assessments and of making the processes formative is not acknowledged in this document, so it continues with the idea that increasing accountability will lead to better exam results, which, in turn, demonstrate better teaching and learning.

Whilst the purpose of this document, appears to argue for change based on assessment data, the thread of thought concerning the use of assessment for tracking performance continues throughout later New Labour policies. The policy processes instigated were positioned as drivers of reducing societal inequalities, with increases in educational performance, especially qualifications, seen as a cornerstone feature (Hall and Raffo, 2009). Despite positioning these policies as a new focus, many from the previous Conservative government were also continued, such as regular inspections and the reporting of school exam performance. The policy does not engage with or acknowledge research theory and the context of schools that are expected to increase their pupil's exam results as a sole indicator of 'success'. The challenge placed on schools to improve was driven through these accountability methods, which were seen as fundamental to New Labour's efforts in school improvement (Muijs and Chapman, 2009) and the assessment results, especially external exams, were a crucial tool.

In October 2004, the Tomlinson report released recommendations for change due to the burden of external assessments, which it summarised as 'too many young people leave education lacking basic and personal skills; our vocational provision is too fragmented; the burden of external assessment on learners, teachers and lecturers is too great; and our system is not providing the stretch and challenge needed, particularly for high attainers' (Tomlinson, 2004, p. 1). Tomlinson recommends a refocus from a high-pressure 14-19 curriculum with a high number of exams to a system where 'teacher-led assessment should be the predominant mode of assessment, though an element of external testing should remain' (p. 9). This recommendation and the justification follow the threads from the previous section, where the literature shows that high-stakes

summative assessments can negatively affect learning and lead to a loss of learning opportunities.

The tension between the Tomlinson report's recommendations and the New Labour government's focus on high-pressure accountability through exam results is clear. These tensions were evident immediately in contemporary accounts of political disagreements and contradictions about the nature of reforms to GCSEs and A-levels, including whether the names of these qualifications would remain and the extent to which teacher assessments would play a part (Shaw, 2004). The academic literature on the use of external assessment, reviewed in the previous section, tends to agree with Tomlinson's conclusion that this can lead to shallow learning and a loss of learning opportunities for students. The report does not acknowledge these tensions directly, however, nor does it acknowledge the policy context of the time, in which New Labour focused on system improvement through high-stakes accountability as a measure of school improvement.

New Labour's focus on using student exam performance became further entrenched with the release of the Making Good Progress consultation report released on 8th January (Department for Education and Skills, 2007). The report brings focus on the use of formal testing to identify students not making 'sufficient' progress and proposes a pilot to fine-tune measures of student performance between key stage assessments already in place. It proposed testing for progress, providing individual tutoring for students in English and mathematics, and setting targets for pupil progress and providing extra funding for schools that achieve them.

New Labour means for these proposed changes to consolidate their use of assessment processes for high support for schools they perceive as performing well, and high accountability for schools deemed to be underachieving. In a House of Commons debate in January, the Secretary of State confirmed 'any changes trialled in the pilot, or adopted subsequently, would not be allowed to compromise the accountability delivered by the framework of tests, targets and performance tables which have helped to drive up standards so sharply over the past decade' (Hansard HD Deb., 16 January 2007). There is also an expectation

that assessment processes can be used for formative purposes, as New Labour described it, Assessment for Learning.

The Assessment for Learning Strategy document was written by the Department for Children, Schools and Families and published in 2008 (Department for Children, Schools and Families, 2008). The document's foreword, written by Jim Knight, the Minister of State, outlines the Government's strategy to increase the use of formative assessment in the classroom. To achieve this, £150 million in support was released over three years to provide the professional development schools across England needed. The investment in CPD offers yet another example of the high challenge, high support approach pervasive in New Labour policy (Barber M, 2001; Brighouse and Waters, 2022).

Despite the promise of increased use of assessment for formative purposes, the document discusses a wide range of summative purposes. The uses of formative assessment to inform learning are discussed, following judgements about progress and attainment, traditionally, summative activities. For example, the following sentence 'every teacher is equipped to make well-founded judgements about pupils' attainment, understands the concepts and principles of progression, and knows how to use their assessment judgements to forward plan, particularly for pupils who are not fulfilling their potential' (Department for Children, Schools and Families, 2008, p. 5). As shown in the previous section (p. 43), much was known at this time about the difficulties of enacting effective formative assessment without understanding its spirit and the promotion of student autonomy. A tension exists between the intentions of assessment for learning national strategies, which aim to embed learning from research on assessment for formative purposes, and the policy environment, which promotes high accountability through summative assessment processes. There is no acknowledgement or attempt to resolve these tensions; however, schools were left to find their own paths through them in the policy.

Schools were challenged by the New Labour policy environment, facing demands to demonstrate their students' continued progress within the high-stakes accountability environment. Whilst assessments were used for formative purposes, they were often framed with the aim of improving performance in external exams, such as the Key Stage 3 SATS and GCSE examinations, on which

school reputation and funding depended. This accountability environment raises questions about the relationship between policy formulation and its assessment consequences, and about the extent to which policymakers engaged with contemporary research evidence and the wider policy landscape when designing these reforms. However, teachers were encouraged, through national strategy documents and other policy processes, to use assessment to define higher standards and to do everything possible to raise them.

One illustration of this encouragement is the Key Stage 3 National Strategy: Making the most of booster classes (Department for Education and Skills, 2002). The document advises teachers on best practice for 'pupils in Year 9 to maximise their achievements at the end of Key Stage 3' (Department for Education and Skills, 2002, p. 2). It was produced amid an intensifying government focus on raising standards through higher attainment in Key Stage 3 SATS examinations. The funding settlement reinforces this priority, allocating £8,500 per school for Year 7 to 9 booster classes in mathematics, English and science (DfES, 2002, p. 2). One reading might frame this funding as targeted support for pupils who need additional help. The distribution pattern, however, concentrates resources in the years immediately preceding the SATS, which suggests that raising measured attainment was the primary objective. Whetton (2009) describes the intense pressure on schools to meet pupil targets in these examinations, with published results generating public discourse about the proportion of pupils reaching the expected level. In October 2008, Ed Balls, then Secretary of State for Children, Schools and Families, announced the end of National Curriculum testing at Key Stage 3.

Following a series of controversies around the marking of the KS3 SATS exams (Curtis, 2008) they were removed without replacement, and an Expert Group on Assessment was commissioned. The remit of this report was to review proposed changes to the assessment and accountability systems in England (The Expert Group on Assessment, 2009) and, overall, to evaluate the main purposes of assessment and how these purposes could be achieved. Whilst the controversy over the KS3 SATS exams may have called into question other government examinations used for accountability, the report did not conclude that there were, at the time, issues with these examinations. This may not be surprising, as

it was widely reported that the review group was not provided with an adequate remit to recommend changes to the SATS examinations that were still being commissioned for primary-age students (Stewart, 2009). This was further evidenced when one of the authors, Sir Tim Brighouse, spoke out about the Sats, calling them 'too crude when it came to assessing primary schools' (Stewart, 2009, p. 6). In its final form, the document places a frequent reference to the use of assessment processes for accountability, stating within one of its conclusions, 'We believe, therefore, that the case for retaining some external testing at year 6 in each primary school is strong' (The Expert Group on Assessment, 2009, p. 5). Whether there was pressure on the group to support external testing as a useful measure endorsed by the government, or whether that was the overall consensus of the group, it is clear that, as with many policy documents, what is written in the document may not reflect the true consensus of all the authors.

New Labour's time in government provided substantial financial support for schools, especially where there were signs that this could lead to measurable benefits, such as increased performance in external testing. Through external government reviews and the use of specific data as evidence, the government of this time sought to create a policy environment in which financial investment in schools was socially accepted, rationalised, and subject to high accountability and pressure on schools. In the latter sense, and in their justifications for increased spending, New Labour offered more of a continuation of previous Conservative policy discourse (Sammons, 2008).

2.5.3 Assessment policy in England under the coalition and Conservative governments 2010 to 2023

Coalition-era education policy in England began with the 2010 general election. Held on Thursday, 6 May, the election returned a hung parliament, leading to a Conservative-Liberal Democrat coalition government by agreement on 20 May. Michael Gove was appointed Secretary of State for Education, having argued for greater school autonomy and parent choice across three years as Shadow Secretary of State for Children, Schools and Families (Shepherd, 2010). His proposals to grant school leaders greater autonomy over budgets and staffing were welcomed by union leaders such as John Dunford, then general secretary of

the Association of School and College Leaders. Shepherd reports that Dunford called for an end to bureaucratic restrictions.

The new Conservative Government presented a policy document titled The Importance of Teaching (Department for Education, 2010). The title of the document highlights, perhaps intentionally, where the priorities for change rested with the Conservative government's teaching. The key purposes for which the document puts assessment to work are accountability and monitoring the performance of the education system, nationally and internationally.

Conservative policy continued the tension between external assessment and learning, framing examinations primarily as instruments of accountability. In The Importance of Teaching, the Department for Education (2010) sets out testing and examinations as mechanisms for holding schools and students to account, while providing parents with information whose intended use remains unspecified. The white paper further presents high-stakes assessment as a necessary process, with outcomes described as serving multiple stakeholders and giving government 'a strong basis for accountability' (p. 11). This framing appears to position examination data as the principal means by which teachers and schools are held accountable for educational standards.

However, later in the Importance of Teaching document, it warns of gaming behaviour by schools, such as teaching to the test and choosing subjects or qualifications that are deemed easier to boost performance scores. These negative consequences are identified as belonging to ‘existing measures of performance’ (Department for Education, 2010, p. 13). These observations from the policy authors suggest that using data for accountability measures incentivises schools to operate in any way possible to achieve accountability targets, even when that means improving their measured performance through inauthentic means at the expense of a deeper, broader education. Schools that do not do this might be labelled as failing, and high-stakes accountability consequences may be deployed.

Having acknowledged these perverse incentives, the document proposes a series of structural adjustments increasing the volume of performance information provided to the public, shaping performance tables to encourage schools to offer

the English Baccalaureate, simplifying inspections, and establishing floor standards for all schools. However, whilst these reforms shift the focus of accountability processes, there is no explanation of how they will reduce the perceived need for schools to act in ways that fulfil the criteria for performance measurement rather than prioritise the needs of their individual students and communities.

One policy discourse from the coalition government is the need to develop the social mobility of disadvantaged students by elevating their performance in high-stakes examinations. Schools were compared based on pupils' performance on high-stakes tests, and these results were used to create Contextual Value Added (CVA) scores that account for differences in student background. The Importance of Teaching document argues that research (none cited) shows that CVA scores are unreliable predictors of school success. What success might mean exactly is not detailed, but it does put forward the idea that the use of CVA might lead to the expectation of different levels of achievement based on pupil backgrounds being 'wrong in principle' (Department for Education, 2010, p. 68). Instead, Michael Gove outlined that they no longer wished to continue with "the soft bigotry of low expectations", and instead follow PISA successful countries such as Hong Kong and Singapore, where school competition is high, would 'put a rocket under our world-class ambitions' (Gove, 2010, p. 25).

The coalition government then encouraged greater competition between schools through redesigned performance tables and floor standards, under which schools would face accountability consequences if judged as underperforming. The message for the profession was that they would be expected to raise the academic performance of all students, regardless of extraneous factors that might affect student progress. If schools were unable to accomplish this, they would face further accountability mechanisms, mediated through Ofsted inspections. Much of the evidence to convince the government of this came from international tests such as PISA and is regularly cited by the education secretary himself (Gove, 2010, p. 25). There is also a commitment to engage further in international assessment programs, including 'PISA, PRILS and TIMSS' (Department for Education, 2010, p. 47). The message of the need for competition is clearly fuelled by the coalition government's concern that

England’s education system should be competitive with other countries, and they saw this as well, as evidenced by international assessments.

The Bew report (Bew, 2011) was instructed by Michael Gove to remain within the government’s accepted position of following the OECD's recommendations. On page four, the document makes this clear, stating, ‘There are two broad positions to which the Secretary of State has asked us to adhere throughout. Firstly, the Government is mindful that the Organisation for Economic Co-operation and Development (OECD) concludes that external accountability is a key driver of educational improvement, particularly for the least advantaged. It, therefore, views a system of objectively measuring pupil progress and holding schools to account as vital’ (Bew, 2011, p. 4). Therefore, although the review is titled ‘The Review into Key Stage 2 Testing, Assessment and Accountability’, it was restricted in its critique of the use of assessment for accountability purposes for fear of breaking this imposed remit. The document, therefore, serves to continue the discourse Michael Gove often uses, arguing that international assessments provide evidence for the increased accountability schools need.

The Bew Report devotes a chapter to accountability, drawing on selected OECD evidence to support 'high stakes' accountability as a lever for improving teaching and learning (Bew, 2011, pp. 21–24). It continues by claiming that 'the most effective systems in the world' use external school-level accountability, and that this approach raises student attainment and progress. No account is offered, however, of what better teaching or learning might look like, despite this presumably being the mechanism through which such gains would be realised. The report briefly acknowledges washback effects on curriculum and teaching but sets these aside as something that 'need not be the case' (Bew, 2011, p. 22). An Ofsted survey of twenty primary schools is cited as evidence that such consequences can be avoided.

Other purposes for assessment besides monitoring and accountability are mentioned within the Bew report, such as formative purposes. Despite the academic evidence for the improvements that can be derived from formative assessment practice, shown in this work’s literature review, it is mostly overlooke. It is hoped that schools will be able to devote ‘as much time, energy and space as possible to use a continual repertoire of assessment techniques as

they see fit' (Bew, 2011, p. 9). However, the academic literature at the time, reviewed in the previous section, evidences that the pressures from accountability based on external exams would mean these 'techniques' are sidelined for exam preparation and practice. Furthermore, the report conflates the formative purposes of assessment by asserting that they involve building spreadsheets of data and can be used for summative purposes. It suggests 'schools having formative assessment arrangements in place, for example, effective tracking of pupils, which will support teacher assessment at the end of Key Stage 2' (p. 10).

The Bew report was part of a large number of changes to inspection, curriculum and assessment at the start of Michael Gove's tenure as Education Secretary in England. The difficulties of using assessment to serve various formative and summative purposes without the undesired washback effects discussed so far were not resolved. This is evidenced by the Department for Education's revisit in March 2014 to the issues of assessment and accountability in Primary Schools. This was later followed by a change in assessment in Primary and lower Secondary education, whereby the tracking of students by assessment level was abandoned.

In 2014, there was a change in how student progress and achievement were measured in Primary and early Secondary phases, whereby national levels were removed, and schools were expected to measure using their own systems. Nick Gibb, the Minister of State, concluded, 'In short, levels were just too vague and imprecise. They were misleading as to what pupils knew and could do. The use of levels was pushing pupils on to new material - in the name of pace - when they had not adequately understood vital content and had serious gaps in their knowledge' (Gibb, 2015a, para. 15). Following the removal of the national curriculum levels, schools were required to construct their assessment processes to measure student progress in other ways. This process meant schools designing entirely new ways of monitoring students' progress, a process that many found confusing and were nervous about collecting robust evidence for accountability (Staufenberg, 2018). It also, however, produced innovative practices, with many schools designing their own assessment monitoring systems where commercial options were not available (Poet *et al.*, 2018). The removal of national

assessment levels is another example of Gove's declared ideology of enhancing school autonomy as a means of improvement.

Prior to 2010, the academisation approach was selective. Encouraging or forcing mainly struggling schools to become academies to drive improvement. This was based on the idea that converting to academy status, with more autonomy from local authority control, would enable these schools to innovate and improve educational outcomes (Chapman, 2015). From 2010, the process of academisation accelerated significantly with the introduction of the Academies Act 2010, which enabled all state schools, not just those deemed underperforming, to apply for academy status (Chapman, 2015). In November 2014, 4,269 schools had converted to academy status, providing evidence that the structural changes to the English education landscape under Gove were making progress. This may indicate greater autonomy for these schools, which are no longer subject to the local education authority's control. Along with the removal of national assessment levels, it was argued by Nick Gibb that the programme of school reform was 'beginning to bear fruit' (Gibb, 2015a, para. 2). However, all schools were still inspected by Ofsted and the national curriculum was still published by the Department for Education, making true autonomy somewhat illusory. In addition, many school Headteachers now work under the direction of the executive headteachers leading most multi-academy trusts, trading local relationships with education authorities for potentially more remote relationships with national academy trusts.

Whilst it has been argued that education was not a prominent feature of the 2010 election (Bailey and Ball, 2016), this section has argued that the agenda for creating an impression of school autonomy was clearly agreed upon and pursued by Nick Gibb and Michael Gove during the coalition. Following this, Michael Gove's influence over education was not direct as a Government Minister, but the discourse of autonomy and accountability continued.

In May 2015, the Conservative government won a general election in the UK, securing 36% of the seats in parliament. This allowed the Tory majority government to enact policy without needing the support of its coalition partners, the Liberal Democrats. The former Education Secretary, Gove, had left office in July 2014 and was replaced by Nicky Morgan. Although there had been a

succession of nine Secretaries of State for Education since Gove, by 2023 there had been greater continuity in the role of Under-Secretary of State for School Standards, with Nick Gibb serving twice in the position for a combined total of 7 years since 2015.

This period followed five years of the Conservative austerity programme, during which overall education budgets were protected, but funding was unevenly distributed in favour of the new academy and free-school programmes (Granoulhac, 2017). The competition and accountability agenda initiated by earlier policies exacerbated the high-stakes effects on schools, in which downward trends in exam performance would result in demotion in league tables and a drop in the school's student intake and funding. The competition for students and the funding they would bring intensified, with successful schools achieving high performance in external exams at any cost to secure better funding. These pressures forced secondary schools to ensure that they could meet Progress 8 performance measures, which marked a return to value-added measures of school progress. The P8 measures of school success have been criticised for their weaknesses as a measure of school cohort success (Prior *et al.*, 2021) and marked a new era of change in school accountability, causing confusion for parents and school professionals.

A research report released in 2017, sponsored by Pearson (Millard, Small and Menzies, 2017) helps to illustrate the confusion felt by Primary and Secondary teachers across England. As previously described (p. 53), the removal of national assessment level descriptors and the introduction of secondary school P8 measures were designed to avert the problems associated with accountability's washback effects on teaching and learning. The report, however, found that due to uncertainties around how data might be used to hold schools accountable, schools focused on producing large quantities of assessment data to 'second-guess how to improve their performance under such measures' (p. 45). The confusion and need for teachers to create data on their students is supported by evidence from Pratt and Alderton (2019) who provide qualitative accounts of how teachers sought assurances of their impact on student learning by producing assessment data to demonstrate progress.

The changes to assessment and accountability during this period led to more autonomy for school leaders in designing systems to measure student progress. However, rather than allowing teachers to develop their own assessment systems, schools still had to demonstrate progress to Ofsted, and exam league tables still influenced the curriculum and teaching. These influences and the workload imposed on teachers to produce data for these processes were cited as a continued source of de-professionalisation (Dorrell, 2018). Teachers' understanding of assessment had not been developed through this time, and teaching unions noted that 'many teachers understandably equate it either with high-stakes tests … or with triple marking and pressure to grade every pupil's [sic]' (Bousted, 2017). With a teaching profession that was deskilled in assessment practice and accustomed to working towards higher grades by almost any means, England was ill-prepared for the challenges of assessing students for school-leaving certificates during the COVID-19 pandemic.

On the 16th of March 2020, Boris Johnson, the Prime Minister of the United Kingdom, announced the closure of all schools due to the COVID-19 pandemic. Two days after this, it was announced that examinations in England would not take place, and a national lockdown then followed. Following these announcements, schools were required to adjust to new operational norms and sought guidance on how their students who would have been taking external exams would be assessed for their qualifications. Ozga *et al.* (2023) have argued that during this time, the English government prioritised avoiding grade inflation. This would follow the Conservative government's ideological focus on accountability driven through external exam grades, as previously described.

In 2020, schools were instructed to submit Centre Assessed Grades (CAGs) to the Office of the Qualifications and Examinations Regulation (Ofqual), the government's exams regulator. The CAGs were then processed by a computer algorithm designed to moderate the grades using the school's prior attainment data, to ensure the grades awarded that year were not dramatically higher than those in previous years. The results from the algorithm were issued to students on the 13th of August, generating public outrage as nearly 36% of grades had been lowered from the submitted CAG grade from schools (Kelly, 2021). Further analysis of which student groups were most likely to have had their grades

reduced found that the CAG process and the algorithm both contributed to increasing disadvantage for students from state schools and those from Black and minority backgrounds (Bhopal and Myers, 2023).

The algorithm details were not made public until after the publication of the results in August 2020, and Ofqual, the independent government department, were blamed for partaking in political secrecy around how the grades had been formulated (Kelly, 2021). The secrecy surrounding the algorithm and the disappointment of students whose grades were lowered eroded public trust. Parents pressured constituency MPs, and media reporting focused on evidence that high-achieving pupils in lower-performing schools were disproportionately affected (Ozga *et al.*, 2023). Following this growth in hostility from parents, the media, and fellow politicians, the minister for education announced that a 'triple-lock' appeal system would be launched. This was an attempt to quell protests from various groups and ease concerns about students being disadvantaged.

The appeal system, however, was quite different from what had been promised and did not, in fact, allow students to receive the highest grade from the algorithm grade (CAG) or from mock exam results. This was initially defended by criticising the Scottish government's abandonment of similar plans and its return to CAG grades, with fears that such a procedure would cause grade inflation. He said at the time, 'the process was necessary to prevent grade inflation which would render the results worthless…' (Guardian, 2020). It was announced on the 17th of August that CAGs would be used to award students their final grade, resulting in grade inflation for top grades estimated to be 12.5% (Ozga *et al.*, 2023) and further confusion for schools, parents and more importantly, students.

The confusion and crises have impacted schools in many ways as they continue to move on from the COVID-19 pandemic. Exam grading concessions for students continued, such as alterations to exam expectations and a relaxation of the grade boundaries. Schools continued to face disruption due to staff and student absence, which further contributed to education disruption beyond the pandemic. Despite this, school inspections returned, and the accountability expectations remain.

2.5.4 Assessment policy in Scotland following devolution 1998 to 2014

Following devolution in 1998, Scotland established the Scottish Parliament in 1999, and Scotland continued its unique nature of schooling, with significant implications for schools and wider education (Walford, 2005). This change in the Scottish political landscape made it possible, although not inevitable, that the Scottish National Party would be able to secure further support and go on to become the first-ever nationalist party to gain power in 2007 (Devine, 2016). The SNP's education policy resisted England's model of academies and free schools and continued its distinctive nature with policies aimed towards attempting to resolve issues for disadvantaged students, free school meals, reduced class sizes and increased widening access to university education (Gillies, 2018).

Scotland has always been proudly unique in its education system. It has been regarded as 'an expression of social unity tied up with Scottish identity; as a reflection of democracy and communal solidarity (schools should serve their neighbourhoods); and as a demonstration that opportunities to succeed should be available to all' (Bryce and Humes, 2018c, p. 46). Ozga (2000) highlights Scotland's resistance to more neo-liberal reforms enacted south of the border, such as school choice and competition. However, this resistance failed to completely reject some aspects of education policy, for example, describing the publication of performance tables and the completion of, albeit moderated, national assessments (Mooney and Poole, 2004).

Scotland is also a comparatively much smaller nation in terms of population compared to England. Despite this, a sense of community in policy making has been questioned, with Humes (1986) asserting that there was a particular group of individuals asserting a normative pressure to policy processes, something later referred to by Gillies (2001, cited in Gillies, 2018a, p. 87) as 'enlightened elitism'. As a small country, it might be argued that this could, to some extent, be inevitable. However, these groups may be identified by their repetitive appearances in this exploration of Scottish education policy.

A ‘National Debate’ was initiated in 2002 to consult on how Scotland might shape this continued independent and unique education system, with the response published in 2003. The review of the curriculum for 3-18-year-olds was announced, and in 2004, Ambitious Excellence Schools (Scottish Executive, 2004) was published to outline the vision for this future and how high standards might be ensured.

In Ambitious Excellent Schools (AES), the document outlines ambitions to ‘benchmark Scotland against international standards, particularly through the work of OECD’ (Executive, 2004, p. 20). This foreshadowed the continued relationship between the OECD and Scottish education authorities in a way that Lingard and Sellar (2014, p. 16) described as ‘Scotland reaching out to the OECD to help constitute a ‘Scottish nation’. The AES document also describes the intention to accept the recommendations to enact a new curriculum for Scottish schools called a Curriculum for Excellence and to incorporate the existing assessment for learning programmes to ‘ensure that assessment supports learning’ (Scottish Executive, 2004, p. 15).

In 2010, the new Scottish Curriculum for Excellence was initially used in schools 2010 after a period of consultation since its initial planning by the Scottish Executive (Priestley and Minty, 2013). The curriculum reform was designed to put teachers in a position of responsibility as curriculum developers, not being prescribed in law as with England’s National Curriculum, but is based on underpinning progressive values and principles (Hayward, Macbride and Spencer, 2016). The model for curriculum reform in CfE was to ‘combine top-down government prescription with bottom-up school-based curriculum development by teaching professionals’ (Priestley, 2010, p. 23). Before addressing potential contradictions, the main components of the curriculum for excellence are discussed.

Unlike England’s National Curriculum, originally designed by Kenneth Baker, which has a focus on attainment targets (Brighouse and Waters, 2022), and the previous 5-14 guidelines in Scotland, CfE is framed around four learner capacities: successful learners, confident individuals, responsible citizens and effective contributors. These descriptions of achievement are broad (Hayward, 2007) and intended to develop from ages three to eighteen (Scottish Executive,

2006, p. 12). In 2007 and 2008, draft experiences and outcomes were published and then reviewed as part of a national feedback process, analysed by the University of Glasgow (Baumfield *et al.*, 2010). This review raised potential tensions between ideas of learning within CfE and accountability processes. Participants from primary and secondary schools talked of 'double vision' and 'different worlds' in describing the multiple and competing demands made of them' (p. 66). Teachers were found to be highly confused with the expectations of CfE, with some reporting that they felt it 'required the replacement of summative assessment with formative assessment' (Priestley and Minty, 2013, p. 49). These tensions are further discussed with other aspects of contradiction in the enactment process later in this text, following a discussion of the other challenges of enacting CfE.

One difficulty in the enactment of CfE was the capacity of teachers to adopt the positions as 'agents of change and professional developers of the curriculum' (Priestley and Minty, 2013, p. 39). Priestley (2010) describes the 'paucity of curriculum theory' (p. 24) for teachers and policy influencers, which was problematic in an environment where curriculum innovation was desired. Priestley and Minty (2013) called for a greater resource provision for teacher learning and discussion of philosophies around CfE to provide a shared understanding of its aims and solve challenges of misunderstandings from some practitioners. Advocates for the CfE reforms emphasised the critical need to engage more deeply with Scotland's contextual factors and mechanisms driving educational change (Priestley, 2010). In addition, it was hoped that such engagement might acknowledge and attempt to resolve inconsistencies in some of the guidance.

A key tension in CfE's enactment guidance lies in reconciling the dual priorities of maintaining national standards for student achievement and fostering teacher autonomy in summative assessment processes (Hayward and Hutchinson, 2013). This duality was further complicated by CfE's stated aim to provide schools and educators with the flexibility to tailor curricula to 'local needs' (Priestley, 2010, p. 23), creating a delicate balance between centralised expectations and localised decision-making. In walking this tightrope through the enactment process, Priestley and Minty (2013, p. 50) found that many teachers in both

primary and secondary schools adopted a ‘wait and see’ approach to developing their curricula, waiting for the publication of new National Qualifications. Later work found that these pressures had not changed, and teachers’ capacity for curriculum change and agency was restricted by school cultures of performativity and accountability related to student attainment (Priestley *et al.*, 2015). The 2016 OECD review of CfE noted that senior-level qualifications were distracting from curriculum reform by dominating ‘public, policy and professional discussion in the past 2-3 years’ (OECD, 2015, p. 99). These tensions persisted and are discussed further in the next section, despite further guidance published by the Scottish Government aimed at resolving tensions around curriculum enactment and the use of assessment to judge and inform learning.

Building the Curriculum 5 a framework for Assessment, Scottish Government (2010) provided guidance for assessment within the enactment of CfE. It provided guidance on how assessment processes might be deployed successfully for multiple purposes. Sometimes confusion and tensions emerged from the document, especially with contradictions between the assessment purposes. One such example is the way the document emphasises the use of assessment processes to support learning whilst also giving parents assurance that students are meeting attainment standards. This can be seen straight away from the forward, where ‘assessment is used to support individual learning and to provide reliable information to learners, parents, employers and further and higher education about the standards that have been achieved’ (p. 1). As described earlier in this literature review, the evidence from research into the impact of assessment is that when it is used for high-stakes monitoring of standards, its use for formative purposes can be minimised (Harlen, 2005a) and can have other unintended negative impacts, such as narrowing the curriculum (Boyle and Bragg, 2006; Polesel, Rice and Dulfer, 2014). For CfE enactment, the focus on standards may put the focus on achievement outcomes rather than the other foci at the heart of CfE.

In building the curriculum 5, there are examples where it reflects varying perspectives or emphases. Although overall, it seems to advocate for a balanced, learner-centric approach to assessment, subtle inconsistencies can be seen at play. It advocates the involvement of external agencies like the Scottish

Qualifications Authority (SQA) and HM Inspectorate of Education for quality assurance and moderation processes. Rigorous and robust procedures will also be important as part of the broad general education and in particular at points of transition in order to ensure the reliability of information shared about progress and achievements' (Scottish Government, 2010, p. 35). These might suggest a desire for external oversight and standardisation of assessment processes rather than them being in the hands of teachers and students. These contradictions may have arisen as a process of compromise between multiple authors or as an attempt to address multiple stakeholders in its intended audience. However, overall, it outlines the intention of the Scottish government at the time that assessment would be aligned with learning and complement the wider philosophies of a Curriculum for Excellence.

The Assessment Is for Learning policies provided teachers with ideas and resources for the enactment of formative assessment practice, but did not prescribe how this practice should work within schools. While this was designed to increase professional competence and agency in teacher practice (Hutchinson and Hayward, 2005), challenges were identified in combining the lessons learned from this programme before moving on to Curriculum for Excellence developments (Hutchinson and Young, 2011). Hayward and Hutchinson (2013) later reviewed the alignment between classroom assessment and the enactment of Curriculum for Excellence practices. They found tensions in the uses of assessment for learning and assessment for accountability had not been resolved through schools enacting Curriculum for Excellence frameworks, especially in the upper levels of secondary. In this way, it began to emerge that external exams might have more influence over teachers' practice and the hope that teachers would use agency to assess to support the learning outlined in their curriculum, as had been envisaged.

2.5.5 Assessment policy in Scotland from the OECD review to 2023

The tensions that emerged between external exams and the ambitions of the Curriculum for Excellence became evident from the conclusions of the Assessment at Transition report (Hayward *et al.*, 2012). The report outlined tensions between the use of assessment for formative purposes and its use to

inform accountability measures. How these tensions played out during the transition for students from primary to secondary phases of education seemed particularly problematic and an example of how assessment purposes had become reductive, with too little trust or development of teacher assessment (Hayward, 2015).

Accountability pressures also extended into the senior phase of Curriculum for Excellence, where assessment guidance lagged behind curriculum announcements. The Building the Curriculum documents, released between 2006 and 2009, set out the senior phase reforms, but the accompanying assessment guidance did not appear until 2011. Details of the assessment and accountability arrangements for the National 2 to National 5 qualifications followed later still. Teachers expressed growing anxiety during this period, seeking clarification of assessment requirements before committing fully to curriculum planning. Buie (2012) reports that teachers requested specimen examination materials and that teacher unions conveyed members' concerns, reflecting the weight the new qualifications carried for accountability purposes.

Further pressure on the importance of the Curriculum for Excellence assessment guidance was placed by critiques of assessment practices at the time. The Assessment at Transition report (Hayward *et al.*, 2012) noted that not enough time was spent on moderation to enable teachers to better agree on assessment standards and the validity of assessment inferences. The report outlined four key areas for improvement in how assessments and the information produced from them might be used:

A better synergy between the curriculum and assessment, along with the validity of what assessment inferences lead us to consider, has been learned.

- Progression in learning would need to be given greater consideration of outcomes beyond fragmented concepts of learning.
- Greater trust and agency for teachers in their role and judgements as assessors.
- A collective understanding of intelligent accountability to empower teachers beyond the confines of exam-related considerations.

These recommendations and the need to further the development of teacher assessment practice were further highlighted in public appeals for the use of assessment information for more formative purposes to improve learning (Donaldson, 2013). The exact nature of the assessment information and how it might be used, however, continued to be debated, with further questioning of how information from assessment processes might be used to improve student learning.

Scotland's education policy has, in recent years, increasingly shifted towards global trends (Gillies, 2018). Gillies argues that, while Scotland previously resisted market pressures and high-stakes testing, these features are now beginning to appear, narrowing what had been a longstanding policy distinction from England. This shift became visible in 2015, when the Scottish Government commissioned the OECD to review the direction and enactment of Curriculum for Excellence. The review framed the principal purposes of assessment as monitoring and accountability, benchmarking Scotland against frequently cited national systems. It noted relative strength in PISA but cautioned that 'on the most recent 2012 PISA surveys, Scotland was similar to the international average after having been one of the leading countries in maths achievement a decade before' (OECD, 2015, p. 10). Key recommendations addressed national system monitoring, observing that existing sampling methods failed to provide 'sufficient evidence for stakeholders to use in their own evaluative activities' (OECD, 2015, p. 23). The report further argued that 'while learner outcomes should not be the only focus of a standards or appraisal system, stronger reference to learners' progress will create improvement' (OECD, 2015, p. 23). In these statements the OECD attempted to balance the formative emphasis Scotland had pursued through Curriculum for Excellence with a renewed accountability function, tying that function to international standing in assessments such as PISA. This framing positioned assessment for accountability as an immediate priority for Scottish policymakers.

Scotland has historically resisted accountability mechanisms such as school league tables, while England progressively adopted them across successive reforms (Johnson, 2016). This divergence shaped the OECD's 2015 review of Curriculum for Excellence, which examined how Scotland balanced formative

assessment with the demand for system-level evidence. The review praised the Scottish approach, noting that 'trust towards teachers and the formative emphasis, with diverse methods to collect information, are very positive aspects of the system' (OECD, 2015, p. 11). The same report, however, cautioned that PISA mathematics rankings had slipped and that existing arrangements did not generate evidence sufficiently robust for policymakers, local authorities or individual schools (OECD, 2015, p. 18). It concluded that Scotland would need to 'maintain the dual focus - on the formative function while improving evidence on learner outcomes and progression' (OECD, 2015, p. 11). The literature considered earlier in this chapter (see page 19), however, suggests that sustaining this dual focus is difficult in practice. Where accountability regimes treat assessment primarily as a tool for generating performance data, the formative purpose tends to recede, and assessment risks becoming a proxy for improvement rather than a means of supporting learning.

The Scottish Attainment Challenge was initiated in 2015 by the First Minister to address educational inequities by focusing on enhancing literacy, numeracy and health and well-being among students from economically disadvantaged backgrounds (Education Scotland, 2023). Ainscow, Chapman and Hadfield (2019) describe the Scottish government's motives for using 'evidence-based interventions' (p. 100) to steer the investment of the £100 million increase in spending over four years with an aim to approach the allocation of resources for specific, measurable educational improvement. This drive to improve key measures and demonstrate improvements in literacy and numeracy can be linked to patterns of curriculum narrowing and a greater focus on accountability related to student attainment data (Shapira and Priestley, 2018).

The politically driven direction to use assessment data to drive improvement was further catalysed by the release of the PISA survey results in 2016. The Scottish National Party faced pressure from a number of fronts to increase the production of assessment data and to encourage its use in school judgement. Nicola Sturgeon, who had been First Minister for a year, had been under increasing fire from political opponents to increase accountability for schools and competition between them (Warrell, 2016; McEnaney, 2021). A conservative spokesman was reported to have described the results as 'a damning indictment

of a decade of failure under the SNP' (Bryce, 2018, p. 600). The political pressures culminated in a speech by Nicola Sturgeon, which launched a change in direction and a pledge to eliminate the attainment gap between disadvantaged students and their more affluent peers in Scotland. Often referred to as the 'judge me on my record' speech, despite never actually saying these words, Nicola Sturgeon outlined proposals to extract more assessment data from schools and attempt to use that data to inform parents and make improvements in education (Henry, 2015). This speech outlined a new direction, with an increase in the use of assessments to survey national education quality and to raise the stakes for schools by reporting their performance in these assessments externally.

The new direction towards increasing the production of assessment data to monitor education quality continued with details of the contents of the National Improvement Framework (Scottish Government, 2016b). The document focuses on narrowing the poverty-related gap in attainment that had presented such an opportunity for criticism of the SNP government. It had a subtitle of 'achieving excellence and equity', which highlighted the focus on educational excellence as defined through academic achievement and equity defined by addressing the attainment gap for disadvantaged students. The use of external assessments was promoted to monitor student progress through local authorities collecting data in a system known as Insight. This tool would allow schools and local authorities to compare themselves and see how their cohort's performance compared to others with similar socio-economic factors. Although the perception of the OECD results and recommendations from the SNP government was that more external assessment measures were needed to drive improvement, it wasn't universally believed that this would achieve the desired outcomes. The use of assessment to inform accountability and encourage competition among schools was cautioned at the time as a possible route towards unintended consequences (Hayward, Macbride and Spencer, 2016).

Scotland's 2017 Governance Review (Scottish Government, 2017)signalled a shift towards a more networked approach to educational reform, even as individual school results continued to be published. The Review proposed six Regional Improvement Collaboratives (RICs), designed to sit between national government

and local authorities and to support educational improvement. By fostering closer ties and enabling the sharing of effective practice among schools, local authorities and other stakeholders, RICs were intended to promote cohesion, equity and excellence across Scottish education (Chapman, 2019).

The RICs have been characterised as a constructive development in Scottish education, particularly for the professional learning opportunities, exchange of practice and collaborative projects they support (Forde and Torrance, 2021). Forde and Torrance also question elements of the model, calling for 'evidence-based approaches and evaluation' to gauge the impact of the collaborative networks and the extent to which they 'move the system forward' (2021, p. 35). Such an emphasis on data-led improvement extended the evaluative apparatus surrounding schools, drawing on what Chapman describes as a 'myriad of data sources' (2019, p. 562) to assess school improvement externally.

Schools and local authorities continued to use external data on student achievement and to compare themselves nationally against benchmarks. Future Improvement frameworks continued to focus on the use of assessment data to evidence the quality of education and seek to evidence improvements in standards. Much of this assessment data relied on external assessments, which fostered a dependence on exams as the most reliable measure of educational quality. This created a vulnerable position for the education sector in Scotland, with a high reliance on external examinations to infer the quality of education.

As in England, the pandemic in Scotland meant schools were closed, except for limited provision for key workers and children with special educational needs. As in England, teaching shifted to online learning, and teachers had to adjust their lessons to alternative formats using online platforms. These school restrictions meant it was not possible for external examinations to proceed in the usual format of students sitting paper examinations whilst supervised in school. Contingency plans for student assessment were particularly important for producing examination grades that determine access to work and further study for many students in upper secondary school.

Initial plans for these assessments were introduced, where teachers and education centres provided estimated grades and rank orders for students. The

school-estimated grades were then moderated by the Scottish Qualifications Authority (SQA), affecting 27% of those submitted. This process was widely criticised at the time and subsequently for both the rigour of assessment in education centres and the transparency of SQA moderation (Priestley *et al.*, 2020). The criticisms escalated when concerns were raised about the 'equity implications of an over-reliance on a statistical approach' (p. 10), which used school historical data that might disproportionately affect disadvantaged students (Peebles, 2020). The impact of these failings in the traditional examination system marked a perceived shift in public consciousness about what is possible in end-of-school examinations.

Future changes for public examinations in response to the COVID-19 pandemic have been documented for internationally comparable countries such as Ireland, which opens the dialogue for what is possible in Scotland (Hayward and O'Leary, 2022). The rapid review of national qualifications experience from the University of Stirling (Priestley *et al.*, 2020) makes many suggestions for future assessment possibilities in Scotland:

- Inclusion of a broader evidence base for awarding final grades in external qualifications.
- Including methods that might meet the need for assessment processes to consider the diverse contexts and backgrounds of students to ensure equitable conditions for all.
- Explore the use of digital materials and assessment systems to support the production, assessment, and moderation of evidence to enhance the efficiency and accessibility of assessments.
- Enhancing the role of multiple stakeholders to co-construct future assessment models.

These suggested changes have been further repeated by some in Scotland, building a movement towards an assessment system that builds trust in assessment systems by all stakeholders (Hayward *et al.*, 2023), equitable and supportive of a greater diversity of skills and knowledge that are more closely aligned with the Curriculum for Excellence's broader educational goals (Stobart, 2021).

2.6 Chapter summary

This literature review has explored research on educational assessment across four interconnected domains: the purposes and processes of assessment, teacher practice and identity, accountability and educational change, and policy development in England and Scotland. Through this multi-layered analysis, several recurring themes warrant investigation in the empirical analysis of the complex relationships.

First, whilst there has been substantial theoretical development in understanding assessment purposes, particularly formative assessment, the translation of this knowledge into sustainable classroom practice appears uneven, a question this study explores. The 'spirit' of formative assessment remains elusive in contexts dominated by high-stakes accountability.

Second, the relationship between national assessment policy and classroom practice is mediated by multiple actors operating at different levels of the education system, each bringing their own interests, beliefs and constraints. Teachers' professional identities and assessment literacy may play significant roles in shaping how policy is enacted, resisted or adapted.

Third, both England and Scotland have increased their reliance on assessment for accountability purposes over the past 25 years, despite substantial evidence of the negative consequences of high-stakes testing. International benchmarking through PISA has intensified pressures for measurable improvement, creating policy environments where short-term gains in test scores can be privileged over longer-term educational goals.

Fourth, the structures and mechanisms of educational change, including inspection regimes, institutional forms, and collaboration and competition, shape the possibilities for assessment practice in ways deeply embedded in each nation's political and cultural contexts.

Finally, the COVID-19 pandemic has created an unprecedented moment of disruption and reflection, opening the possibility of reconsidering assessment

practices whilst also revealing the powerful institutional and political forces that resist fundamental change.

The next chapter will outline the methodological approach for exploring these themes through interviews with two distinct yet interconnected groups of participants. Policy influencers, those individuals as members of organisations operating at supra, macro and meso levels who shape assessment policy discourse and direction, will provide insights into the intentions, constraints and political contexts that inform policy development. School-based teachers and leaders, who operate at the micro and nano levels where policy meets practice, will illuminate how assessment policies are understood, mediated, and enacted within the distinctive contexts of their schools and classrooms.

By bringing these perspectives into this analysis, this study will explore the relationships between policy intentions and practical realities, between official discourse and professional knowledge, and between the aspirations for assessment reform and the enduring structures that shape assessment practice. The methodology must therefore be sufficiently flexible to capture the complexity and nuance identified in this literature review, whilst maintaining the analytical rigour necessary to generate insights that can inform both scholarly understanding and professional practice in this critical domain of educational policy and practice.

3 Methodology

3.1 Introduction

This thesis explores how research evidence, policy development and classroom practice interact within systems of England and Scotland. It examines where domains align and where they diverge. Drawing on the perspectives of teachers, school leaders and policy influencers, the research explores how these dynamics operate within and across the two countries. Engaging in in-depth interviews with these stakeholders, it studies the mechanisms by which individuals in the field navigate and adapt to an evolving policy landscape. By analysing their accounts, this study examines the tensions between policy intentions and professional understandings within countries. This understanding was developed through three key questions:

1. Within the context of policy reform (1998-2023) in Scotland and England, what are the intended purposes of assessment?
2. To what extent are assessment policies in Scotland and England aligned with assessment research?
3. How do teachers, school leaders and policy influencers perceive educational assessment policy in the Scottish and English education systems?

These research questions direct the methodological design and guide the selection of participants and data collection instruments.

This two-tiered approach enables systematic comparison of policy trajectories across both jurisdictions, tracing how national assessment policies are interpreted, mediated, and enacted at different levels of the education system. By triangulating policy influencers' articulated intentions with classroom practitioners' lived experiences, the research critically examines the extent to which assessment policy achieves its intended outcomes and identifies points of divergence and convergence in the two jurisdictions' enactment processes.

3.2 Ontology and epistemology

This section explores the theoretical framework that underpins this work and how this drives the research questions. It then discusses how knowledge is sought in this work to develop a better understanding of research, policy, and teachers' perceptions of practice. The approach to this study required questioning a range of paradigms and assumptions about the nature of the research questions and how they might be explored.

In addition to my previous training as a science teacher, my axiology and position as a teacher educator in England may influence how they interpret education systems in Scotland and England. Axiology can be defined as the values and beliefs held by a researcher (Cohen, Manion and Morrison, 2018c) and may be formed by my prior experiences and training in England, which inform the values that they hold. As a former teacher and beneficiary of the English education system, it is arguably easier for me to comprehend and accept its educational conventions than the less familiar Scottish system. Therefore, particular care is taken when observing and analysing the workings of the educational system in Scotland, and any potential biases or orientations towards this system are considered reflexively. Given the need to understand what is meant by teachers operating within Scotland's distinct policy environments, I spent a week observing lessons in Dundee schools to gain an appreciation of the contextual nuances and unique terminology in Scottish education. This approach, proposed by Coe *et al.* (2017, p. 185), is useful to 'prepare for and conduct each interview with that [culture] in mind. When undertaking observations and discussing findings within these contexts, it is necessary to remain cognisant of my position as an outsider in some contexts and as a more familiar entity in others.

As a researcher within the social sciences, I have come to recognise the constraints inherent in strictly positivist ontological positions for understanding social phenomena. This research necessitates an understanding of how individuals construct and experience social reality, with an appreciation that multiple, contextually-situated interpretations may coexist rather than a single, objective truth (Walshaw, 2007). It is worth describing the possible offerings of positivism and quasi-experimental approaches before explaining the decision to

reconsider the usefulness of these theoretical approaches in the application of this work.

As an undergraduate student, Training in the scientific method meant that I was taught what is sometimes referred to as ‘the scientific method’ (Cohen, Manion and Morrison, 2018, p. 10). A hypothesis is created to explain phenomena that may be observed and then later explained. The extent to which the hypothesis is accepted in predicting what will be observed determines whether we can accept the predictions proposed as fact(Cohen, Manion and Morrison, 2018b). The natural science paradigm is associated with a French philosopher, Auguste Comte, who forwarded the doctrine of positivism, where knowledge is advanced by human observations and empirical methods (Cohen, Manion and Morrison, 2018c). This work made the assumption that the causality of events and the extent to which these causes can be identified through experimental testing. This assumption is known as determinism and, along with empiricism and parsimony, forms the background of the author’s epistemological understanding from prior scientific training (Cohen, Manion and Morrison, 2018b).

Initially, I approached the social science of education with attempts to explain causal links between classroom behaviours in a deterministic fashion. It was expected that these behaviours could be understood through experimentation, with a group serving as a control and another as a treatment group, a method commonly referred to as ‘empiricism’ (Cohen, Manion and Morrison, 2018b, p. 11). Finally, I assumed that it would be possible to use data obtained from these experiments to explain the links between observed behaviours with statistical data referred to as ‘parsimony’ (Cohen, Manion and Morrison, 2018b, p. 11). However, the methodological approaches of this paradigm did not align with my classroom experience and the social nature of how processes work within schools.

Positivism, as a theoretical framework, guides me to separate method from my own biases and from the object of study. When studying chemicals or other species, this framework and the approaches derived from it are useful for identifying patterns and fundamental laws that explain phenomena observed. With social behaviours and constructs, it is more difficult for me to position myself outside of what is being studied. Randomised controlled trials are used in

some areas of education research to understand social phenomena. However, this work aligns with Gale (2018), who argues that such trials conflate the physical and social world, failing to recognise the ontological differences between them. In fact, as a teacher only exposed to natural science methodologies, I expected that an empirical study might well suit this research project, until careful consideration of the research questions and the knowledge necessary to explore them was undertaken.

As a former classroom teacher and now a teacher educator, I have direct experience of the practices and contexts under investigation in this study. This proximity means I cannot be considered a straightforward outsider for all aspects of the research, and it carries implications for my epistemological positioning. Biesta (2007) notes that following the best approach to answering research questions, rather than being led by prior epistemological training, is a common challenge for teachers. Similarly, Scotland (2012) warns that adopting an objectivist approach to knowledge production whilst disregarding ontological assumptions may lead researchers to inappropriate conclusions. Acknowledging my position as an insider researcher for some aspects of this study was therefore an important step in ensuring that my methodological choices were driven by the research questions rather than by the assumptions I brought to them.

3.2.1 Researcher positionality

Educationalists and many other professionals can complete educational research within their own institutions. These research professionals, of which the author is one, must be mindful of how this position may influence their research practice and the interpretation of their findings. Being a research professional can bring challenges associated with anonymity for those they are investigating and the outcomes of their work (Atkins and Wallace, 2012; Elliott, 2023). Where the research occurs within the research professional's institution or where I am invited to conduct research, I might be considered an insider. In contrast, when I have recruited external participants for my work or examined the practices of institutions outside my work, I may be said to be an outsider. To be aware of how this might influence my work, I must be mindful of my position as a former teacher in England and as an insider researcher, as well as an outsider when researching education in Scotland.

The insider/outsider position may be considered on a continuum. However, Trowler (2011) argues that a researcher is never truly one or the other. A researcher can be said to operate along various parts of this continuum, as demonstrated by the questions posed by Atkins and Wallace (2012). From this thesis, I would not position myself as an insider researcher, as it does not take place within my own institution. However, the English education system will be more familiar due to their professional experience. At the same time, as a former teacher, I may not be viewed fully as an outsider in some school contexts, but as an education researcher external to the school and country. I occupy a dual position as both insider and outsider within this study. This positioning necessitates sustained reflexive practice throughout data collection and analysis. Specifically, the research design incorporates a systematic interrogation of potential bias arising from professional familiarity with educational contexts, alongside a critical examination of how insider status may shape the interpretation of participant narratives. These reflexive strategies aim to mitigate the methodological challenges inherent in occupying multiple positionalities within educational research. Using this position as a previous natural scientist who is now conducting education research in participating institutions, several theoretical positions were considered.

3.3 Theoretical framework

Social theory provides analytical frameworks for examining the lived experiences of educational actors. For this research, theoretical frameworks are employed to analyse the relationships between policy formation and enactment and to examine how social actors mediate policy enactment processes. Bradbury *et al.* (2022), working through the COVID-19 pandemic demonstrates that schools may not always be impacted in the way that policy creators intend, and some may act in opposition to it. In exploring these events, many research tools and paradigms may be used. This section first establishes the theoretical position underpinning the study. It then explores how other researchers have investigated the relationship between policy and practice, distinguishing the approaches they adopt and the forms of understanding they generate.

As already described, based on my previous perspectives in natural science research, positivism may have framed this work. This approach, born from the

Enlightenment period, has been attributed to Auguste Comte, who argued for research to collect objective evidence in the form of facts that could be scrutinised and replicated (Cohen, Manion and Morrison, 2018b). Positivist research begins with hypotheses about observable phenomena, which can then be tested empirically to establish verifiable facts or universal laws. The overall aim is to achieve proof in the form of a general law. In natural science subjects, where the variables of investigation might be reduced, the production of universal laws is an achievable aim and yields impactful conclusions. In research fields, where there might be less control over variables, such as in social environments and where the research context can add to the meaning of any conclusions, positivist paradigms may not produce the most impactful findings (Cohen, Manion and Morrison, 2018b). I cannot occupy a position of complete neutrality within the research process; their presence and interpretative framework inevitably shape the generation and analysis of data.

Max Weber challenged the assertion that positivism was the only, or best way to explore the social world (Cohen, Manion and Morrison, 2018b). He argued that the subject of sociology aims to interpret social phenomena, which may not lend themselves to universal laws of understanding. The notion that human beings could be studied in the same way, at a distance and without affecting their behaviour, did not seem appropriate or likely. The behaviour of those being studied is open to personal choice and agency and therefore may not be consistent with any universal law. Not only that, but it will likely be affected by culture, age, and disability, to name a few variables, but also the behaviour or social position of the observer (Cohen, Manion and Morrison, 2018b). The realist paradigm seeks to understand how social actors operate within specific contexts. This approach requires situating research findings within their temporal and spatial circumstances to develop contextualised explanations of social phenomena.

How policies are interpreted and responded to within specific national contexts is not accessible through any single research tradition. This thesis examines assessment policy enactment across two jurisdictions, and that comparative purpose requires a research approach flexible enough to capture both the

intentions behind policy decisions and the experiences of those who enact them in schools.

Pring (2015) provides a useful orientation here by distinguishing between positivist approaches, which can identify correlations and patterns warranting deeper investigation, and interpretative approaches, which explore how people make sense of the social world in education. Rather than adopting one paradigm in advance, this thesis takes a pragmatist position, selecting theoretical tools guided by the demands of the research questions (Robson and McCartan, 2015). Pragmatism allows lines of enquiry to be pursued in whichever way is best justified to explore the social situations under investigation.

When considering the purpose of research, a better understanding of the subject matter or answers to predetermined research questions may be regarded as useful justifications. However, researchers inspired by the Frankfurt school of social theorists have suggested that another function of research is to advocate for greater equity and emancipate human beings from societal constraints based on race, class, gender and other social factors (Creswell and Poth, 2018). Critical theory and social justice research offer an opportunity to examine how economic and political structures influence the school assessment system, thereby interrogating common policy claims about advancing equity (Elliott, 2023). Critical Theory enables researchers to explore how social constructs lead individuals to a state of ‘relative powerlessness’ and others to positions of enhanced power (Cohen, Manion and Morrison, 2018b). Critical theory provides an analytical lens for examining how policy exerts power through the school system. It supports critical appraisal of whether policy aims are enacted in schools as complex social institutions, while also offering grounds for considering how those realities might be changed (Troyna, 1994). Using critical theory as a lens would promote the questioning of how teachers exist as professionals and the extent to which they may be empowered to provide the best education for their students.

Considering my position reflexively, a paradigm was chosen that would best facilitate the understanding of the research questions. As a researcher and teacher educator operating in communities with multiple social deprivations, I inevitably had an interest in exploring these questions from the perspective of

how power and control operate through the processes of assessment policy enactment. My position as a teacher would also affect how respondents may interact when revealing their struggles in achieving professional aims of equitable educational experiences for their students to someone sharing their occupational identity.

Reflexively engaging with fellow teachers further necessitates the offer of findings that contribute to understanding how to improve equity in education. This study aims to highlight educational assessment discourse in Scotland and England and explore the extent to which the fields of policy, academia, and practitioners align in their ideals of how assessment should be conducted. By comparing Scotland and England, it becomes possible to discern contextual variations with greater precision, as the countries' shared cultural background enables a more distinct focus on differences in political dynamics.

In the aim of comparing both countries, the concepts of teacher assessment identity, equity in education, and the multiple purposes of assessment will need to be considered. The complexity and potential perspectives of these components offer several paradigms for this research, as discussed previously. A critical theory paradigm is used to achieve the aims of this research, authentically considering my position as both a researcher and a practitioner in education.

Using critical theory approaches does not mean that a pragmatist approach would not be suitable, as it would also allow an exploration of the perspectives on assessment policy, research, and practice in England and Scotland. However, the use of critical theory provides an intrinsic prompt to question the extent to which assessment is purposed to enable or limit social justice and equity in education. Critical theory, therefore, offers the framework best suited to examining how assessment policy frames professional practice and to interrogating the assumptions about equity that policy texts encode.

3.3.1 Structuralism and discourse

It is important to establish what is meant by assessment discourse before proceeding to explore this in policy documentation and conversations with those

who engage with policy. Fairclough (2013, p. 3) outlines discourse as the ‘relations between communicative events’, which can be analysed through the exchange in a variety of these events. Walshaw (2007) writing earlier had used the notion of what can and can’t be said as an included definition of discourse. Discourse around assessment can be found in the media and everyday conversations, particularly in conversations about examinations and the degree to which they may affect students and their futures. Using discourse theory to explore how thoughts on assessment might be conceptualised provides a framework for how assessment processes might be perceived by all those who engage with them.

The research questions required interpreting policy texts to identify the purposes of the assessment processes being advocated. Such texts are typically authored by multiple policy influencers who seek to convey particular meanings. School leaders and classroom teachers, as readers of these documents, must then interpret what the policy changes intend and require of them. This raises a central interpretive question. Either meaning is embedded within texts in ways recoverable through systematic analysis, or it exceeds what writers can convey through the words on the page (Belsey, 2002).

Structuralism offers one route into the analysis of policy texts, though its tools rest on assumptions that the following paragraph will explore. As an analytic posture, structuralism is an anthropological model that draws upon early twentieth-century French linguistic and social theory. It treats behaviour and meaning as expressions of an underlying system that constrains and shapes human action (Silverman, 2022). Within this tradition, Ferdinand de Saussure proposed that the linguistic sign joins a signifier, the word or sound-image, to a signified concept (Walshaw, 2007). This relation between signifier and signified is arbitrary rather than natural, so meaning does not arise from words pointing to objects that exist in the world prior to language. It emerges instead from the systematic differences between signs within a language (Walshaw, 2007). Reading a policy text in this way directs attention to which terms have been chosen, which have been omitted and how the terms used are differentiated from the alternatives available to their authors. Atkins and Wallace (2012) describe this form of reading as textual analysis. The unit of analysis is the

document itself rather than the wider Foucauldian discourses examined on page 113 in this chapter.

However, whilst this work seeks to interpret the intentions of policy influencers, such analysis is complicated by the collaborative nature of policy text production. Individuals within olicymaking groups exercise varying degrees of influence over final documents. This presents challenges for structuralist approaches, since decisions regarding word selection, placement and omission may reflect the preferences of only one contributor among many or may arise from editorial processes rather than deliberate policy choices. Given these analytical challenges, this study examines the social construction of meaning in policy texts. From this perspective, the significance of terms and phrases emerges through processes of negotiation among policy actors, resulting in texts that, whilst collectively established, may not reflect unanimous agreement.

To help understand how assessment is conceptualised within policy texts and professional discourse, it is necessary to examine the purposes attributed to it. Such purposes typically address three dimensions: what assessment information is intended to achieve, when it should be collected, and for whom it is designed to be useful. Through this description, not only is permission given for assessment tools to be deployed, along with the circumstances under which they are used, but also, through their omission, some purposes are deprioritised or discouraged. In this way, both the boundaries of assessment discourse, what is discussed and what remains unspoken, and the conceptualisation of assessment itself can reveal the intentions underpinning policy texts and stakeholder contributions.

Discourses may be shaped by political environments, with particular actors wielding significant influence over what can legitimately be discussed (Ball, 1993). Broadfoot (1999) argues that dominant groups may direct discourses to align with their interests. Within educational assessment, this enables powerful actors, including policymakers and governing bodies, to control public understanding by determining which aspects of assessment practice enter legitimate debate (Richardson, 2022). Policy texts function as key instruments through which this influence operates, framing permissible discourse and thereby affecting changes in assessment practice (Hatcher and Troyna, 1994). To

examine this influence, it is necessary to analyse both the discourse embedded within policy texts and the discursive experiences of practitioners working within these policy environments.

Discourse analysis in education has been critiqued for its traditional focus on policy texts as vessels of discourse rather than as products (Ball, 2015b). As originally described by Michel Foucault, discourse refers to the social knowledge of what can and cannot be said, rather than to what is contained in texts (Foucault, 2001; Walshaw, 2007). This distinction has significant methodological implications for how assessment policy might be investigated. Accordingly, this thesis explores assessment discourse not only through policy texts as products of this discourse but also through semi-structured interviews with those involved in policy creation and enactment.

3.3.2 Critical theory in educational policy research

A critical-theoretical lens allows the research to examine how policy documents and participants categorise the actors involved in assessment reform. Of interest is not just what is said but also what is allowed and not allowed to be said and by whom. Creswell and Poth (2018) define critical theory as the alignment of methodological positions. Critical educational research is explicitly designed with the deliberate political intention of emancipating individuals and groups in an egalitarian society through identifying and interrogating illegitimate power structures and promoting freedom and social justice (Cohen, Manion and Morrison, 2018a). This approach is fundamentally influenced by Paulo Freire's work on participatory action research, which commences with explicit social and political intentions that articulate with the dominated and poor classes, emphasising that research must be educative and involve popular participation as a democratic process of conscious action (Cohen, Manion and Morrison, 2018a). Context is crucial for educational change, and policy can have differential effects depending on who it is enacted for and when (Braun *et al.*, 2011).

Critical theory explores interpretations of social constructs through a critical lens (Cohen, Manion and Morrison, 2018c). It analyses these structures and the effects that they can have on individuals and groups within society. With a

critical lens, judgements can be made about the extent to which these people in society may be empowered or disadvantaged by the social constructs that are explored.

Within this work, the context of Scotland and England allows an analysis of two different policy fields. This provides a novel comparison of not just how policy might be enacted within one government, but a comparison of any differences there may be in classroom practice. When analysing the discourse created by policy enactment, there are several methods that may be used. Critical discourse analysis draws on a diverse set of methods, reflecting research in the field (Fairclough, 2013).

Critical discourse analysis offers a qualitative approach to exploring how language choices shape and reflect the social world. This methodology employs analysis of written texts, spoken language, and communicative practices to interpret the intent and enactment of policy as discourse. It also enables the examination of themes within policy documents, contextualised within the broader societal frameworks in which they operate, and through interviews with those who create and enact policy. Through this process, themes are identified and interpreted, and their differential impacts on various groups are considered. Its critical orientation encourages researchers to ask whether these differential impacts advance greater equity or instead consolidate the interests of more powerful groups (Fairclough, 2013).

This critical orientation encourages the researcher to ask whether such impacts advance greater equity or instead consolidate the position of a privileged few.

Intersectionality questions whether individuals can meaningfully be grouped by apparent disadvantage alone. Within any such grouping, internal differences may mean that disadvantage is experienced unevenly and to varying degrees. This thesis therefore examines how assessment policy may affect different groups of teachers and pupils in different ways. The analysis attends in particular to how interview participants describe the reach of assessment processes and the consequences of accountability for teachers' working lives.

Together, post-structuralism and intersectionality offer a foundation for methods capable of answering the research questions while engaging with existing theory. They support examination of the social constructs through which assessment policy is interpreted, and of the extent to which policy advantages some groups while constraining others. The following section sets out the specific methods adopted and explains how they were applied to the structures considered above.

3.4 Research design

To explore how policy texts construct possibilities for educational assessment practices, it was necessary to interpret the meanings embedded within policy discourse. Using a post-structuralist approach, it was acknowledged that an author's meaning cannot be directly drawn from the words as they appear on the page (Belsey, 2002). Language is an imperfect vessel for meaning, and given that policy documents can have multiple contributors and influencers, they may not always reflect the intended meaning of what is written (Belsey, 2002). Along with other works in policy, such as Ball (1993, 2015), interpretations of policy are undertaken, not just from the words that are used, but also by the ways in which they have been read and understood by the people who interact with them (Ball, 1993, 2015b).

The analysis of policy documents traced how assessment purposes were defined, positioned and related to wider themes such as accountability and school improvement. Reading the texts iteratively, the study attended to which purposes were elaborated at length, which appeared only in passing and which were absent altogether. This attention to emphasis and silence is consistent with the post-structuralist position adopted above, in which what a text omits carries as much analytic weight as what it states (Foucault, 2001; Belsey, 2002). Policy documents were therefore treated as products of discursive negotiation rather than as transparent records of authorial intent, allowing the analysis to consider how such texts may shape rather than merely reflect educational practice.

This created the chronology that is found in section 4 of the literature review. These policy documents were supplemented by contemporary media coverage, which provided insight into both the professional discourse surrounding their

publication and the broader educational context of the period. For the researcher's position, this was particularly important when examining the Scottish context, as the researcher's understanding of England was naturally deeper, given the professional experience discussed on page 109. Exploring how professional reporting portrayed these themes at the time provided a better understanding, which was then explored more critically by professionals working in policy environments.

Understanding how policy actors experience assessment reform requires direct engagement with their professional contexts. The field of sociology has developed methodological traditions that encourage interaction between researchers and the social world they study (Yanow, 2011). This interaction can involve the researcher in dialogue with the communities they are exploring to gain a 'local' appreciation for the social phenomena they wish to understand (Yanow, 2011). For the research questions in this work, it was most appropriate to engage in conversations with teachers, school leaders, and policy influencers who engage with national assessment policy in England and Scotland. This allows me to interview each layer of enactment in policy processes, from those that influence policy intent down to the teachers in the classrooms whose practice the policy may intend to impact. By exploring policy enactment from multiple perspectives, an opportunity for triangulation arises to ensure a richer, more valid understanding of these processes. In addition, as Ozga (1990, p. 359) suggests, in trying to understand policy processes, it is vital to:

> 'bring together structure, macro level analysis of education systems and education policies and micro-level investigation, especially that which takes account of people's perception and experiences.'

Through these interviews, this research sought a better understanding of what intentions policy documents might have why certain assessment purposes might be prioritised for discussion within the published policy documents. Throughout the analysis of these texts, a cautious approach was taken regarding authorial intent. Ball (1993) cautions that policymakers are not always in sole control of what is written in their policy documents, so the texts alone could not fully account for the meanings policy carries in practice.

Building on this caution, the research sought to trace the cascade of policy processes from national formulation to classroom enactment, examining the relationships among policy influencers, school leaders, and teachers. To engage with this process, it was first necessary to understand how policy influencers describe the purposes of assessment and how they perceive policy as having shaped educational assessment practice. This exploration was best facilitated through semi-structured interviews, which enabled discussion of the various themes influencing both policy formation and classroom assessment. Policy influencers may have interests and wish to discuss these openly and at length. To explore how all influencers might perceive other aspects of educational assessment, however, it was necessary to provide prompts in these conversations. In particular, the literature review prompted an exploration of the themes of accountability and the political implications of assessments such as PISA. It was important to elicit any thoughts that policy influencers might have around these topics and how they might wish to influence how schools operate.

3.4.1 Case study approach

With a better understanding of how those influencing policy may deem assessment useful, this work then needed to follow how this might be enacted in schools. The flow of instructions delivered in policy documents can encourage meanings to be interpreted and reinterpreted, in order to negotiate the shared understanding (Bradbury, 2019; Braun and Maguire, 2020). To elicit these understood meanings, the critical theory framework suggested qualitative discussions with Headteachers and classroom teachers were undertaken, and an interpretation was sought of how they describe the purposes of assessment. As Cohen, Manion and Morrison (2018) described, qualitative methods allow for an exploration of what underlies behaviours such as classroom practice, and provide an opportunity to prove what is observed on the surface. This thesis examines how enacted meanings of assessment align with, or diverge from, those set out in policy documents.

In addition to exploring how teachers described the purposes of assessment, it was essential to examine how teacher assessment identities were formed. Using the assessment identity model outlined in the theoretical framework, it was

possible to classify the influences on teachers' conceptions of assessment processes. This approach provided not only an understanding of teacher conceptions but also insight into how professional practice was shaped. The analysis sought to elucidate how teacher practice could be influenced and the extent to which this was attributable to policy imperatives or to dimensions of teacher agency.

3.4.2 Case selection

This work is interested in how schools and teachers might interpret different national policy texts and discourses. To explore how these discourses may differ under the influence of national policies, it was useful to compare two nations within the United Kingdom. Through this comparison, an understanding of how different policies may influence classrooms can be sought. Furthermore, discussions with teachers and policy influencers in these nations can help elicit understandings of teachers' lived experience and how this might influence their practice.

The choice of Scotland and England was influenced by several factors. Working in the northwest of England and studying at a Scottish university, I had access to expertise and resources in these geographical areas. The two nations also serve as contrasting cases: whilst geographically proximate and sharing membership of the United Kingdom, they exhibit divergent approaches to education governance and assessment policy. Interviews with school-based participants in these two jurisdictions would provide insight into how teachers perceive their roles in assessment processes within the scope of this study. The decision was made to focus on England and Scotland, excluding other nations, such as Northern Ireland and Wales. Wales was developing a new national curriculum (launched in 2022), while Northern Ireland has a distinct cultural and historical context that would introduce additional variables beyond the scope of this comparison. To enable meaningful comparison whilst maintaining feasibility, one city in England and one in Scotland were selected as research sites.

3.4.3 Policy influencers

In addition to selecting schools from these two nations, policy influencers were recruited to explore how policy has been shaped within each context. The inclusion of policy influencers from both Scotland and England facilitated an examination of each jurisdiction's distinct context and its potential impact on policymaking and enactment. Policy influencers were purposively selected from various spheres to capture multiple perspectives on how policy is influenced and mediated between policy document creators and classroom practitioners. Table (3.2) presents the different policy spheres represented within this study, accompanied by jurisdictions for their inclusion and the anticipated analytical contributions of each.

Table 3.2: Rationale for categories of policy influencers.

Policy Sphere	Rationale for choice
Academic government advisors	Academic advisors are often used by politicians to provide an understanding of theory and theory and advise on enactment. Three advisors from Scotland and three from England were chosen to allow for an understanding of how the current assessment policy has developed.
Government policy critics	Policy critics form a vital function in helping to evaluate the success or limitations of policy intentions and impact. Two advisors were selected from each region as interviewees, providing insight into their roles and approaches to olicymaking in England and Scotland.
Unions	Unions can offer critical appraisals of national and local policies and their effects on schools. As providers of professional development, they may also work to mediate and develop meaning from policy texts. This work was discussed with union representatives to better understand their perspectives on assessment policy and the extent to which they feel it has influenced classroom practice.
Government politicians	Government ministers will receive recommendations from advisors and civil servants to make policy decisions. This group of participants was chosen to help understand how decisions were made and how they might have influenced policy creation in recent years.
Professional organisations and teacher Professional development	Professional organisations can work at the Meso-level to interpret policy for practitioners and connect policy requirements to practice. The discussions with these organisations and individuals will help develop an understanding of how policy messages might be translated, to focus on practice in a way that might be better understood.
Local Policy Leaders	Local policy leaders will mediate national policy and work with schools to enact it in local contexts. With an understanding of their local educational needs, they would help shape and focus policy documents to drive change within schools. The leaders were approached to better understand how they worked with policy processes to create beneficial change for their local schools.

Interviews with policy influencers provided valuable insights into the multiple layers of policy actors who may interact with assessment policy processes. As described in the literature review, this interaction is not a linear process; accordingly, the present study did not categorise these influencers hierarchically but rather conceptualised all participants as contributing to discourses around assessment policy that are subsequently received and interpreted by school practitioners. Analysis of these discourses for themes relating to the purposes for

which assessment might be used, such as accountability, motivation, and informing learning, helps to shape participants' conceptions of assessment. Understanding these discourses is useful for identifying 'damaging discourses of duality' (Richardson, 2022, p. 51), wherein certain purposes are presented as coexisting harmoniously with no acknowledgement of the tensions between them or the consequences that may emerge. These discourses may continue to be mediated at the school or micro level of policy enactment; consequently, it was essential to explore the influence on classroom practice with school-based practitioners.

The research questions posed meant that it was necessary to explore the perceptions of teachers working in the two policy environments of Scotland and England. To gain an understanding of these environments, it was useful to choose schools with similar settings and comparable characteristics. This meant that differences between how teachers perceived assessment processes might be attributed, to some degree, to the influence of national policy. The intention was not to make generalisations about how policy impacts upon a whole nation but instead focus on case-study schools working in one region of each nation. This would allow for rich data from interviews with participants about how policy is mediated in these case study schools.

Liverpool was chosen as the case-study city region for England, while Dundee was selected for Scotland. With an estimated 148,820 residents, Dundee ranks fourth in terms of population in Scotland (National records of Scotland, 2021). Life expectancy in Dundee for males is 73.92, which is the second lowest in Scotland, and 30.1% of children in Dundee reside with families experiencing low income and material deprivation (Dundee City Council, 2021). The Liverpool City Region includes the Liverpool City Local Authority, including the local boroughs of Halton, Wirral, St Helens, Sefton and Knowsley and was established in April 2014. This region is England's third most deprived local authority area (Liverpool City Council, 2019) with the Liverpool City Council alone having a population of 505,000 (Demographics headline indicators, 2021). Whilst the Liverpool City Region is considerably larger than Dundee in absolute times, each city represents a comparable proportion of its respective national population, making demographic comparison meaningful within limits. Both cities share industrial

histories as maritime trading ports, followed by periods of economic decline as these industries contracted. The participating schools serve communities affected by these regional challenges; however, there remain differences in their specific student catchment areas, reflecting localised variation within each broader urban context.

To answer the research question concerning teacher perceptions of assessment purposes, it was essential to seek an understanding of practitioners' ideas within their context. Hamilton and Corbett-Whittier (2013) explain the benefits of using case study research to understand the complex nature of environments such as schools. This allows the researcher to experience school working ‘within the world of those being researched’ (Hamilton and Corbett-Whittier, 2013, p. 11). Case studies can provide details of the rich research context to explore teachers' assessment practices and decision-making within different policy spheres (Bassey, 1999; Yin, 2018) There are limitations inherent in a case study approach to research, including the tendency towards particularisation and challenges regarding the generalisability of research outcomes (Yin, 2013). Additionally, the production of a large volume of data can present challenges when drawing conclusions (Yin, 2018). These characteristics mean that the findings from this research, which examine specific challenges within the selected schools, may not be directly generalisable to all schools within England or Scotland. However, producing findings generalisable across countries as large and diverse as England and Scotland would exceed the scope of this study. Rather, the intention is to provide detailed insight into how national policies are enacted within specific school contexts, thereby contributing to a broader understanding of this process. As Ball (2021, p. 12) explains, policies are not enacted as such … they are inflected, mediated, resisted and misunderstood, or in some cases simply prove ‘unworkable’. These case studies thus offer insight into policy mediation within different Scottish and English school settings.

This work uses the two contexts of Liverpool and Dundee and situates cases at the level of individual schools. It aims to explore the rationale and descriptions of local assessment policy processes in these institutions. Schools may operate local (‘little-p’) policy processes within national policy environments, negotiating the enactment of national, or ‘big-p’ policy. This creates a unique

microclimate within each school, which will be explored through individual case studies, allowing analysis of materials for teachers with different levels of responsibility within each school.

The selection criteria specified secondary schools within each city region whose catchments served socio-economically disadvantaged local communities. Schools serving such communities were a particular focus of New Labour policy reform and associated research seeking effective improvement strategies (Muijs *et al.*, 2004). Research examining schools in disadvantaged contexts has demonstrated that national policies frequently fail to address their distinctive needs (Passy and Ovenden-Hope, 2020). Consequently, school leaders must often integrate externally imposed policy demands with locally determined improvement priorities (Gu, Sammons and Chen, 2018). The two schools selected for this study exemplify institutions navigating these tensions: both serve socio-economically disadvantaged communities and operate within policy frameworks that apply uniformly regardless of local context. To ensure coherence and rigour across both case studies, the semi-structured interview schedules and subsequent analysis were informed by common themes derived from the literature.

Table 3. below identifies the categories of school professionals who were interviewed as part of this work. To gain an insight into how policy processes might affect the practice of professionals working inside a school, it was necessary to interview members of staff from various levels of experience and leadership. There were restraints on the number of staff that could be interviewed due to the pressure they faced during the COVID-19 pandemic and recovery. Due to these restraints, it was only possible to interview a few members of staff within each category and explore their practice within assessment processes. It had been hoped to explore how teachers of different subjects might experience working within their school or nation's policy processes; however, it was not possible to work with the number of teachers this would have required. It was possible, however, to work with members of staff from practical and theoretical subjects so that comparisons could be made around the experiences of these teachers.

Table 3.3: Rationale for categories of school professionals.

Teacher Career Stage	Rationale for choice
Early Career Teacher / Probationer	Secondary Teachers in the early stages of their careers have undergone a period of training which will have provided them with a basic knowledge of assessment and pedagogy for their subject. Interviewing early-career teachers will allow a perspective of how teachers view their role and agency in assessment processes. It is of interest to explore/investigate how early teachers feel as they start to work with these policies and their students.
Experienced Classroom Teacher	Experienced classroom teachers are described as having had experience of teaching and assessing their subject for a period of at least three years. They will have experienced the various assessment processes within their school and will be expected to follow their school policy. Experienced classroom teachers may have a more developed understanding of their subject and pedagogy. Further, they will have experienced policy changes and may have opinions to offer on the extent to which these policies affected their practice and perception of assessment processes. They will have specific perspectives on their student's needs and knowledge of the impacts of policy on these students.
Head of Subject / Principle Teacher	Teachers who have leadership responsibility for their subject area are often more experienced members of staff. They not only have a depth of knowledge of the demands of their subject but are also charged with enacting wider school policies into their department. Interviewing teachers with responsibility for a particular subject will allow a perspective of subjects across a whole-school cohort. They will have a wider awareness of the impact of policy across a whole school within their subject and will perhaps have an opinion on the extent to which they have positive, negative, or neutral consequences.
Senior School Leader	Senior school leaders may be responsible for whole-school policies, including assessment. They may be informed by the subject they have taught and the national policy that they have been asked to enact in school. They will also have a greater role in school inspections and responsibility to monitor the school's progress in meeting inspection targets. Senior school leaders may be able to describe the balance they aim to achieve between working with whole-school targets and supporting subject leaders in enacting policies in classrooms.

Head Teacher	Headteachers have experience with changes in national policy and enacting policy within their schools. They will also have a wider knowledge of, and demands from, key school stakeholders' interests. Headteachers may be able to describe the effects of national changes in assessment policy and how their school has enacted them. They may talk about the challenges faced by their school and the extent to which changes in policies enable them to serve their school community.

3.4.4 Data collection

The discussions with policy influencers were facilitated through semi-structured interviews to explore alignment with perceptions of assessment and the research literature. Semi-structured interviews allowed the interviewer to direct the conversation, ensuring that all research questions were addressed while also inviting policymakers to openly and freely discuss their thoughts and perceptions regarding assessment (Cohen, Manion and Morrison, 2018a). The interview questions were developed with a representative policymaker to ensure they were relevant and used language that prompted discussion of the various priorities and experiences of those involved in policy creation.

The interview protocols (Appendices A and B) provided sufficient structure to ensure that all requirements of the University of Glasgow ethics approval were met. Any question prompts explored the necessary aspects of research questions in conducting the interviews. Using a semi-structured protocol with consistent questions and prompts avoided imposing the interviewer's views beyond the questions or seeking an external truth beyond that experienced by the interview participants (Morris, 2009). Using a loose structure that allows participants to talk freely helps ensure that the research questions' priorities shape the conversation rather than the interviewer or interviewee (Morris, 2009). Throughout the interviews, it was necessary to allow policy influencers to discuss their perceptions of the rationale behind enacted policies as freely as possible.

In following the development of discourse through policy enactment, it was possible to explore the extent to which these policies have impacted school policy and classroom practice. This impact may affect policy documents in schools, as they interpret the national policy to craft their own assessment

policies for local use (Ball *et al.*, 2011). Studying school policy? documents may indicate how the national policy documents influence school policy but may not reveal whether this has an impact beyond the sole author. It was crucial to try and understand how teachers' perceptions of the purposes of assessment had changed. As suggested in the literature, teachers may not always follow policy as written at national or school levels and may not believe in the policy that they obligingly enact in their own classrooms (Ball *et al.*, 2011).

For investigating the use of assessment and perceptions of school professionals, semi-structured interviews were chosen. Classroom observations of teacher practice in case study schools were considered; however, they were not possible due to COVID-19 restrictions. While these classroom observations might have shown how teachers use assessments, understanding why teachers use assessments and what informs that use was pertinent to the research questions and could be explored through semi-structured interviews (Yin, 2018).

Qualitative interviews are useful to better understand the perceptions of participants, such as classroom teachers and school leaders (Cohen, Manion and Morrison, 2018b). The teacher participants had to be able to talk freely about their own practice and what might influence it, whilst for school leaders, it was necessary for them to describe what they would encourage and like to see happening around their school. Preliminary work suggested that using artefacts would provide teachers with the opportunity to think more specifically about their teaching practice and how they use assessment for various purposes. In semi-structured qualitative interviews, a number of predetermined themes and suggested topics on assessment could be used to elicit discussion. It also allows an 'openness to changes of sequence and forms of questions to follow up on the specific answers given and the stories told by subjects' (Brinkmann and Kvale, 2015, p. 150). Semi-structured interviews allowed for a dynamic approach, whereby the themes of enquiry from the literature could be used to set a topic of conversation and be led by how teachers described their practice. It was possible to follow up with questions about what informed the practice and how interviewees had acquired or developed it in their classroom or school. Teachers' descriptions and their choices about what to raise could help build an understanding of what influences practice within schools.

Cohen, Manion and Morrison (2018) outline the need to consider the importance of comparable data when designing interviews, as this is related to the extent to which interview questions and style may need to be standardised and generic. Overall, in this study, the fitness for purpose of interview styles was considered and semi-structured interviews were chosen to ensure that important themes were uncovered, whilst the direction of participant discussion throughout the interview remained flexible.

To be sure of eliciting a rich dialogue, trial interviews with volunteer teacher colleagues were conducted as part of a pilot study. Four methods of eliciting discussion of practice were trialled. Each method consisted of a single stimulus artefact and the simple question, 'Using this as a prompt, what can you tell me about the purposes of assessment in your classroom'. The interviews were voice recorded and transcribed using Zoom software.

The purpose of this pilot was to find out which of the artefacts produced the richest discussion about assessment in the teacher's classroom. The artefacts were used to allow the teachers to reflect on their practice, prompting their thoughts on classroom actions that utilise assessment processes. This means not just the volume of discussion that teachers offer? but also the depth of analysis they provide in discussing how they use assessment in their classroom. The most effective artefact would allow teachers to talk freely about the multiple assessment purposes in their classroom.

The first method involved showing the teachers a generic classroom situation. The image was downloaded from an open-source website and displayed in a normal classroom environment. The teacher was asked to look at the picture and discuss how it made them think about the purposes for which they use assessment in their classroom. Teachers were able to talk freely about the multiple assessment purposes in their classroom.



Figure 3: A stimulus image of a classroom. Image sourced from: www.unsplash.com. Accessed: 21/6/2021

The second stimulus material was a transcript of a classroom interaction, sourced from (Wiliam, 2011). Participants were first asked whether they found any aspect of the transcript difficult to read or understand, and were then invited to reflect on how the interaction depicted connected to their own classroom assessment practices.

For the third method, the views of teachers were explored using a video of classroom interaction. The video was of an American classroom, where the teacher went around the room and viewed students' work, made comments on it and gave verbal feedback to each student. The teachers were asked to observe the video and consider how it might compare to their own practice and how they used assessment in their classroom. The video was then stopped, and the teachers were asked if they were happy to begin the conversation around assessment in their classrooms.

The pilot study incorporated an additional elicitation tool: the framework table from Wiliam and Thompson (2007), which categorises formative assessment processes according to the roles performed by teachers, learners and peers (see Table 3.2). Although the original framework focuses on formative purposes, pilot participants found it an effective stimulus for discussing summative assessment

purposes. This unexpected utility informed subsequent refinements to the interview protocol, though the table itself was not used in the main study.

Table 3.2: The roles of classroom participants in assessment. Adapted from Wiliam and Thompson (2007).

	Where the Learner is Going	Where the Learner is Right Now	How to Get There
Teacher	Clarifying and sharing learning intentions and criteria for success	Engineering effective discussions and tasks that elicit evidence of learning	Providing feedback that moves learners forward
Peer	Understanding and sharing learning intentions and criteria for success	Activating students as instructional resources for one another	
Learner	Understanding learning intentions and criteria for success	Activating students as the owners of their own learning	

The pilot discussions with teachers revealed that all four methods allowed teachers to relate to their classroom practice. Teachers were asked a single question,

Looking at the images in front of you, can you tell me how you use assessment in your classroom and what influences your reasons behind these purposes?

There was variation in the depth to which teachers discussed their practice and how the approaches allowed them to discuss relevant topics around assessment

practice. Some of the artefacts were very leading, which meant that teachers were not able to discuss wider issues or any part of their practice that was not reflected explicitly by that artefact. This made artefacts highly appropriate for use in the main project, and some of them of less useful in encouraging teachers to talk freely about their practice.

3.4.5 Pilot study

Artefacts one and four proved most productive in eliciting discussion with participants. Despite artefact one being a photograph of a specific classroom, participants identified a diverse range of classroom practices from the image. Participants were able to elaborate about their own role and the role of students in these assessment dialogues, as well as the purposes underpinning such practices. Participants also discussed the role of accountability and its influence on teachers' decision-making.

According to Wiliam and Thompson (2007) Artefact 4 (Table 3.2) was intended to show strategies for formative assessment; however, many participants spoke at length about using this artefact for summative purposes. In addition, participants described the roles of multiple stakeholders in their teaching, including Ofsted, parents, and their school leaders. This artefact was by far the best at eliciting a diverse conversation about multiple purposes of assessment.

The third artefact appeared to constrain participants' responses, directing their attention exclusively towards the specific classroom practice depicted in the video rather than broader assessment considerations. Whilst discussion of the practicalities of enacting assessment was relevant to the research aims, participants tended to describe and evaluate only what they had observed, rather than engaging with the wider purposes of assessment or the factors that might shape teachers' assessment practices. Obtaining a diverse range of video artefacts would present practical challenges, as permission from schools would be required and individuals would need to consent to being recorded. Given that this artefact elicited limited personal discussion of broader assessment themes, video-based stimuli were not employed in the main study.

The second artefact, a classroom dialogue, did not elicit much discussion among teachers about assessment. The teachers were able to discuss times when they had similar conversations and how frustrated it made them feel. They focused on the exact scenario given in the conversation and were not prepared to discuss anything beyond it. There was a brief conversation here about the purposes of assessment, so it was the least useful.

Following the pilot project, artefacts one and four were selected for use in the main study. These were presented together to elicit discussion, with participants invited to use either artefact as a stimulus for reflecting upon how assessment processes might occur in their own classrooms. On reflection, the absence of school uniforms in artefact one may have limited participants' ability to identify with the depicted scene. Consequently, the image was replaced with Image 2 (below), which shows a similar classroom conversation. In this revised image, students are visibly engaged in, or have recently completed, written work. This detail was intended to suggest that assessment processes depicted in the scene might encompass written work as well as assessment of verbal contributions, rather than implying reliance on oral contributions alone.



Figure 4: Photograph of students and their teacher in a conversation. Obtained with permission for use from Pixabay.com

Online recorded interviews were conducted with individual classroom teachers and school leaders at various levels of institutional hierarchy. This approach enabled examination of how national assessment policy is interpreted and enacted differently across organisational positions, recognising that school leaders' perceptions of what policies demand may diverge from classroom teachers' enacted practices (Ball *et al.*, 2011). Such differences can manifest as a spectrum of assessment practices that constitute individual teachers' assessment identities, as discussed in the literature review.

A single interview protocol (Appendix B: Interview schedule for educational professions) was used with both teachers and school leaders, incorporating the same stimulus images and core questions. Consistent with the semi-structured rationale outlined above, the researcher adjusted the emphasis of interview questions to reflect each participant's professional context. In practice, school leaders were invited to elaborate on how they interpret national policy and frame assessment expectations through school-level documentation, whereas teachers were encouraged to discuss assessment processes in their own classrooms and the factors they perceived as influencing them.

The interviews were recorded via Zoom conferencing software, which includes an audio recording function and a transcription tool (www.Zoom.com). The transcripts were checked for errors and sent to the participants to ensure they were happy and supportive of the statements contained therein. Where participants wanted to redact or correct statements, this would have been welcome, although no participant requested it. Once the accepted transcripts were collected, it was possible to begin exploring the emerging themes concerning the purposes of assessment.

3.4.6 Data analysis

This study employs a variety of participant categories to capture diverse perspectives on policy processes pertaining to educational assessment. The first cohort, the policy influencers, was analysed to develop an understanding of the policy creation process. This analysis also examined their perspectives on how policy documents function within broader policy processes to enact changes in schools and the wider education system.

Subsequently, comparisons were undertaken both between the various groups of policy influencers and between the English and Scottish contexts. These analyses illuminated the assessment purposes prioritised by policy influencers, as well as the mechanisms intended to drive changes to assessment processes. The findings also revealed social factors that may have impeded change in assessment policy and in education more broadly.

The second cohort of interviews enabled analysis of the assessment processes in two schools in England and Scotland. Gaining the perceptions of teachers and leaders at various levels within the school hierarchy allowed for an understanding of what informs their practice and how they perceive these impacts on their students. In particular, this study was interested in eliciting ideas about the teaching professional's assessment identity, which is discussed further below. Once all participants had consented to the use of their transcripts, they were analysed using thematic coding software.

NVivo, a computer-assisted qualitative data analysis software, was used to code transcripts from interview participants. To compare the themes from policy

influencers and themes from teachers and school leaders, it was useful to use similar lines of analysis for both sources of evidence. Therefore, the same questions were applied when analysing the interview transcripts from school professionals and the participants from the policy-influencing sphere. When analysing the transcripts, the following questions were used:

1. Within the transcripts, what are the key purposes for assessment processes that are advocated or mentioned as possible?

2. What influences the participant's views of assessment processes and their purpose? Is it what they have learnt, their personal experience as a student or their experiences as a professional?

3. Are these multiple purposes harmonious, or are there possible tensions between the demands placed on the assessment process, those taking part in the process, or the use of information produced from these processes?

Thematic analysis was selected for its theoretical flexibility, enabling deployment across epistemological positions without being tied to a particular theoretical tradition (Braun and Clarke, 2021). This was essential for a study operating across two jurisdictions and two participant cohorts. Unlike interpretive phenomenological analysis or grounded theory, thematic analysis permitted the identification of patterned meaning across diverse accounts while accommodating the study's critical theoretical orientation.

When considering the purposes of assessment processes, as advocated by the participants, several classification approaches were available. While established taxonomies for assessment purposes exist within the literature (Newton, 2012, 2017), adopting these frameworks risked constraining the analysis to predetermined categories that might not capture purposes unique to participants or to the specific school cases within this study. Instead, an inductive analytical approach was initially used to code the transcripts. This approach treats the dataset as the starting point for theme generation (Braun and Clarke, 2022, p. 56), with the researcher actively constructing themes through iterative engagement with participants' accounts. Initial themes were

identified from a selection of interviews to create a provisional coding framework, which was subsequently applied across all transcripts. Where new themes emerged, they were added to the code list, and earlier transcripts were revisited to check applicability. This study adopted a reflexive approach to thematic analysis (Braun and Clarke, 2021, 2022). The provisional codes generated from initial transcripts were not treated as a fixed framework but were revised, merged, and expanded as subsequent transcripts introduced new patterns. This recursive process distinguishes the approach from codebook or coding reliability variants, in which themes function as analytic inputs determined prior to engagement with the data (Braun and Clarke, 2021).

The NVivo software facilitated the exploration and analysis of relationships and patterns between the coded purposes and respondents' categories. This approach enabled interpretation of the perceptions of policy influencers and school professionals on their own terms, rather than constraining them within a pre-existing framework. Within the semi-structured interviews, it was important to create space for participants to articulate how their own conceptions were shaped by their context. This approach held the potential to reveal novel purposes for assessment not yet described in the literature, whereas an approach using predetermined themes might risk ‘confining the data analysis to my own agenda rather than “the other”’ (Ezzy, 2002, p. 85). For this reason, it was important to adopt a reflexive approach to analysis, one grounded in the meanings of participants, whilst acknowledging the difficulty of fully separating my values from those of my participants (Atkins and Wallace, 2012). The use of NVivo supported coding and the management of multiple themes simultaneously; however, the transcripts were first read outside of the software, with written notes recording emerging themes, rather than relying solely on NVivo functions, which risk distancing the researcher from participants’ situated accounts and them (Braun and Clarke, 2022).

When identifying and interpreting themes within the transcripts, it was essential to acknowledge my former position as a teacher in England. Within reflexive thematic analysis, this professional background is understood as an analytic resource that informed interpretive sensitivity to the conditions within which teacher participants operated (Braun and Clarke, 2021). Within a critical theory

approach, the reading and analysis of this work involves meaning-making to develop an understanding of how power is asserted through policy processes and accountability mechanisms to restrict or liberate teachers' choice of practice. To enable this process of understanding and interpreting what teachers communicate, reflexivity is required, with consistent reflection on what is revealed through the process of examining power relations between participants (Braun and Clarke, 2021). To enhance the understanding of Scottish policy environments, the school visit (discussed on page 104) took place.

However, contextual familiarisation alone cannot ensure reflexivity during analysis, and it was therefore essential to employ additional strategies to enhance the trustworthiness of the data (Lincoln and Guba, 1985). These included recruiting participants from a diverse range of national policy influencer groups and sharing transcripts and recordings with participants to verify that their contributions had been accurately captured.

In analysing interview data, this study adopted a position between two tendencies identified in qualitative research (Parker and Knight, 2025). Participants' accounts were approached as situated perspectives shaped by professional context, positional interests, and the dynamics of the interview encounter itself. This required interpretive work that went beyond verbatim reportage to examine what participants' language might reveal about the structural conditions within which they operated (Parker and Knight, 2025).

The advantage of using a case-study methodology is that the contextual experiences of participants can be considered (Yin, 2018). Specifically, contextual factors help to explain why school professionals employ particular assessment processes and why policy influencers seek to shape policy development in specific directions. From the perspective of these policy influencers, such an approach also enables exploration of how policy has been formed and of the underlying intentions during the development process. By situating participants' accounts within their respective contexts, this methodology supports analysis of the construction of assessment policy at the national level, examined through the perspective of policy influencers, and its enactment within schools, explored through the experiences of teacher

participants, for whom Ball's policy enactment framework provides the primary analytical lens.

The semi-structured interview schedule was also designed to elicit how school professionals come to understand and enact their identities as assessors. When analysing these transcripts, the assessment identity framework proposed by Looney *et al.* (2018) was employed as an analytical lens (See Figure 5). This framework facilitated the identification of factors contributing to individual teachers' understanding of assessment processes. These factors were then coded within the transcripts to explore which formative experiences were most frequently associated with teachers' conceptions of assessment purposes. Through subsequent NVivo analysis and the generation of visual data representations, comparisons across the different formative events in teachers' personal and professional lives became possible. In this way, Looney *et al.*'s (2018) framework functioned as a sensitising concept, directing analytical attention towards particular features of teacher accounts without constraining the thematic analysis within predetermined categories.



Figure 5: Framework for conceptualising teacher assessment identity extracted from Looney et al. (2018).

In addition to the assessment identity framework, Le Grand's (2006) typology of knights, knaves, pawns, and queens provides a further analytical resource in the later interpretive stages of this study. The typology was not applied as a coding

framework during the initial thematic analysis of interview transcripts, where the primary task was to identify and interpret the meanings that individual participants attached to assessment processes and their professional roles at the level of their own accounts. Le Grand's framework operates at a higher order of abstraction; rather than illuminating individual contributions, it examines the assumptions about professional motivation and agency encoded within governance structures as a whole. As with many typologies, these are idealised types, and the same individual may be interpreted differently depending on context and perspective.

To ensure the rigour of this study and demonstrate that the chosen methods meet appropriate standards, careful consideration was given to questions of reliability and validity. As a former natural sciences undergraduate with an understanding of these concepts within assessment contexts, I might have approached this study with more positivist interpretations of quality criteria. However, for an interpretive study seeking to elicit the perspectives of professionals working within schools facing challenging circumstances in a specific UK nation, such traditional conceptions of reliability and validity would be neither practical nor appropriate.

Within the literature on case study research, there is considerable discussion regarding validity and reliability (Cohen, Manion and Morrison, 2018c). Cohen, Manion and Morrison compare the generalisability of case studies to single experiments within natural science, noting that both research approaches can appear isolated and without replication. However, Yin (2018) argues that findings can demonstrate reliability and consistency when multiple case studies addressing similar phenomena yield comparable results, thereby enabling researchers to corroborate and contrast findings to develop robust qualitative interpretations. Establishing this discussion, Yin (2013) addresses the generalisability of case study findings, acknowledging an inevitable compromise between achieving an in-depth understanding of individual cases and selecting cases sufficiently representative of broader populations. These considerations of reliability and generalisability informed the design of this study and were balanced against the need to authentically address the research questions.

This study required access to schools whose contexts would illuminate the challenges inherent in the enactment of assessment policy. It was also necessary to identify school professionals in England and Scotland willing to discuss how assessment processes operated within their classrooms and the factors shaping this practice. For these reasons, and due to constraints imposed by the COVID-19 pandemic, data collection was confined to one school in Liverpool, England and one in Dundee, Scotland. Whilst these findings cannot claim generalisability across either nation, the issues faced by these schools are unlikely to be unique; rather, the study may reveal common factors affecting policy enactment processes in both jurisdictions.

One crucial advantage of case study research is its internal validity, which permits inferences about causal mechanisms and pathways (Gerring, 2017). This strength is particularly suited to the present study's research questions, which seek perspectives from teachers and policy influencers. These participants offer causal explanations, grounded in their observations from within the education systems of Scotland and England. Such in-depth explanations, grounded in lived experience, are better suited to addressing this study's research questions than approaches prioritising statistical generalisability. The teacher participants serve as illustrative examples of practitioners operating within their respective policy contexts, rather than as a representative sample from which broader claims might be extrapolated.

3.5 Ethical considerations

Data collection involved interviews conducted in two stages: the first with policy influencers and the second with educational professionals working within schools. Throughout the research process, awareness was maintained of the demands that participation might place on interviewees; these demands were communicated transparently, and interview timings were strictly observed to minimise additional burden. Interviews were held online using Zoom software to facilitate ease of participation for both policy influencers and school staff. Initially, it was anticipated that school visits would enable observation of teaching practice and the school environment. However, due to COVID-19 considerations, online interviews were deemed safer for both researcher and participants.

When inviting participants to participate in an interview, the University of Glasgow provided consent forms, participation information sheets, and privacy statements. These documents had been approved through the College of Social Sciences' ethics process prior to the commencement of data collection. Providing these materials in advance ensured that participants understood the purpose of the study, the nature of their involvement, and their rights as research participants. Additionally, participants were informed how their data would be anonymised and stored following their involvement.

Participant personal data were initially stored for the purpose of invitation and arranging interview times. Access to this information was restricted to my sole use to protect participants' confidentiality. The online Zoom interviews were recorded using Zoom's software, and the recordings were securely stored on the University of Glasgow OneDrive in accordance with the University of Glasgow data protection guidelines (2021). Following this, the recordings were transcribed, and the transcriptions were sent to participants for their consent to be included in the research. Participants were also offered the option to redact part or all of their transcriptions from the study before these were used for thematic analysis. Once themes had been drawn from the transcripts, participants were invited to attend focus groups to hear the anonymised themes that had been identified, in an effort to enhance the trustworthiness of these interpretations (Buckler and Moore, 2023). This opportunity allowed them to add further data as to their considerations of the themes.

Participant details were then deleted, and the transcripts were assigned a code to protect the anonymity of participants. In this work and future publications, participants, schools or other organisations were identified using code pseudonyms. This is with the exception of the local area of school participants, which were identified to allow for comparisons of the contexts.

Throughout the ethical considerations of this study design, it was necessary to be mindful of the workload and organisational difficulties associated with working alongside schools. Initially, it was hoped to be able to work with a particular local authority in Scotland, but they had many layers of applications to seek to work. Unfortunately, although the local authority approved the research design, they were not allowing further research in their schools due to

the unprecedented pressures placed upon them throughout, and since the pandemic. It was therefore necessary to seek approval from another Scottish local authority and work with them to apply the research design in a way that was beneficial to both parties. These changes to the research design were necessary, but the new authority did not match the English local authority as well as originally intended due to demographic differences, such as the size of the cities and their populations.

The drawbacks encountered when approaching local authorities and individual schools highlighted the pressure these institutions face. Public institutions may often be placed under high pressure to provide efficient services, but these pressures were particularly apparent and more severe due to the effects of the pandemic (Kim and Asbury, 2020). For me as the researcher, the personal rewards from conducting this study will be a PhD award, publications and perhaps promotion within the workplace; however, there is a risk that resources are taken from these already pressured institutions without a thought of reciprocity (Cohen, Manion and Morrison, 2018c).

Sikes (2006, p. 112) uses the term ‘rape research’ drawn from Lather (1986) to describe research where the ‘researcher gets what they want and then clears off’. It was therefore essential to ensure that participants derived meaningful benefits from their involvement in this study. During preliminary investigations, participants indicated that they found the opportunity to discuss their professional frustrations with an external researcher might support clarity in their thinking and practice through structured reflection. Additionally, all participating schools were offered a reciprocal benefit: I would use anonymised data to address questions the school might have about their assessment practice, potentially informing future improvements (Taylor, 2023). This approach aimed to provide meaningful reciprocity without creating inducements that might compromise voluntary participation (Cohen, Manion and Morrison, 2018c).

3.6 Research limitations

In the final research design, one Scottish local authority was substituted for another, meaning that the English and Scottish local authorities were less well-

matched than originally planned. The original intention was to select two local authorities that were similar in terms of history, size, and socio-economic status. Despite this methodological limitation, the comparison between the two local authorities remained viable because each operates within a distinct national policy context. The capacity to compare practitioners working within these different policy frameworks enables examination of how participants attribute observed differences to their respective policy environments, precisely as the original design intended. Further, this pairing of Liverpool and Dundee constitutes a new comparative contribution to UK educational policy research.

This comparative study of assessment policy enactment in Scotland and England addresses the research questions through two complementary strands of inquiry: interviews with national-level policy influencers from both nations, and focused case studies involving education professionals working within one local authority area within these nations would have enabled further exploration of how national policies manifest across different local contexts. Such intra-national variation exists due to differences in socioeconomic conditions, local culture, and, in some instances, school funding arrangements. I chose not to include the other two nations within the United Kingdom, Wales and Northern Ireland, in this study. This decision enabled a more sustained focus on the policy history and trajectories that have developed across an extended period in England and Scotland and facilitated richer case studies through in-depth interviews with multiple practitioners in each location. This scope, however, limits the extent to which findings can be generalised across the United Kingdom as a whole. To address the research questions effectively, it was essential to foreground the voices of the education participants and to examine the policy contexts that may have shaped their professional identities as teachers.

The COVID-19 pandemic constrained several methodological approaches that would ideally have been deployed during this project, particularly restricting access to schools and teachers in ways that limited opportunities to explore classroom practice. School availability for participation was affected by high levels of staff absence and an institutional focus on student welfare. Ethical considerations required attention to how research presence in schools might risk

disease transmission and add to teacher workload during an already demanding period.

When approaching local authorities and headteachers, it became apparent that schools were experiencing substantial operational pressures due to staff absences or illnesses. Consequently, any additional demands that participation might create for teachers required careful mediation. One local authority initially identified as a potential research site declined involvement, citing the significant pressures facing schools in that area; this necessitated relocating the case-study schools. These circumstances indicated that research proposals involving extended physical presence in schools were likely to be declined, given the potential burden on already stretched staff resources.

To minimise the burden on teacher participants, interviews were conducted remotely using Zoom software. Participants were asked about their perceptions of assessment practice rather than being directly observed. This approach offered flexibility around teachers' schedules and enabled interviews to proceed without visiting schools, thereby avoiding any potential disruption to COVID-19 protocols. However, this method precluded direct observation of classroom practice, meaning the study relies on teachers' self-reported accounts of their work. This limitation did not significantly compromise the research, as the research questions sought to understand teachers' perceptions of working within particular policy environments rather than to analyse their actual pedagogical practice. As Cohen, Manion and Morrison (2018) noted when discussing qualitative approaches such as interviews, such methods foreground participants' voices and provide explanatory accounts of the classroom practices that might otherwise only be observed.

The use of semi-structured interviews with teacher participants enabled a deeper exploration of the motivations underpinning their practice and how these had been formed. However, the time-intensive nature of this approach necessarily limited the number of teachers who could participate from each school. Consequently, patterns identified through these interviews may be specific to the regions studied and cannot be generalised to all practitioners working under assessment policies in either nation. This was not, however, the intention of the research questions, which sought to explore relationships

between practitioner understandings of assessment purposes and how these might align with or diverge from the perspectives of national policy influencers. While it was not possible to capture all such relationships, the detailed nature of the interviews facilitated rich discussion of how teachers engage with policy and the factors that shape this engagement. Furthermore, the study design aimed to strengthen the validity and reliability of findings by triangulating the perspectives of multiple actors across the chain of policy enactment.

Future work may look to increase the scope of this project and address some of the limitations discussed above. One area of interest might be a comparison across more nations of the United Kingdom, offering an interesting analysis of a broader range of policy environments. This would be especially of interest in the case of Wales, where a large curriculum reform known as Curriculum for Wales aims to create dramatic changes to assessment and accountability processes (Evans, 2022). Following how these policies are developed and enacted would allow for an analysis of how new policies are introduced and the challenges the government faces when trying to enact such changes.

Additional challenges not explored in this research include the response of school communities and students to assessment policy reforms. The reception of such reforms by parents, governors, and pupils may pose enactment challenges for both policymakers seeking systemic change and practitioners attempting to modify classroom assessment practices. These stakeholder responses, alongside broader community expectations, can exert pressure that either facilitates or constrains changes to classroom practice.

The research design precluded direct observation of classroom practice. Consequently, the study relied upon practitioners' self-reported accounts of their assessment approaches, which may not fully reflect enacted practice. Future research employing classroom observation methods could usefully explore how teachers interact with students during the assessment process.

3.7 Chapter summary

This chapter has set out the methodological framework for exploring the central research questions. The decision to adopt a critical theory paradigm, situated

with an interpretivist ontological position, reflects the nature of the research problem itself. Educational assessment policy is produced, mediated and enacted within complex social environments. It is shaped by competing interests, historical trajectories, and structural conditions. Understanding these dynamics requires sensitivity to both how policy is discursively constructed and to how it is experienced by those who work within it.

The research employs a two-tiered comparative design. Semi-structured interviews were conducted with national policy influencers in Scotland and England. Focused case studies were then undertaken with education professionals in Liverpool and Dundee. Together, these strands trace how policy moves from national formulation to institutional enactment. This structure reflects Ball's (1993, 2015a) contention that policy enactment is an active, socially situated process. At each stage, it is shaped by professional identities, institutional contexts, and relational dynamics.

These analytical lenses inform the interpretive framework. Le Grand's (2006) typology of motivational assumptions supports examination of the values underpinning policy-design. Ball *et al.*'s (2012) policy enactment framework provides a vocabulary for understanding how policy is received and contextualised at the school level. Looney *et al.*'s (2018) model of assessment identity is employed as a sensitising concept. This informs the interpretive reading of teacher accounts without constraining analysis within predetermined categories. Inductive thematic coding in NVivo structured the identification of patterns across participants' accounts, with the researcher's interpretive engagement mediating between participants' situated perspectives and the broader analytical framework.

These choices constitute a coherent methodological response to a research problem that operates at multiple levels of the education system. Chapter Three presents the findings, beginning with the perspectives of national policy influencers in both nations.

4 Findings

4.1 Introduction

Previous chapters have explored evidence from twenty years of assessment policy development and research. They reveal connections between educational change, assessment policy imperatives, and classroom assessment processes. This chapter explores how education policy actors in England and Scotland conceptualise current and recent assessment reforms, including attempts to embed formative assessment practices and changes during the COVID-19 pandemic. It will examine both the perspectives of national policy influencers, teachers and school leaders within the two case study schools. Finally, this chapter will explore the relationships between the processes of policy creation and enactment at the national and school levels in the presented cases.

Exploring England and Scotland separately enables assessment processes to be examined within each context. This builds an understanding of how policy processes and educational practice interact during policy enactment. Ball (2015b) explains that policy texts influence discourse to create a truth that is translated and enacted in various ways by teachers' creative agency. It is therefore important in this work to compile the words of policy influencers and practitioners in the case study schools and infer the relationships between policy and practice.

The overarching ideas from the literature review in Chapter 2 include accountability, the need for clarity in assessment, and the challenges posed by multiple influences on assessment policy. These ideas informed the questions posed to participants during the qualitative interviews. The participants then expanded upon these ideas, developing them to build the core themes below:

- The diverse purposes of assessment processes in education for and between Scotland and England.

- The understanding and perceptions of assessment processes and their impact.

- The impacts of policy processes on assessment practice.
- The influence of political thinking and discourse on assessment policy as perceived by policy influencers in Scotland and England.
- Consequences of high-stakes testing on teaching and learning in schools.
- The impact of COVID-19 on perceptions of assessment processes.

These six themes cut across the three research questions. The first and fourth themes address the intended purposes of assessment predominantly (research question one). The second and fifth themes contribute to understanding the alignment between policy and research (research question two). The third and sixth themes illuminate how participants perceive assessment policy within their professional contexts (research question three). This structure reflects the interconnected nature of the themes; however, each section contributes evidence relevant to multiple research questions simultaneously.

This chapter discusses the key themes in England and Scotland separately, drawing on interview participants' responses. It then draws these together to make comparisons between the two countries. In the first section below, this chapter explores common ideas about how assessment is used for many different purposes.

4.2 Common themes within this chapter

Although this chapter describes the findings from both countries separately before highlighting common themes, it is important to clarify some key ideas. This provides a transparent basis for the analysis of participant interview data that follows. One such theme is the multiple purposes for which assessment processes might be used.

Table 4.3 below presents the key purposes of assessment and their definitions, as described by participants. Each assessment purpose is organised into a Meta-purpose, which summarises the broad use for which many such processes are deployed in education.

Table 4.3: Definitions for the multiple purposes of assessment processes.

Purpose	Description
Accountability Purposes - When assessment processes are used to hold individuals or institutions, such as students, teachers or schools, to account with consequences such as rewards or sanctions.	
General judgement	The use of assessment outcomes to hold individuals or institutions accountable for their performance, often with consequences or sanctions. Although these were undefined by the participant.
Judgement of the middle tier	Discussions of local authorities or MATs being held to account using assessment processes.
Accountability to school inspectors	Discussion of data from assessment processes being used to hold schools and teachers accountable for the results.
Judgement of schools	Where participants describe the role of accountability in judging schools.
Judgement of teachers	Using assessments to judge the efficacy of teachers based on the results from external exams or internal assessment processes.
Motivational tool	When assessment processes are described as increasing teachers' and students' motivation towards educational improvement. This can be achieved by providing students with a sense of progress or by refining goals to make them challenging yet achievable.
Political motives and justification	When politicians use local and national data from various assessment processes to justify a political position.
Formative Purposes - The use of assessment to inform and improve instruction and/or student learning.	
Checking for understanding	The use of informal, ongoing assessment to monitor student understanding and inform instructional decisions. This is often informal and ongoing.
Diagnosis	Using assessment to diagnose educational needs and abilities.

Demonstrate the need for assessment adjustments for disability and additional needs	When assessment processes are used to demonstrate the benefits of using adjustments has an impact for the purposes of appealing to awarding bodies for reasonable accommodations to assessment routines.
Driving policy and to make change	The use of assessment to drive policy changes in education.
Future guidance	When the results of assessment processes are used for course selection by parents and students.
Informing parents	Where participants discuss the use of assessment processes for the purpose of informing parents about their children's progress and achievement.
Predicting future performance and setting targets	Where student assessment data are used to predict their future performance in exams or set targets against which this performance will be judged.
Inform teaching	The use of assessment to inform and improve instruction.
Provide feedback to students	The use of assessment to provide students with timely, actionable feedback on their learning.
Retrieval practice	When teachers describe using assessment processes for retrieval practice, known elsewhere as the testing effect.
School improvement	Where assessment processes are described as being used formatively to inform school evaluation and improvement.
Student progress monitoring	Where participants describe the use of assessment processes to check that groups of students are making good progress.
Undefined formative	Descriptions of using assessment processes ‘formatively’ without qualification of what this means to the participant.

Summative Purposes - Use of assessment processes to create a summary of student learning.	
Demonstrating student success – unspecified	Using assessment to show a successful educational outcome, although not defined what this may be, for whom it may be created, what it is being used to demonstrate or how it might be shown.
End of period of learning	This is where assessments are used to summarise attainment at the end of a period of learning.
Evaluating research interventions	Discussion of using assessment processes to evaluate the effectiveness of an intervention or educational change.
International comparison	Descriptions of using international assessments, for example PISA, to compare country performance in various subjects.
National quality assurance	Where Assessment processes are used to inform the state about the quality of judged through student performance.
Producing data	Where assessment processes are used to produce data for national, local or personal purposes.
Qualification	Where assessment processes are used to measure student attainment and award qualifications.
School choice	The use of assessment data, especially from external exams, provided to parents to allow for school comparison and school choice.
Selection	Assessment processes that allow some students to be admitted to certain competitive streams, programmes or institutions. Also to inform future employers about potential candidates.
Setting and streaming	Using assessment processes to judge a student's ability and place them in different taught classrooms based upon their comparative performance.
Showing the impact of an intervention	Where participants describe the need to use assessment data to demonstrate a positive or negative impact from an intervention.
Summative - unspecified	When assessment processes are used summatively to judge student attainment, although it is not specified how or for what purpose beyond this.

In the following sections, these ideas will be discussed for England and then for Scotland.

4.3 England

4.3.1 What are the diverse purposes of assessment processes in education in England?

Understanding which assessment purposes dominate professional discourse matters because the purposes that receive sustained attention from policy influencers and teachers tend to shape how assessment operates in practice. Across the English interviews, policy influencers and teachers emphasised markedly different purposes. Policy influencers focused on system-level accountability and the maintenance of standards, reflecting the structural imperatives of English education governance. Teachers, by contrast, discussed assessment predominantly in relation to pedagogical decision-making and pupil progress, constrained by the accountability mechanisms that framed their professional context. The extent to which these perspectives demonstrated convergence or divergence between national and school-level environments illuminates how assessment processes are positioned to serve fundamentally different, and at times conflicting, institutional purposes across policy environments.

4.3.1.1 Measuring and tracking progress in learning

A common theme of agreement between policy influencers, school leaders and teachers was the potential for assessment to help understand what progress in learning was being made and decide what should happen next. Policy influencers had hopes that assessment could help inform students to understand what is needed to improve their learning:

> The kids are able to say well I'm a level whatever, but I'm working towards level the other, and I know how I'm going to get there. (England, Professional organisations and teacher professional development)

In agreement with this, other participants hoped that assessment processes may help improve students' ability to reflect on their learning and to regulate their

next steps toward learning goals. One policy influencer saw this as an improvement to assessment practice that should be a goal of educators:

> Looking forward, the ability to capture what a child's understood or not understood very quickly, in live time, in classrooms. (England, Former NCSL CEO, policy advisor)

These perspectives indicate a broader shift in policy thought towards reconceptualising assessment – moving away from purely summative judgements of past learning and toward a more dynamic, formative approach that actively informs the learning process.

School leaders and teachers articulated similar aspirations for a developmental approach to assessment. In the Liverpool school, there had been a recent change from producing assessment data at certain key points in the year to allow teachers more flexibility over when assessment processes operate to inform subject teaching. Liverpool Teacher 7 emphasised the volume of the old data collection process:

> I think where schools, over the last number of years have gone wrong with that is they've used assessment as a purpose to create data and this kind of vast volume of data, don't get wrong, we're as guilty as any school to create this vast volume of data; five tracking points across a year to create all these data drops which, the whole purpose in the assessment was just to create data rather than actually the sole purpose, which is what have the students learnt? What haven't they learnt? And what misconceptions they have picked up on the way so we can intervene correctly with those students.

The school leader attributed this practice to serving accountability mechanisms

> I think the reason why schools used to have five data drops a year [inputting summative assessment data to a school's central database] was very much Ofsted driven. (Liverpool, school leader)

These accounts suggest that previous school inspection frameworks in England had driven schools to develop assessment policies that generated large amounts of summative data. Liverpool teacher 2 described how teachers in their department were working within the system but acting to undermine the demand for the production and collection of data:

> We've changed our assessments this year, so the way that we assess the students now - it's quite tricky in school because we have calendar tracking points in which we have to deliver data at that set point to report home to parents; what we found in English is it doesn't necessarily always match up those points to what we're teaching. So, we've kind of abandoned those tracking points, and we just report on the newest set of data that we've got from our learning. (Liverpool, School middle-tier leader)

This teacher participant also continued to describe their refocus away from producing large amounts of data and toward trying to provide more teacher feedback, as appreciated by students:

> We've given whole class feedback, and we've given the students their assessments for them to take home, to rewrite and redraft their assessment again; we've had some really positive, lots of positive feedback from the students. (Liverpool, School middle-tier leader)

This school policy transition from rigid data collection schedules to more responsible assessment practices appears to be fostering teacher autonomy and student engagement with feedback.

4.3.1.2 Accountability purposes

Overall, policy influencers felt that assessment processes in England were routinely used to provide data for accountability systems and to shape institutional evaluations. In this system, assessment outcomes operate as both raw material and interpretive evidence, feeding accountability frameworks by providing the quantifiable metrics. Simultaneously, they inform evaluative decisions by establishing the evidence through which school effectiveness, teacher performance, and student achievement are determined and publicly reported. Policy Influencer 1 identified the extent to which exams permeate the education system:

> The exam system is just the backdrop, which is the water in the Goldfish Bowl. Nobody can imagine a system without those exam scores. (Policy Influencer 1, England, Senior academic)

The metaphor of assessment suggests how embedded this national policy influencer considers high-stakes assessment to be in the English context; it is not always seen, but without the water, the fish would die. Removing the exam

system would, to many people, cause the education system to not function. Not only is data from assessments used for accountability, but the frequency with which they are deployed for this purpose is also increasing. Policy Influencer 2, an expert in teacher professional development, explained how this appeared in the system:

> Assessment now, it largely drives the system. We've really bought into that old adage about making pigs fatter by weighing them more often. (England, Professional Organisations and Teacher Professional development)

While these perspectives highlight the pervasive influence of exam-based accountability, questions remain about whether these processes occur without conscious challenge from teachers and students.

The interview data from teacher participants in England revealed concerns about the instrumental use of assessment practices to provide data for accountability mechanisms. Liverpool Teacher 7 described the trend for:

> I think where schools, over the last number of years have gone wrong with that is they've used assessment as a purpose to create data and this kind of vast volume of data. (Liverpool Teacher 7, School leader)

This describes a recent focus on using assessment to compile data sets about students and using this data as an endpoint for later purposes, perhaps for accountability. The tracking of data for accountability was also reflected upon by Liverpool Teacher 4:

> We're using the summative within our department in my opinion, probably a bit too frequently, but also not actually using the information it gives us other than to put into the tracking. (Liverpool Teacher 4, Early career teacher)

This demonstrates a newly qualified teacher who is able to reflect on their practice and is driven to use assessment processes for summative purposes, delivering data that is not used to improve student learning. A recurring theme across the teacher interviews was apprehension about the use of assessment processes to produce data and metrics for accountability rather than to help identify areas for student and teacher development. This apprehension was not unfounded. Teachers' unease about the primacy of using assessment processes

to produce data reflected broader pressures that shaped how these processes were used across the system. The following section examines these pressures, beginning with the accountability and marketisation agendas that framed assessment as a tool of competition rather than of pedagogy.

4.3.1.3 Accountability and marketisation

Many participants were keen to discuss how assessment processes were used as summative judgements to inform accountability processes and school choice. These purposes create league tables to foster competition between schools in England. Policy Influencer 1 described this process:

> The government compel schools and local authorities to publish their results… it became a kind of League Table…schools were valorised from being at the top and criticised at the bottom. (England, senior academic)

The high-stakes environment created by this competition was described by school leaders and teachers as especially harmful. Policy Influencer 9 felt particularly strongly about this:

> When you combine that with high-stakes accountability for individual teachers and for schools, then you corrupt everything. (England, professional organisations and teacher professional development)

These participants represent shared concerns amongst many participants that the political drive for accountability may produce multiple negative consequences. These consequences were further exacerbated when assessment processes were used to inform parents' school choice.

Policy influencers were concerned that encouraging parent choice had created a market-based system for schools. Policy Influencer 1, cited above, explained that the intentions behind these school-choice policies were to create a market in which schools compete for students and parents apply to individual schools based on each school's exam performance. There were multiple criticisms of this approach, especially regarding accountability, as well as the extent to which school results could reliably inform parental choice. Policy Influencer 1 highlights one particular issue:

> How good a school is now is a very poor predictor of how good a school is going to be in five years' time when your child takes their exams. (England, senior academic)

These findings suggest that effective parental choice requires access to authentic indicators of school quality drawn from multiple sources beyond high-stakes assessment data alone.

Policy influencers were not advocating against any accountability for schools, but a broadening of what informs it from just the final exam performance. Policy Influencer 6 suggested that:

> Accountability is much broader than that. It includes inspection, it includes community relationships, it includes governance. (England, academic government advisor).

Policy Influencer 9 (professional organisations and teacher professional development) recommended a reduction of the external high-stakes aspects: 'We don't need that set of external assessments… we could do some very rigorous in-house across-school internal assessment'. This suggests some policy influencers' desire to retain accountability while expanding it beyond examination results to incorporate multiple sources of evidence. Furthermore, they indicated openness to alternative assessment approaches that could reduce reliance on external high-stakes testing and move towards a broader range of quality indicators to inform accountability processes.

4.3.1.4 Formative purposes

Many policy influencers were keen to emphasise the purpose of assessment to improve teaching in classrooms. Policy Influencer 12 promoted the role of the teacher to directly instruct and correct student work. They described teachers using assessment for this process:

> The point is what happens when you say something incorrect…the skill is in how they [teachers] then correct and improve understanding. (England, Policy Advisor and Civil Servant)

This highlights the role of the teacher in using assessment to inform and improve teaching, whilst operating within a policy environment that holds teachers accountable for student results. The role of assessment processes to provide

feedback was also discussed in terms of how the teacher should improve, according to Policy Influencer 6:

> Rich questions to form a much, much greater part of oral exchange and pedagogy activity in the classroom. Providing a lot more feedback to kids, parents and teachers on how each child is thinking. (England, Academic government advisors)

This again highlights the aim of assessment to improve the role of the teacher as the deliverer and assessor of learning, through understanding students' ideas and providing feedback to them and their parents. Both contributions emphasise the role of the teacher in assessment processes, with Policy Influencer 12 particularly emphasising the teacher as having the 'correct' ideas. With both focusing on the role of the teacher, which enhances the teacher's responsibility and accountability for developing students' thinking and what they can do.

4.3.1.5 Increasing the formative purposes

When considering the possible purposes of assessment, policy influencers in England advocated greater emphasis on formative assessment practices while reducing the emphasis on summative assessment. Many influential national policy-makers in England strongly advocated an enhanced role for formative assessment, believing it would lead to better learning for students. Policy Influencer 9 strongly advocated for the integration of formative assessment in teaching:

> Assessment as being an integral part of what it means to teacher… you need to understand what they've understood and what they already know in order that you can plan what you're going to introduce next. (England, Professional organisations and teacher professional development)

This participant was keen to make clear that their conception of what it meant to teach effectively depended on using assessment to inform that teaching and direct the next steps in learning. The effects of these formative purposes were supported by Policy Influencer 22:

> I'm quite a strong believer in kind of formative assessment… you just saw the impact of that in terms of pupil learning. (England, Inspectorate, senior academic)

Policy influencer 22 emphasises the value of formative assessment in improving learning. This emphasis is notable given their role in school improvement, a domain in which high-stakes examinations have come to dominate practice. As earlier policy in influencers observed, that dominance can narrow the learning it purports to measure.

In agreement with increasing the role of assessment for formative purposes, Policy Influencer 10 criticises an overemphasis on summative accountability:

> Accountability is only interested in certain things that go on in the classroom, and it privileges those things and encourages teachers to privilege those things. It only values that which it measures, as many people have talked about, and that's where the distortion enters it… it gives only a partial view of what happens in a classroom. It gives only a partial view of the teacher and, in particular, only gives a very, very partial view of the student. It treats them or produces them as a subject of accountability. Makes them calculable and comparable, rankable and is uninterested in what is sometimes called the whole child. (Policy Influencer 10, England, senior academic)

This emphasises the current focus of assessment processes on accountability rather than formative purposes to inform learners and their teachers about what should take place next in learning. Policy Influencer 11 agreed with this perspective:

> So-called assessment is often used to satisfy the aims of a system… in that way, it's not being true to itself. (England, Head of QCA, policy advisor)

Both sources highlight that, at the national and local levels, accountability systems based on summative assessments diminish the opportunities for assessment to inform learning.

Many participants advocate for increasing the use of assessment for formative purposes while reducing the emphasis within the classroom on summative purposes. To do this, Policy Influencer 25 promoted the idea of making assessment processes more immediate and responsive:

> The ability to capture what a child's understood or not understood very quickly, in live time, in classrooms. (England, former NCSL CEO, policy advisor)

National policy influencers who work closely with teachers, such as Policy Influencer 9, highlighted that they see:

> *Assessment as being an integral part of what it means to teach.* (England, Professional organisations and teacher professional development)

These suggest a need to integrate assessment into everyday teaching practice and a move away from assessment processes operating as a separate activity. Using assessment formatively as part of day-to-day teaching ensures it underpins what it means to teach effectively, even in a policy environment that promotes more summative purposes. Yet the policy environment does not just promote summative assessment for pedagogical reasons; it also harnesses assessment data for purposes that extend beyond the classroom.

4.3.1.6 Political purposes

Another theme emerging from the interviews with policy influencers identified the role of assessment as a political tool. Rather than the assessment tools themselves being useful in political discussions, it seemed to be suggested that the grades produced via the assessment processes were particularly useful in political discussions. For example, Policy Influencer 11 suggested:

> Politicians find them [grades] useful because they can abuse them and they can compare like with like to make statements. (England, Head of QCA, Policy advisor)

This participant's deliberate use of the term 'abuse' signals a fundamental concern about the integrity of how assessment data are deployed in political discourse. Specifically, the term 'abuse' suggests that grades are being manipulated or misrepresented beyond their intended technical purpose, potentially distorting their evidential value to support predetermined political narratives. The phrase 'compare like with like' further illuminates this concern, implying that politicians engage in spurious comparisons that obscure contextual differences between educational settings, student populations, or assessment conditions. Such practices risk reducing complex educational processes to crude metrics suitable for political rhetoric. The policy advisor's statement, therefore, reveals a professional awareness that grades, whilst technically designed for

representing student achievement, become rhetorical instruments when appropriated for political purposes, potentially divorced from the evaluating contexts that give them meaning.

It is not only national assessment grades that are used for political purposes; Policy Influencer 12 highlighted the use of international assessment performance as a political tool:

> Politicians like to use international assessment data and indicators to often show the progress they're making. (England, Policy advisor and Civil Servant)

Analysed together with that of Policy Influencer 11 above, these testimonies reveal a pattern of assessment data being strategically deployed across multiple scales, national and international, for purposes of political legitimation rather than educational improvement. Policy Influencer 12's observation, while employing more neutral language, nonetheless reveals a similar pattern of selectively appropriating international metrics to narrate stories of policy success, irrespective of whether such data genuinely evidence the claimed outcomes. What renders these testimonies particularly significant is their source, both emerging from policy advisors embedded within the assessment governance system itself. Their explicit acknowledgement that politicians 'abuse' or strategically 'use' such data to 'make statements' or 'show progress' suggests a professional community that is aware that the systems they administer may be vulnerable to political appropriation.

In trying to evidence system improvement, schools are also regularly compared and expected to use exam performance in England. Policy Influencer 1 explains how the use of performance in high-stakes exams may not be the best evidence of school quality:

> The system has worked to actually maximise the scores. Those grades are no longer as useful as they used to be. (England, Senior academic)

This account articulates what is, in effect, a version of Goodhart's Law operating within the education (see literature review p. 32). The participant's language, 'maximise the scores', frames schools as strategic actors responding rationally to the incentive structures created by accountability policy, rather than as

institutions primarily oriented towards learning. Policy Influencer 10 further supports the idea of school accountability and competition informed by exam performance:

> It was intended that there would be a market within which school performance would be used as a surrogate for quality. (England, Senior academic)

This contribution highlights how exam results are currently relied upon to inform multiple stakeholders about the quality of individual schools, aiding comparisons and measures of student experience. Yet this systemic focus on what assessment results communicate about schools raises the further question of how assessment processes themselves shape students' experiences within them.

4.3.1.7 Motivational impacts of assessment processes

Although participants frequently raised potential negative effects of certain assessment processes, such as exams and other external assessments, they also raised many beneficial aspects. One aspect which demonstrated a complex tension between perceived benefits and drawbacks of assessment practice was the extent to which assessment processes may be motivating for students. Policy Influencer 12 was one such policy influencer, raising this as an advantage:

> They also have some other purposes around motivation and frankly, giving kids a bit of a kick up the backside if they need it. (England, policy advisor and civil servant)

Other policy influencers spoke less strongly but were often in agreement that students would work harder in a positive fashion when they know they will be assessed, especially with exams.

However, teachers and school leaders' accounts presented a different perspective, highlighting concerns about student well-being. Liverpool Teacher 5 had a personal reflection of assessment processes on their own child:

> I've got a son, my son's autistic... when assessment week comes around, the change in his state, his emotional state, his mental state, his ability to function at home, I often ask myself, is it worth it? (Liverpool, School middle-tier leader)

For some students, there are periods of their schooling dedicated to assessment processes which create a psychological and emotional strain. For students who may not experience success or be deemed successful in these assessed periods, Liverpool Teacher 3 describes their experience:

> …if it's not evident to them that they're making progress, that can be really damaging on their motivation, it can be really damaging on their self-esteem. (Liverpool, Early career teacher)

This evidence points to an important dichotomy in how assessment processes impacts different learners. For some students, assessment activities may be something worth studying especially hard for. Still, for others, such processes may be an insurmountable obstacle and present significant psychological barriers that potentially undermine their educational engagement and aspirations.

The evidence collected from policy influencers and teacher participants highlights how assessment processes in England encompass multiple purposes, reflecting broader tensions in education policy and practice. The dominant accountability narrative, manifested through high-stakes testing and performance metrics, exists in tension with teachers' desires to emphasise assessment's formative purpose. This dichotomy reveals a fundamental challenge in England's assessment system – how to maintain appropriate accountability for standards in education while creating space for assessment practices that genuinely support teaching and student learning. The transition observed in the case study school moving from rigid data collection toward examples of more flexible, learning-focused assessment perhaps indicates an example of how schools may shift how assessment purposes are conceptualised at the school level, even within the constraints of England's high-stakes assessment environment. This was particularly posited by participants in practical subjects, where the design of assessment processes advocated in national policy created tension with how they would choose to authentically assess learning in their subject area, as discussed later in this chapter. The accounts of policy influencers and teachers described here point to unresolved tensions between accountability mechanisms and pedagogical needs. Participants across both countries identified the absence of a shared understanding of assessment purpose as a persistent barrier to coherent practice.

4.3.2 Stakeholder perspectives on assessment processes and their impact in England

Three areas of contention emerged from participant account of assessment in England. The role of teachers in awarding qualifications, the accountability pressures attached to assessment data, and the consequences of high-stakes testing for students.

4.3.2.1 Teacher assessment for qualifications

One area of contention over the purpose of assessment processes included the role of teacher assessment as a process to award qualifications such as GCSEs or A-Levels in England. There were many reflections on how these processes had worked during the COVID-19 pandemic, where teachers felt pressured by parents and students to award certain grades for students. Policy Influencer 8 highlights the reported national grade inflation:

> I think what happened in the pandemic in England, anyway, is a good example of why we shouldn't use coursework because it [grade-awarding] kind of got out of hand. (England, senior academic)

Their description of ‘it got out of hand’ reflects their opinion on the rise in the number of higher-level passes awarded to students during the pandemic. Their quote also demonstrates a recent pressure on policy in England to reduce the amount of coursework contributing to the assessment of school qualifications in England. For policy influencers in England, there was a strong assertion that teachers face a systematic challenge in balancing different assessment demands. Policy Influencer 1 described this tension:

> The government in England realise that you probably couldn't combine assessment for formative purposes and assessment for summative purposes. (England, Senior academic, consultant)

These contributions reveal structural challenges in positioning teachers within national assessment frameworks. The reflections on COVID-19 experiences and subsequent policy directions suggest that pressures to award higher grades and accountability demands make it increasingly difficult to sustain dual teacher roles for both formative and summative purposes.

The teachers interviewed also describe pressures from interpreting large volumes of assessment data. Although they may want to learn from assessment data to improve teaching, teachers often reported being ill-equipped to do so. This was described within the context of the COVID-19 pandemic and also more generally. Liverpool Teacher 4 describes this challenge:

> So we, as a department, use summative assessments probably quite often... But sometimes I think, with the summative in my department, I'm not sure if we're actually... we're testing them and we're seeing that's the gaps... But I don't know if we're actually using that in a productive way moving forward. (Liverpool Teacher, early career teacher)

Particularly for new teachers in this school, it would appear there is such a high frequency of summative processes for assessment that teachers struggle to develop formative insights from this assessment data. Later in their interview, Liverpool Teacher 4 elaborates on this:

> Obviously, because I'm an ECT [Early Career Teacher], I'm still training, still growing, still progressing... The tracking is something that I'm trying to engage with a lot more so that I can learn more about it, the numbers, where they're going, and stuff like that. (Liverpool Teacher, early career teacher

This teacher participant has achieved qualified status but is still describing a steep learning curve that they are experiencing in the balance of producing data for summative processes and trying to use it meaningfully to improve learning. This does not seem to be just an issue with this teacher's assessment literacy and is confirmed to be the case for other newly qualified teachers by Liverpool Teacher 1:

> We've got an ECT who is starting from scratch - doesn't know how to mark, doesn't know how to assess. The examination course isn't until November, so automatically, you know, she's teaching a course that she's never been … learned how to assess on for three months; no actual training on it. (Liverpool Teacher, School Leader)

This teacher participant is suggesting a deficiency in the training received by newly qualified teachers. This means they perceive themselves as not having the ability to use assessment processes effectively, and this is corroborated by their more experienced colleagues. These teachers may not have trained in the same

institution, which reveals a potential issue with assessment literacy beyond this school or the Northwest of England.

Policy influencers also raised concerns about teachers' abilities to use assessment to gauge learning. Policy Influencer 9 provided some historical context for what may have created these systemic issues:

> We then got a new Chief Inspector of Schools, Amanda Spielman, who was somebody who understood data. And what had been going on, of course, was that there had been this whole industry with teachers who didn't readily understand statistics, didn't readily understand the impact of data, working towards other people who are trying to hold them to account, and they were people that didn't understand it either. (Liverpool, Professional organisations and teacher professional development)

This participant is describing previous accountability systems under which Ofsted, the school inspectorate, would request assessment data from schools to evidence student progress and make judgements based upon what was provided. They then describe the creation of systems to create metrics and statistics that were not widely understood, according to this participant, but were produced to maintain a system of accountability. Other participants then described how improving assessment literacy in England's teaching profession might help mitigate some of these complex pressures from summative assessment processes serving accountability in England that lead to misunderstanding of assessment or, as Liverpool Teacher 7 described it, 'fudging' internal assessment results. Liverpool Teacher 6 made such a suggestion for improvement:

> We need to improve assessment literacy massively... I think you can close the gap between the adverse impacts of accountability and high-quality education by establishing good professional development in respect to assessment literacy. (Liverpool, school middle-tier leader)

This call for enhanced assessment literacy from teacher participants resonated with perspectives articulated at the national policy influencer level, yet the specific role of teachers in high-stakes qualification assessment remained contested. Views on teacher involvement in assessment for qualification diverged sharply. Some participants argued that strengthening assessment capacity could equip teachers for a greater role. Others contended that

structural accountability pressures made such involvement untenable, regardless of the professional development offered.

Analysis of interview data revealed divergent positions amongst policy influencers regarding the appropriate role of teachers in high-stakes assessment for qualification purposes. These positions appeared to centre on fundamentally different assumptions about the relationship between accountability, equity and assessment practice.

One policy influencer put forward an argument that teacher involvement in summative assessment should remain restricted due to concerns about bias and inequitable access to coaching:

> There are strong proponents of the English Education Assessment system, which are about lack of bias, that the disadvantaged kids will do worse if you had a different system because the advantage kids would get coached to do any teacher assessment in advance whereas the disadvantaged kids were less likely to get a coach so and also with biased against disadvantaged kids, because of that they're strongly proposing that the best system is written assessment. (England, former NCSL CEO, policy advisor)

In contrast, other policy actors emphasised the potential for enhanced assessment literacy to address these tensions. One policy advisor suggested that regular professional development could reconcile accountability pressures with high-quality educational practice:

> We need to improve assessment literacy massively. We do a lot of work here, actually out of Cambridge, in terms of Cambridge Assessment Network by [name removed] and we got huge interest from teachers, you know, it's just escalating, actually post pandemic. I think you can close the gap between the adverse impacts of accountability and high quality education by establishing good professional development in respect to assessment literacy. Yeah. Absolutely fundamental. (England, academic government advisor)

These contrasting positions reflect broader debates within the literature regarding the feasibility of teacher-mediated assessment for qualification purposes in high-stakes contexts (Sturrock, 2024).

The significance of assessment capacity challenges identified in policy discourse was substantiated by empirical evidence from the Liverpool school case study. Teacher and school leaders felt they needed further professional development in assessment and described their current understanding as the result of informal, on-the-job learning that had taken many years of trial and error. Their accounts revealed a tension between two competing demands. Teachers were expected to serve as assessors for internal accountability purposes and, at the same time, to interpret assessment data in ways that improved their teaching and their students' learning.

4.3.2.2 The impact of accountability mechanisms

As indicated earlier in this section, participants frequently discussed the use of assessment processes to produce data that is then used to hold people and institutions accountable. This analysis focused on teacher assessment functions and the importance of teachers understanding this data for their role as assessors. However, the interview participants also described the use of assessment data by multiple stakeholders both within and outside of the education system to hold individuals and institutions accountable.

Accountability in general was mentioned by teachers, school leaders and policy influencers as a fundamental aspect of securing improvements in education. It was often mentioned that in England, tensions are caused by the structure of accountability, which is heavily informed by assessment data. Policy Influencer 25 articulated this tension:

> I think accountability is really important in education. I don't think any government could invest the amount of money it invests without knowing that that money was well spent... but I think the emphasis has been too much on academic excellence and not enough on some other aspects of education. (England, former NCSL CEO, policy advisor)

This quote is representative of the sentiment of many participants who described a system where accountability is essential, but the current format is heavily informed by external examination data. While some of the negative effects of these accountability pressures were raised in the literature review,

the extent of their impact on teachers' sense of accountability emerged as a significant theme from the case school, warranting further analysis.

Liverpool Teacher 4 described their perspective on the relevance of external accountability to their school:

> So if we start off with external, it's a top-down approach. Massive neoliberalism. Ultimately, the school has pressure to get good grades. Good grades mean funding. Funding means better education for schools, and ultimately if we don't get a good Progress 8 or Attainment 8, the school are accountable. (Liverpool, Early career teacher)

This participant's explicit naming of 'massive neoliberalism' is analytically significant. An early career teacher deploying this ideological vocabulary suggests that the accountability system is not merely experienced as institutional pressure but is understood by some practitioners as an expression of a particular political logic. Their capacity to articulate the funding-grades-accountability chain reveals an awareness and developing assessment identity, already shaped by the performativity system they have entered.

Not all teachers and school leaders in the school explained the mechanisms with detail of metrics such as Progress 8 data scores from external examinations, but all described feeling a sense of collective importance of positive data for the school's, and potentially their own, future. Some participants in the case study school felt that some senior school leaders, not necessarily in this school, put pressure on them to ensure assessment data were favourable. This was especially demonstrated by Liverpool Teacher 7:

> I think there's definitely legacy issues in the senior leadership team, you have legacy issues with staff whether that's faculty leaders or teachers because for 10-15 year period, it was drilled into them how powerful this data are and about how that we needed to collect this data because we needed it and if we didn't have it, it could cost us our jobs when Ofsted came and all those type of things. Even when you have meetings with the local authorities, the local authority - a number of those are colleagues who have not worked in schools for a number of years and so they still kind of talk about what is your Progress 8 Score looking like currently for year 11? They are wanting kind of predicted grades and those type of things. Trustees and Governors again, because there's so much training gone on previously

> in terms of “oh well we want to see what this data looks like”. (Liverpool, school leader)

This is notable because this participant is a school leader themselves, yet they describe difficulty in changing the culture in their school and note difficulties in doing this with changes in accountability mechanisms. Their descriptions identify many stakeholders who may struggle to understand new accountability metrics or find it difficult to avoid using measures that were previously given very high priority. The school professionals have formed assessment identities that are now in conflict with the information the headteacher is providing them.

This excerpt illustrates the difficulty of communicating a student's progress toward an achievement goal when a single grade on a report cannot capture the trajectory of improvement. One insightful quote, again from Liverpool Teacher 4, illustrates the complexity derived from an increased use of data in communication with parents:

> If they are a Band B but they've got a B-, it just says they're underachieving, and it flags them up as like an orange colour on the report hub, and the parents are immediately like “oh, you're underachieving.” And then there's been a little bit of stigma around that, and parents and kids as well have been a little bit almost distressed. But in reality they might have had a target grade of an 8, and they got a 7 in the assessment. (Liverpool, early career teacher)

This excerpt provides an insight into the difficulty of communicating with students who are making progress towards an achievement goal. The use of the word ‘distressed’ also suggests that negative effects may be produced for parents and students, not just from the accountability process itself but also through a misunderstanding of what the data may demonstrate about the student’s learning. Liverpool Teacher 4 comfortably uses ideas such as target grades, numerical scores, colour ‘flags’, and letter grades as diverse proxies for achievement, which may be difficult for parents and external stakeholders to fully comprehend. This perhaps demonstrates a high potential for confusion in the education system when reporting the results of assessment processes.

4.3.2.3 The levels of data literacy

The general understanding of data by stakeholders in education, especially when used for accountability processes, was frequently raised as an issue. Policy Influencer 22 identified a fundamental issue with data interpretation:

> Data literacy in schools and in local authorities... is decidedly a mixed picture... people do take data and run with it and say, look at say, for example, value-added data or something and don't know that that's a model and there is a confidence interval around that, and that's not an absolute number. (England, inspectorate, senior academic)

These ideas present a sense from a former school inspector that there may be a propensity for assessment data to be used with little understanding or true ability to interpret it. This was a common consensus among policy influencers, despite agreement on the importance of being able to competently use data as described by Policy Influencer 23:

> You got to understand data, and so many people don't have a sophisticated understanding of it. (England, trade union representatives)

Policy influencers, school leaders and teachers were all unable to explain why the meaning of assessment data remained persistently unclear across the system. None could account for how the lack of clarity for any stakeholders had become embedded as part of the system design.

Policy influencers recommended improvements to be made to the public's ability to interpret education data so they can make informed choices for their children. Policy influencer 18 described what they see as a part of the solution:

> I think you need a far more sophisticated approach to the publication of data. And part of that is to try to actually help the population become intelligent consumers of education. (England, Policy advisor and civil servant)

Liverpool Teacher 7 described efforts made in the case study school to try and improve parent understanding of assessment data:

> We've then put on webinars for parents to try and explain why we've moved away from this now; that you're tracking reports going to say

> what percentage they've achieved in that assessment, how that compares to where we feel they should be achieving. (Liverpool, school leader)

These improvements demonstrate the case study school's willingness to effectively explain assessment data to parents and to make changes to the data it provides. The changes in assessment measures and data produced are discussed later in the context of policy changes, but it is important to note here that continual external and internal policy changes mean that all school stakeholders require ongoing updates on how to interpret the assessment data. Revised reporting formats also placed demands on the school's relationships with external stakeholders. School leaders described needing to train parents and governors to interpret assessment information that no longer relied on National Curriculum levels. Without that training, external stakeholders lacked the fluency to engage meaningfully with pupil progress data, limiting its value as a basis for dialogue about learning. The result is continued confusion and concern about learners' progress, as described by the school participants.

4.3.2.4 Impacts of increasing assessment demand on students

The effects of assessment processes having such a focus on data and the production of metrics were described as having a particular impact on students. Both teacher participants and policy influencers noted several negative effects on students' mental health and learning, with teacher participants reflecting on specific observations made within the case study school. Both groups discussed the mental health impact on students completing numerous exams as part of the assessment process. Liverpool Teacher 1 explained their perspective on these effects:

> I can see the students here being literally pulled left, right and centre into boost and secures and to mock exams and more mock exams. I would say it does have a negative effect on probably students' mental health, particularly after COVID as well. (Liverpool, School middle-tier leader)

This teacher is describing the use of exam papers completed in the same conditions as external examinations completed at the end of schooling, such as GCSEs and A-levels, but marked by teachers to track student progress and produce predicted grades. In the Liverpool case study school, the use of mock

exams was aimed to be beneficial to students who may otherwise have been nervous about taking exams in such conditions and provide data for subject leaders to drive interventions with groups of students who may need further lessons after normal school time. This extra support may be appreciated and valued by some students, but others may feel extra pressure from having to attend further lessons after school as a result of disappointing mock exam results.

Chasing certain grades by both students and their teachers were described as having negative effects on learning. Liverpool Teacher 3 described students losing a love for learning:

> It [school focus on grades] can sometimes feel like it's stripping away the passion of some of those subjects, the enjoyment, the emotion that you can get from reading a good text.

They also added that this may be detrimental to feedback processes:

> There's this absolute kind of focus on the numerical result for a piece of writing. The reams and reams of feedback that you can jot for hours on end in a pupil's book almost become invisible as they look at the grade. (Liverpool, School middle-tier leader)

This teacher’s account reveals a perceived zero-sum relationship between the affective dimensions of learning and the summative demands of the assessment system. When teachers described feedback becoming ‘invisible’ once a grade was attached, they identified a recurring pattern. Pupils filtered assessment information through the evaluative frame that surrounded it, attending to the mark and discarding the commentary. Feedback designed to support learning was, in practice, read as a judgement of performance. This account provides empirical evidence for the displacement of formative by summative purposes at the level of student reception, extending the pattern already identified among practitioners.

Policy influencers tended to discuss effects on students more abstractly; however, Policy Influencer 10 described the limited opportunities and exposure to stress posed by the exam system:

> The student's experience is impoverished and narrowed. Their capabilities are underdeveloped. And at the same time, they are increasingly subject to stress...Our school students are among the most stressed of any in Europe. (England, senior academic)

Policy Influencer 10 highlights that concern for the stressful impact of exams is reaching the level of policy influencers, but it isn't quite clear why these negative impacts are discussed less frequently.

Interestingly, there was a striking lack of reference to the positive effects of assessment processes for students in England. The formative purpose, as discussed previously, was one of the few positive outcomes, often described as a useful tool for teachers to provide their expert feedback and inform others what should happen next. Policy Influencer 6 described this process as providing:

> Feedback to kids, parents and teachers on how each child is thinking and what they're doing. (England, academic government advisor)

Many policy influencers described the formative use of assessment processes as teachers acting as experts to inform their students what to do next. In England, this was emphasised in terms of the teacher's role, excluding other aspects of formative practice such as activating students as learning resources for themselves and each other.

The accounts in this section point to a reinforcing cycle of problems with England's assessment processes. Weak assessment literacy limited teachers' capacity to contribute meaningfully to external qualifications and reduced the credibility of the data that accountability mechanisms relied upon. That data were widely regarded by both teachers and policy influencers relied upon. That data were widely regarded by both teachers and policy influencers as lacking validity, while its complexity prevented parents, governors, and other stakeholders from interpreting it. Proving resistant to change, accountability mechanisms attracted high public visibility and directly affected school leaders' career trajectories, making reform politically difficult. Teachers, meanwhile, system that generates acute stress with few compensating formative benefits. The section that follows traces these effects back to the specific policy decisions that participants identified as their source.

4.3.3 To what extent does assessment policy impact practice in England?

Policy influencers commonly identified England's assessment policy as producing problematic effects, attributing these largely to its overemphasis on data protection and use. The previous section examined how assessment processes affected students and teachers. This section turns to how Policy Influencers traced those effects to specific policy decisions and reforms. Their accounts aligned with those of teacher participants on several points, most notably concerns about the volume of data generated by assessment and its consequences for learning and accountability.

4.3.3.1 Policy to increase the use of assessment for formative purposes

One policy strand, the sustained drive to embed assessment for formative purposes, drew detailed reflection from policy influencers about its effects. Policy influencers indicated that many policy processes had attempted to embed practice using assessment for formative purposes, with limited success. Policy Influencer 1 considered government efforts to combine formative and summative purposes a failure:

> The government in England realises that you probably couldn't combine assessment for formative and summative purposes... this is a fundamental tension, and I think the government is now pretty much given up. So they're basically saying to schools, just sort yourselves out, and we'll test you with these performance targets. (England, senior academic)

This participant was very critical of any progress that had been made to date on reforming classroom assessment, summarising their answer with:

> I think you could make a very reasonable argument that there have been no positive effects on classroom assessment practice of any government initiatives, simply because they got so much of it wrong. (England, senior academic)

These contributions suggest that despite multiple policy initiatives, they have faced significant challenges in their enactment. Taken together, they provide further evidence of a widely perceived incompatibility between formative and summative purposes of assessment due to multiple challenges in the education

system and an apparent government retreat from direct intervention in some areas.

The difficulty and frustration with system-level barriers have not resulted from a lack of desire or effort on the part of policy influencers. Policy Influencer 18 described their efforts:

> I tried really hard with the QCA to put more emphasis on assessment for learning or learning skills and moderated Teacher Assessment because I felt it was those sort of policy levers was a way of sort of reconciling the concerns. (England, academic government advisor)

This policy influencer suggests that, so far, attempts to limit teacher assessment and shift the focus of their role to formative purposes of assessment have been difficult to achieve.

Within this national policy field, teachers in the Liverpool case study described a parallel displacement of formative work by summative routines. Teacher 6 located this displacement in their own department practice:

> So we, as a department use summative assessments probably quite often. I would feel probably too much because as a department we do an exam style assessment at the end of every half term which we would then use to check the knowledge of what they've acquired in that unit of work... So, we're using the summative within our department in my opinion, probably a bit too frequent, but also not actually using the information it gives us other than to put into the tracking and then would go back to the parents. (Liverpool, Experienced Teacher).

Two strands of critique sit within this account. The first concerns volume. An exam-style assessment at the end of every half-term registers, in the teacher's own framing, as routine over-testing rather than formative use. The second concerns purpose. The resulting data feeds tracking systems and parental reports, but does not return to the classroom as feedback to inform teaching. Assessment information, therefore, circulates outward and upward, but not back into pedagogy. The account illustrates how a summative rhythm can crowd out formative intent even where school policy advocates the latter, with the data produced serving institutional and communicative ends rather than instructional ones.

Other teacher participants felt that there had been a shift within the school toward adopting national discourses on what was considered knowledge worth teaching and assessing. Practical subjects especially describe tensions between a focus on exams used to measure student understanding, and what matters in their subject. Liverpool Teacher 5 described this:

> I don't think we're doing much around the assessment of the practical side now because of the way our assessment policies in school are being directed. (Liverpool, school middle-tier leader)

The value of their teacher judgement is described tentatively against that of objective tests:

> We do assess and we do include some elements of the practical work...But generally speaking, what's reported home is the score they've achieved on the test. As a reporting measure, we can include elements of the practical work if we believe it's of consistent value. (Liverpool, school middle-tier leader)

There is a tension created here between what the teachers may value in their subject, practical ability, and what is able to be assessed and reported to parents through assessment tools that are available. Liverpool Teacher 5 again summarises their frustration with this:

> Sometimes there's a little bit of almost like a lost element of learning there because sometimes a student might not remember all the knowledge from factual recall from a test...but actually, they performed really well in the practical elements. (Liverpool, school middle-tier leader)

Teachers also expressed frustration at being unable to assess in ways that felt meaningful to them. For these participants, meaningful assessment aligned closely with the curriculum's intent, yet their desire to 'mirror' the final examinations pulled practice in the opposite direction. Liverpool Teacher 4 describes their desire to change assessments to be more authentic for their subject:

> I think there's work to be done on the assessments that we currently give in food technology. I think sometimes a lot of it's very surface level. I do think the practical one we do at the end of the year is a good measure of assessment for what it means to be good practically

> at food tech. I'd probably like to see that half-termly, probably. (Liverpool, Early career teacher)

When asked about their motivations for this, Liverpool Teacher 5 (Experienced Classroom Teacher), also in a practical subject, explains,

> …because it doesn't mirror what we're doing at GCSE. We often find we're starting again. It will mirror exactly what Maths needs and what English needs, but it doesn't work for us. (Liverpool, school middle-tier leader)

Teacher 5's account suggests that the design of internal assessment in practical subjects is shaped less by subject-specific pedagogy than by the demands of high-stakes GCSEs. The desire to 'mirror' GCSE requirements points to a washback effect in which external examinations dictate the structure of lower-stakes assessments. Whether these teachers also seek greater scope for practical assessment on its own terms remains unclear, because the gravitational pull of GCSE accountability dominates how they describe their concerns.

4.3.3.2 Policy to promote fairness in assessment

Further discussion of the impact of assessment policy focused on issues of fairness. The evidence from interviews suggests that neither national policy influencers nor teacher participants in the Liverpool case study school believe the recent assessment policy is promoting fair opportunities for students. Despite all participants acknowledging that fairness is commonly stated as an aim of changes to assessment processes, these goals are not currently being met. Policy influencers focused their concerns on the wider system construction and aspects such as comparable outcomes as a form of norm referencing for national assessments.

In discussing the challenges of current awarding procedures for GCSE grades in England, Policy Influencer 23 explained:

> Within the year, they [Ofqual] norm reference the results, and you're guaranteed that roughly speaking, 1/3 of GCSE students doing English and Maths get a grade that is regarded as failing...And we tell 1/3 of the children in this country who speak English as their first language that they have failed in English, and I think that is disgusting. (England, trade union representative)

This policy influencer clearly felt that the current system administered by Ofqual, which ensures that similar results are awarded each year nationally, is unfair and ‘disgusting’. The perceived inevitability that a predetermined proportion of students nationally will not achieve a GCSE grade considered the pass threshold for a passing grade was a difficult aspect of the current national assessment policy that many participants found unfair. These methods, known as norm referencing, are used to ensure comparable outcomes across cohorts. Proponents argue that norm referencing provides a necessary safeguard against grade inflation, maintaining the value and credibility of qualifications over time. However, participants in this study perceived norm referencing as less fair than criterion referencing, where students can meet a grade threshold by demonstrating specified knowledge or skills regardless of how other candidates perform.

At the middle-tier local authority level, there were implications of the comparable outcomes policy that may be felt by many stakeholders. Policy Influencer 11’s account particularly illustrated these:

> Once you realise that, actually, the only way my children can succeed is if others do less well, you're bound to be disillusioned. (England, head of QCA, policy advisor)

This contribution reflects two imperatives for this policy influencer. Firstly, the commitment to enhancing educational outcomes for learners in their care, Secondly, the pragmatic need to maintain competitive advantage in an increasingly marketised policy environment. These aspirations, however, exist in tension with England’s comparable outcomes policy. Under this model, Ofqual anchors the national grade distribution to prior attainment within each cohort, effectively fixing the proportion of candidates awarded each grade in any given year. One institution can therefore improve its examination results only if other institutions’ results decline.

Some policy influencers advocated for a change in the comparable outcomes policy. Policy Influencer 11 described why this was needed:

> Sadly, not enough teachers understand the way the testing system is arranged. They still believe it's criterion reference when it's

> not...Political interference is likely to lead to massive disillusionment in teaching. (England, head of QCA, policy advisor)

This statement provides insight into the perceived lack of understanding by many teachers of how grades are awarded. They fear that other stakeholders would be disillusioned should they realise that improvements in outcomes are moderated nationally, something that is not widely discussed in public discourse or in the design of accountability mechanisms. Experienced teachers in the Liverpool case study, however, appeared to understand this. Liverpool Teacher 7 described their understanding of the current policy:

> 72% of your year group across the country are going to leave without a Grade 4 in Maths and that you're going to have 72% are going to leave without a Grade 4 in English. (Liverpool, school leader)

This participant clearly understands the implications of this for their students. Policy Influencer 25 advocated for a cautious approach to changes:

> I'm a bigger fan of Criterion referencing, but I also strongly believe that they have to have really strong moderation; otherwise, you'll end up with grade inflation, which is what happened at universities. (England, former NCSL CEO, policy advisor)

This excerpt supports changing grades to reflect students meeting certain standards, but cautions that this could lead to higher rates of top grades due to accountability pressure. The possibility of changing this policy and tensions with other accountability policies were also described by the Head Teacher of the case study school:

> I remember when I first taught Maths, what coursework was like and how it could be fudged and then the pressure then on teachers then to do things with the coursework or to make students do things to the course which is just in that grey area of whether it's right or wrong. (Liverpool school leader)

There appears to be agreement among many leaders in education that, under the current accountability system, there would be too much pressure to award students grades that exceed their merit. The use of the word ‘fudged’ suggests a judgement that this process was not sufficiently rigorous for the awarding of national qualifications.

Analysing teacher participants' perspectives on fairness within the Case Study School revealed tensions between contemporary assessment policy frameworks and their use as measures of school effectiveness. Liverpool Teacher 7 discussed the socioeconomic factors that made these judgements unfair from their perspective:

> I've only ever worked in highly deprived areas. Now the framework has always been unfair on me when it used to be just on outcomes...How many of your students were getting 5A*-C's, including English and Maths. When I worked in one of the most deprived areas in Wakefield in the country, that framework was never fair on me then. (Liverpool, school leader)

This school leader felt that they were being judged unfairly as their students faced many more challenges due to socioeconomic disadvantage than students in other schools. This meant that they would be judged on the same measures as teachers who worked with students who did not face these barriers and perhaps had access to resources that would advantage them in preparation for exams.

It was common for teacher participants to feel as though their lived experiences were detached from policy designers and that politicians would not understand the challenges faced by the school to try to ensure their students were able to achieve. Liverpool Teacher 5 described their views on this:

> We've got people sitting around the top end, and I'm talking at government level now, who are making the most bizarre decisions around education and particularly around assessing...I dare to think when was the last time they actually come to a comprehensive school in like a traditional comprehensive school in Liverpool and sat with a cohort of 44% Pupil Premium students. (Liverpool school middle-tier)

In this, they discuss the community of students they serve with a high level of disadvantage, as reflected in the 44% Pupil Premium measure, and how they feel policy design is not currently made with the circumstances of either pupils or teachers in mind.

Both policy influencers and teachers in the case study schools were asked whether they thought the current exam system in England was fair. There were mixed views on this, and some participants focused on how assessment through examinations particularly disadvantaged students from economically under-

resourced backgrounds. Policy influencers, such as Policy Influencer 25, described competing arguments around this:

> There are strong proponents of the English Education Assessment system, which are about lack of bias, that the disadvantaged kids will do worse if you had a different system because the advantage kids would get coached to do any teacher assessment in advance whereas the disadvantaged kids were less likely to get a coach...they're strongly proposing that the best system is written assessment. (England, former NCSL CEO, policy advisor)

This contribution describes both sides of the argument, as commonly presented by participants, in which exams are seen as an imperfect yet best way to assess students' learning at the end of their courses. Although some students will not be able to prepare for exam demands in the same way as those with certain advantages, this ensures that the work being assessed is produced solely by the student. Although this contradicted issues some participants had raised about the validity of exams for practical subjects, the concern about ensuring a robust system to prevent potential cheating was a key priority.

Teachers in the Liverpool case study school raised many barriers and complications within the system. These barriers include the detrimental aspects of high-stakes exams and the previously raised questions about norm-referenced exam results in this work. However, they also question whether some students may not be able to perform equally well in exams due to aspects of the assessment that challenge students in their understanding beyond the construct being assessed. Liverpool Teacher 7 explains:

> I had a student, really bright mathematician but came from a poor background. His vocab was limited, and it was a value money question around tea, coffee and around the use of an urn. And he had no idea what an urn was... we should be testing on the mathematical knowledge not on that. (Liverpool, school leader)

This teacher participant was concerned that their students might not perform as well as other students who lacked knowledge relevant to the context discussed in some exam questions. This teacher participant suggests that this is a problem in maths, which an English teacher in the school considers to be a learning problem for such students. Liverpool Teacher 2 discussed the vocabulary barriers for some students in their English exams:

> In our school in particular, there's a real Maths English divide and lots of our students prefer maths...because it's quick wins, and with English they really struggle because of their lack of cultural knowledge, their lack of vocabulary. They really, really struggle to access it sometimes. I've got students in my Year 11 class with a reading age of 13, and the text is supposed to be 16. (Liverpool, school middle-tier leader)

Whilst levels of literacy and vocabulary might be the types of knowledge an English exam aims to assess, this teacher describes their students as at a disadvantage due to the broader demands of knowledge and the ways in which they are questioned. In this way, the assessment appears to incorporate broader aspects of literacy and background knowledge beyond the key knowledge and skills claimed to be assessed, potentially raising concerns about its validity.

4.3.3.3 Off-rolling to meet the demands of policy change

A final concern from teacher participants that relates to previously described accountability uses of exams is off-rolling. Policy Influencer 22 describes this practice from some schools:

> It is a tendency to try and make certain pupils disappear from the assessments, so we need to make sure that we have mechanisms in place that ensure that doesn't happen. (England, inspectorate, senior academic)

The language of making students ‘disappear from the assessments’ unveils the extent to which accountability metrics have inverted the relationship between schools and their stated purpose. When institutional survival is dependent upon aggregate test data, individual students become liabilities to be managed rather than learners to be educated. The head teacher of the case study school describes this as being particularly problematic for students in the exam preparation years:

> That's where you see off rolling and that's where you see schools trying to kind of persuade the students to leave, to go to another school because they don't want them to impact on their exam results. And then there's that level of suspicion between schools when there are transfers between schools, particularly in Key Stage 4. (Liverpool school leader)

Schools serving disadvantaged communities may feel no choice but to opt for these practices to maintain their exam results and reputation. Participant 25 suggested this as a possible enhanced pressure:

> It's going to create lots of stress, even for successful schools to create stress, but if the odds are stacked against you, it's even more stressful. (England, trade union representative)

The pressure for schools to ensure their exam cohorts consist of students with the highest likelihood of success can also be seen before students enter the school. Liverpool Teacher 7 (Head Teacher) describes how students identified as having SEND needs may be discouraged:

> We're incredibly inclusive - have a really large number of SEND, for example, but if I wanted to improve my schools Progress 8, a strategy would be me persuading parents with children who got SEND not send their children to my school, and I mean how wrong is that? I absolutely refuse to do that. But you know, there's some heads who probably feel that their job is on the line and some trusts - would they ever admit to it openly? No. But we know that's how they operate. (Liverpool, school leader)

The head teacher clearly positions themselves as refusing to engage in these practices but acknowledges the pressure driven by accountability and competition between schools mediated through examinations.

So far, this section has highlighted how assessment policy in England has a strong focus on data collection and accountability. This is putting pressure to make changes that demonstrate improvement in this data, without necessarily advancing teaching or learning. In some cases, it changed schools' behaviour detrimentally. To permit changes to these policies, enhancing positive effects and mitigating negative implications, many participants argued that there was first a need for cultural changes.

4.3.3.4 Policy change inertia presented by cultural factors

Some of the cultural changes proposed by both policy influencers and teachers centred on a required shift in society's view of what education means and what is expected of an education system. For example, within assessment systems, participants questioned the role of national examinations and the grades they

produce. Policy Influencer 9 made an observation about England's cultural attitudes:

> When you look at our society compared with other nations, you notice that, for example, we love lists, we love cues, we love ranking. We like to know where people are and we have this whole thing about, don't get above yourself... We've got a very kind of rigid class system, although we don't like to talk about it... all that stuff is about keeping you in your place. (England, professional organisations and teacher professional development)

This policy influencer's framing reveals deeply embedded assumptions about assessment's social function. The reference to 'keeping you in your place' positions assessment processes not as pedagogical tools, but as mechanisms of social reproduction. This language suggests that the policy architecture rests upon and reinforces a stratified cultural logic predating and transcending any individual policy intervention.

The aspects of culture were also described by teacher participants as expectations communicated by parents. When trying to adapt school policy to focus assessment on wider aspects of learning and feedback, and to move away from grades, they say this does not meet parents' expectations. Liverpool Teacher 7 describes the challenge of these expectations:

> Parents have found it difficult. Parents from the start of year seven want to know what their child's going to get at the end of year 11. They want to know if they're going to leave with the top grades... Parents have found it really difficult, particularly those parents who are very kind of academically minded. (Liverpool, school leader)

The participant's framing of 'academically minded' parents as those fixated on grades reveals an exclusion between academic engagement and grade attainment. This distinction matters as the account suggests that parental expectations operate as a constraint on assessment reform, reinforcing the grade-centric culture that Policy influencer 9 identified on page 184. The convergence between these accounts indicates that the demand for grades is not solely a top-down policy imposition but is co-produced through parental expectations shaped by decades of cultural normalisation.

People who have emerged through an established exam system, such as that in England, may expect it to continue to operate in the same way as they experienced, regardless of how fair they found it to be. Such is the power of the cultural expectations for this system to be robust. Policy Influencer 11 describes how this operates:

> Parents trust the exam results... If they did poorly in life, but well at school, they think it was their fault? They don't think it was a school's fault because they did well at school... So, they always believed the exam system was right. Because there's something about the notion of an authoritative grade telling you where you are that keeps you in your place. (England, head of QCA, policy advisor)

This statement further supports the idea that the exam system forms part of a wider culture in England, where it is used to sort people into a correct 'place' within society.

The interview participants also described how the exam system is used politically as a proxy for maintaining educational standards. Policy Influencer 25 describes how attempts to change this would be difficult:

> It's hard for a Secretary of State or Cabinet Secretary to stand up in Parliament and defend a move away from grades because they'd be accused of reducing accountability and going soft on standards and that's hard politically. You got to be really strong politically to be able to make that call. (England, former NCSL CEO, policy advisor)

From this participant's perspective, attempts to change the influence of exams and grades would be difficult, as they serve as proxies for educational standards. Such is the power of these measures that a move away, even if to a more diverse and comprehensive system, may draw accusations of lowering standards. Policy influencers understood that changing assessment policy away from the current systems could be difficult, as it might attract criticism from the opposition, which may be why this has remained a constant theme since the inception of national examinations.

Overall, these contributions suggest that frequent changes to assessment policy may have had far less impact on how assessment processes take place in England than their stated ambitions. The evidence from interview participants suggests this may be due to embedded cultural attitudes that rank, measure, and judge

success primarily through examination results. Policy influencers suggest that attempts to change how assessment processes are used would also incur political criticism, further disincentivise radical change and perhaps limit the extent to which policy change has succeeded in altering how assessment processes are used in schools. Despite this, political discourse often includes assessments and data, and politicians may use them as political tools.

4.3.4 In what ways does political thinking impact assessment policymaking in England?

Political thinking can be communicated to the education field in a number of different ways, and the participants referred to many examples of these during the interviews. The findings from the interviews suggest that political priorities significantly shape assessment policy through accountability mechanisms and funding pressures, often creating tensions between policy intentions and their enactment in schools.

4.3.4.1 Opacity in Political Thinking and Communication

The participants describe a varied picture of policy creation, communication and enactment. Policy Influencer 18 described the complexity of creating a sustained direction of travel in policy with political changes:

> There was a sort of a shift in the politics. You know, the Iraq War destabilised things, the interface with Brown and Blair became more destructive. David Miliband left… And then certainly when the Tory party took over in 2010, you know, it was as if none of what we were developing had happened, and we went right back to pretty didactic. (England, academic government advisor)

This relatively short excerpt displays some of the complex political influences at work during the latter New Labour years and the transition to the Conservative and Liberal Democrat coalition in 2010. They described this transition and the interruptions in the later New Labour years as being obstructive to real systemic change:

> You know, systems that are really effective have to stay with it for quite a period of time. You can get a change in a relatively short period of time, like within three years, right? I'm talking system-wise now, but you can't get sustained change unless you keep at it for, say,

> you know, up to a 10-year time horizon. (England, academic government advisor)

Ultimately, this participant described the changing nature of politics in England, where changes in government are more frequent than they assert is necessary for long-term policy creation and enactment.

4.3.4.2 Politicians' Selective Use of International Assessments

As well as frequent changes in the political parties controlling education systems, international assessments were frequently cited as influencing political decisions. These comparisons were named as drivers for political agendas. Policy Influencer 22 describes such policy drivers from PISA:

> I think, you know, PISA is probably the best example in terms of internationally. Every time the PISA data comes out, how everybody dashes to the country that's doing particularly well and tries to work out what it is they do. And, of course, they've already missed the boat by the time they get there because it's probably what they did 10 years ago that actually makes a difference. (England, inspectorate, senior academic)

This participant is highlighting the influence of these international assessments on political direction and how this can lead to rapid attempts to adopt policies internationally.

There were many mentions of policy borrowing, but some participants felt that PISA didn't always push politicians in one particular direction. Policy Influencer 21 describes how politicians selectively use international comparisons:

> So there's no doubt that those international comparisons have been used to push arguments more or less, sensible arguments. But the fact is that nobody knows why countries are successful...if you believed in the importance of rigorous ability grouping, you'd pick Singapore; if you believe in the importance of mixed ability, you pick Finland...It's so there's enough variation here for people to be able to pick a model and be able to argue from the evidence for that model. (Scotland, education leadership consultant)

This participant highlights how useful international assessment results can be for politicians seeking to justify a change in policy direction.

Both participants highlighted the difficulty of using international assessments to direct policy change. When politicians take the results of these comparisons,

they may filter the policy implications and select only results that confirm their preconceptions about what should be done to improve educational standards. They may, however, be copying policies that have not yet taken effect in the results and ignoring past policies that have put the country ahead in the assessment rankings. Policy Influencer 11 was most critical of this use of international comparisons for political purposes:

> The most blatant use or misuse of the exams is the way the PISA results are used... we compare spurious examples between different nations to demonstrate the effectiveness of a system so that we can promote the policies we want to promote. (England, senior academic)

This participant highlights how exam data may be used when politically expedient, even when this may not align with the overall picture of a country's performance.

4.3.4.3 Temporal Tensions in Policy Enactment Between Political Cycles and Embedding Reform

When considering the role of government and policymakers in the design of assessment processes, policy influencers discussed longer-term strategic needs. Policy Influencer 12 felt that the government should develop and train the profession by telling them how best to use assessment. They explained:

> The thing that government got wrong, which hopefully they're attuned to now, is understanding that you need to train and support people better to put better things in place. Part of that is the training professional development I talk about, but part of it as well, is showing models and exemplars of what great and good looks like so people don't have to reinvent the wheel and come up with something really low-quality because reinventing the wheel is stressful, time-consuming, and you know sadly often not particularly high quality and so I think government needs to do much better at providing exemplification of different aspects of school life, and that includes in assessment. (England, policy advisor and civil servant)

This participant had taught in the classroom for 2-3 years and explained how frustrated they felt at not being told exactly what would work for their assessment and teaching practice. When they moved from their brief classroom experience to an advisory role in the Department for Education, they were keen to change the need for time-consuming development. Rather than having to work and evaluate this, they wanted the government to solve potential workload

implications from this process by directing what should be done. They linked this to recent DfE policies with high levels of centralised control across many areas of education, from teacher education (now called teacher training) to curriculum reform.

Other policy influencers felt that there had been recent, significant changes in education policy and that the government would be better off allowing a period of time to embed current changes. Policy Influencer 18 particularly advocated for this:

> Systems that are really effective have to stay with it for quite a period of time... but you can't get sustained change unless you keep at it for, say, you know, up to a 10-year time horizon. (England, academic government advisor)

As discussed previously, this is particularly difficult in a political system such as England, where the political longevity of a government may be short, and there has been a lot of change in education secretaries leading the Department for Education.

Teacher participants tended not to discuss specific examples of political control or national policy changes. This may indicate they are not as aware of political changes and discussions taking place within the DfE, where policies are designed and reformed. Instead, the national policies are mediated and translated into inspection frameworks by Ofsted and other policy mediators. Discussions that edged into matters of national policy focused on changes in how policies were being checked for enactment, especially through Ofsted inspections. Liverpool Teacher 7 was able to discuss this:

> I think Ofsted has been a driving force behind that. Before Ofsted had really got into that framework, we'd already started to kind of move that way anyway... give Ofsted credit where it's due... Ofsted now trying to move to a more research driven approach and what the framework is now, is a far fairer framework there ever has been. (Liverpool, school leader)

This teacher is also a school leader and, perhaps, more likely to be aware of national changes such as the new inspection framework from Ofsted. They praise the new framework for the positive changes to make school judgements fairer.

The interview data presented reveal several factors that shape how political thinking influences assessment policy in England. Firstly, they highlight a tension between the time required for meaningful system change – estimated by participants at around 10 years - and the relatively rapid turnover of political leadership and policy direction in England. These conflicts appear to create significant barriers to sustained reform, with policy initiatives often being abandoned or substantially altered before their full enactment can be evaluated. Secondly, the data suggests international assessments, particularly PISA, serve as powerful political tools that can both drive and justify policy decisions. However, policy influencers highlighted that these international comparisons are often selectively deployed to support pre-existing policy preferences rather than to inform genuine, evidence-based reform. The findings also indicate a disconnect between national olicymaking and classroom-level understanding, with teachers generally showing less awareness of broader political influences on assessment policy than those in policy roles. This gap between policy design and enactment contexts suggests potential challenges in achieving coherent system-wide change, particularly given the complex accountability mechanisms described by participants.

As England's assessment system faced these ongoing challenges of political influence and policy enactment, the arrival of COVID-19 in early 2020 would present unprecedented disruption to established assessment processes. The pandemic would test not only the technical capabilities of the assessment system but also the relationship between political tension-making and assessment policy in entirely new ways.

4.3.5 What has the impact of COVID-19 been on assessment processes in England?

The interviews informing this work took place between January and June 2023. At this point in the global pandemic, schools in Scotland and England were navigating a complex recovery environment where persistent pandemic-related disruptions continued to undermine education despite efforts to return to normality. Assessment arrangements in both jurisdictions sought to balance support for students whose learning had been interrupted with public confidence in the qualifications awarded. In England, GCSEs and A-levels returned to pre-

pandemic formats while retaining support measures such as formula sheets in mathematics and physics. Scottish learners, by contrast, sat modified course assessments under an adjusted grading approach.

4.3.5.1 Presenting a Potential for Change

Within this context, both teacher and policy influencer participants acknowledged that the pandemic had created potential for system change in assessment processes. Policy Influencer 10, however, felt that these opportunities were not realised:

> I think it [COVID-19] did create a momentary opportunity where there might have been the possibility to think about education differently, but enormous effort was put into reconstituting the system as it was before. (England, senior academic)

This contribution refers to the government's calls for a return to external examinations as soon as it deemed it safe to do so following the pandemic. Following the recommencement of external GCSE and A-level exams, there was no consultation on whether anything positive could be learned from the changes brought about by the pandemic, and a public discourse emerged around getting back to the 'fairness' of exams (Meadows *et al.*, 2025).

During the COVID-19 pandemic, there was a rapid shift from qualifications awarded on the basis of external exams to those awarded on the basis of teacher-assessed grades (TAGs), or, rather, Centre-Assessed Grades (CAGs), as schools adjusted teacher grades. This caused a significant workload for teachers, requiring new exam arrangements and creating tension between teacher assessment and students' expectations. Liverpool Teacher 7 explained this:

> We found that the results days were far more difficult for TAGs and CAGs than they are normally because normally when the students come in, they know that basically, it's been down to them. (Liverpool, school leader)

Teachers felt pressure from students and parents, especially around the release of results, as they were responsible for awarding students' full grades for the first time. This teacher also noted:

> And I think it's very difficult when you put those emotionally charged teachers in a place where they're looking in the eyes of the child that last day when the signing the shirt, knowing what grade they have assigned them. (Liverpool, school leader)

At this time, teachers were placed in new positions as assessors and awarders of qualifications, a role new to them in England and one for which they felt unprepared.

Teachers felt that removing the exams changed their relationships with students with whom they had worked towards their future. One policy influencer, Policy Influencer 1, however, felt that teachers may also be put off by not having exams for a different reason:

> Interestingly, this is in terms of COVID; the idea that there wouldn't be exams terrified so many teachers, and I was pointing out that when I started teaching in the 1970s, 40% of our kids did not even take a GCSE essay exam... for many teachers, the idea of actually having kids who they couldn't motivate by the threat of an exam is horrifying. (England, senior academic)

The use of “terrified” here may well stem from teachers' concern about the accountability they may still face if their students do not achieve certain grades in assessments.

4.3.5.2 Policy Decision-making and Missed Opportunities During the COVID-19 Assessment Crisis

The Head Teacher of the case study school also felt that some schools bowed to pressure from parents and students more than others:

> I spoke to off-the-record conversations with head teachers from other local secondary schools, whose results absolutely rocketed during TAGs and CAGS... we went through ours with a fine-tooth comb. (Liverpool, School leader)

Some schools were aware that their grades were moderated by the national algorithm and also increased the centre grades they sent to OFQUAL to ensure any students who were downgraded received what their teachers had awarded. Other schools, aware of a conversation in the profession about the algorithm comparing cohorts to previous years, downgraded teacher assessments in cohorts where teachers had expected improvement.

The mixed information and strategies schools deployed during the pandemic highlight the difficulty of navigating the fast-paced changes in assessment processes. This would involve complex decisions at any time, but at this time, the government in England were also trying to navigate other major changes to national systems and policies. However, participants clearly stated that there were significant issues with government-school communication at this time. Policy Influencer 12 observed:

> When the pandemic hit, I think it became apparent that the government had no way of just contacting all schools like quickly... the amount of stuff that government is trying to communicate is vast and often done in a not a very good way. (England, policy advisor and civil servant)

The changes, therefore, could not involve thorough consideration or consultation with schools and the public before being enacted and communicated.

This communication deficit meant policy changes were enacted and disseminated to schools without comprehensive stakeholder consultation or evaluation. Consequently, the design of the assessment processes would not represent a substantial trial of how such changes might work if teacher assessment were to be established. Decisions were made to revert to public examinations, which had been an integral part of the education system for many years.

Despite calls for a return to examinations, some stakeholders did not seem concerned about the lack of external exams. Policy Influencer 9 highlighted this attitude from the public:

> Parent Kind did a survey during the pandemic because, of course, there weren't any SATs in primary schools during the pandemic, nobody missed them and parents, didn't miss them either. But parents are a lot more worried about GCSEs and A-Levels. (England, Professional organisations and teacher professional development)

This indicates that some external examinations may be perceived as more valued by the public than others. Particularly GCSEs and A-levels, which may be more valued as they lead to qualifications that qualify for entry to further institutions and employment.

Overall, the temporary national assessments enacted during the COVID-19 pandemic did not lead to long-term change. Campaigners for a return to examinations and the status quo were successful, resulting in a return to national systems in much the same style as they were before. Some participants felt that this was a lost opportunity to reflect on the system changes and any positive effects that may have been realised. Liverpool Teacher 7 suggested that the profession missed an opportunity to build public confidence in teachers as assessors

> There could have been a really good opportunity there to say as a profession, this is how we can be trusted and how we can do it effectively. I feel that we let ourselves down a little bit by how overinflated the grades were. (Liverpool, school leader)

This reflects a national discourse that, during the pandemic, the number of students achieving high pass grades increased due to their teachers' generosity. However, this could also have been attributed to Ofqual's lack of grade suppression, which ensures that each year's cohort of grades is comparable to previous years'.

This section highlights the role of assessment processes in England, drawing on national policy influencers and teachers from a case study school in Liverpool. It highlights the purpose for which they consider these assessment processes, as well as the effects they continue to have on the education system and the people within it. The contributions from participants highlight the sometimes-extreme pressures created on stakeholders in England's education system and the way in which exams, in particular, are used as a political tool.

Particular political and cultural influences over the system can be seen within these interview contributions, but are better understood through comparison with Scotland. These comparisons will form the focus of the later discussion section, following an analysis of findings from interviews with policy influencers in Scotland and the teachers from the case study school in Dundee.

4.4 Scotland

4.4.1 What are the diverse purposes of assessment processes in education in Scotland?

During the interviews with teachers at the case study school in Dundee and national policy influencers, diverse goals of assessment processes were described. They revealed tensions between formative assessment purposes to support learning and the uses of assessment for accountability purposes. Whilst many purposes described fall under these broader categories, there were also more granular discussions of how assessment should function in Scotland to meet diverse goals.

4.4.1.1 Formative Purposes

Accounts from both education practitioners and policy influencers converged on a shared articulation that assessment should primarily serve learning. However, there was variation in the consistency of this stated commitment across participant groups. The ease with which policy influencers across both jurisdictions endorsed formative purposes may reflect a discursive consensus detached from enacted assessment practice. Accounts from participants suggest accountability demands displaced formative assessment work (see Policy Influencer 3 on page 196).

Despite accounts of formative purposes of assessment displacing assessment processes for accountability, both teachers and policy influencers converged on a shared articulation that assessment should primarily serve learning. The ease with which all parties endorsed formative purposes may itself reflect a discursive consensus that has become detached from enacted practice, see page 206 for example.

One policy influencer, in particular, Policy Influencer 15, articulated the importance of the assessment's formative purpose:

> Teachers knowing their children and their learners, well, knowing the next steps in their learning, how to assess where they're at and where they're going. (Scotland, general teaching council CEO, professional organisations and teacher professional development)

This was aligned with teacher perspectives in the case study school, for example, Dundee Teacher 1:

> Assessment for me is essentially a diagnostic tool. It's a part of the process of learning and educating in which the teacher requires pupils to be able to show their understanding or their skill set at three particular stages of that learning and teaching process: one before, one during and one after. (Dundee, Experienced)

In Scotland, teachers discussed how assessment was used for formative purposes through peer and self-assessment strategies. Dundee Teacher 2 emphasised the importance of peer-assessment:

> I'll do peer assessments with them quite often, because the classes tend to be a bit bigger and that's quite an easy way to make sure that we're getting a bit of assessment and also that they're able to kind of share their knowledge and understanding with the peers and hopefully learn from each other as well. (Dundee, experienced)

This teacher clearly saw value in students working collaboratively and building understanding from one another in the learning and assessment process. These contributions demonstrate a shared understanding of assessment processes principally serving a formative purpose, informing what should follow in teaching and learning. Both groups also noted, however, that in Scotland this primary purpose was moderated by accountability-related assessment processes.

4.4.1.2 Accountability Purposes

Both policy influencers and teachers in the case study acknowledged the significant role of assessment processes for accountability. Policy influencers were particularly critical of this purpose, with Policy Influencer 3 noting:

> Accountability has shaped assessment... teachers, I think, believe that assessment is accountability, that they are the same thing. (Scotland, Professional organisations and teacher professional development)

This participant is describing the wider effect that using assessment for accountability purposes can have on other processes and education in general. The uses of assessment for accountability of professionals seem to have developed over time, as described by Policy Influencer 5:

> What we've seen since the late 1990s is a shift away from a system premised upon professional support for practice to a system based upon measurement and evaluation of practice or its outcomes. (Scotland, senior academic)

This participant is describing a fundamental shift in how accountability is framed, from accountability with institutions to accountability of institutions using assessment data.

4.4.1.3 Hierarchical Accountability Structures

The ways in which institution leaders are held to account by data-driven processes in a hierarchical system were described by Dundee Teacher 8:

> We're obviously communicating it to the City Council, my bosses and the assessment results are used by me; I have termly reviews with my education officer, and we discuss broad general education attainment... I use the assessment information and the data to hold attainment meetings twice a year with all my PTs, and I also have an attainment meeting with the Chief Education Officer every year. (Dundee, school leader)

The school leader here clearly demonstrates many layers of accountability and multiple processes of holding people to account through the use and discussion of assessment data. In Scotland, student performance in assessment processes is being used extensively to justify the appraisal of multiple people's careers.

Much of the focus of this use of assessment-related accountability can also be seen at the highest levels of Scottish politics, with the First Minister. Policy Influencer 20 reveals how political priorities have intensified accountability pressures:

> The media also can fixate on the outcomes of that kind of assessment and so do politicians... because the First Minister, when she came into office, declared education to be her number one priority, and closing the poverty-related attainment Gap. That then immediately turned everyone's heads towards exam outcomes and qualification outcomes. (Scotland, Trade union representative)

This policy influencer describes the election of Nicola Sturgeon and focuses on data following her 'Judge Me on My Record' speech. This not only highlights the use of assessment data for accountability but also a political purpose for

assessment processes, attempting to demonstrate the effectiveness of political interventions. Many participants discussed how this was a turning point in how assessment data were used to judge the performance of all in Scottish education.

4.4.1.4 Tensions Between Formative and Accountability Purposes

Participants discussed both purposes for assessment processes of moving learning forward and for accountability as being in conflict with one another. Many reasons were provided for this, such as the pressure from accountability driving what was taught and how as well as the time spent collecting data for accountability processes. Policy Influencer 3 provided an introduction to this tension:

> There are also assessment impacts where teachers are making assessments of pupils' progress, and they are identifying what comes next... So you know there's two kinds of opposing impacts. I think there's the one that is about moving learning forward and progress. And there's one that is very much linked with accountability and the performance judgements. (Scotland, professional organisations and teacher professional development)

This contribution positions formative and accountability purposes not merely as different but as structurally oppositional. The language of 'opposing impacts' suggests a systemic incompatibility rather than a manageable tension, framing the relationship as zero-sum within the current policy structure. This structural reading was reflected in the classroom experience described by Dundee Teacher 1:

> Pupils come into this room and come into this classroom with a plethora of ideas and experiences, which I'm never really able to tap into because of the bureaucracy and the nature of the assessment process. (Dundee, experienced)

Here, the participant's reference to 'bureaucracy' locates the constraint not in assessment itself but in the administrative apparatus surrounding it. The phrase 'never really able to tap into' signals a perceived foreclosure of pedagogical possibility, suggesting that accountability demands do not merely compete with formative purposes but displace them from the available professional repertoire.

In addition to teachers using assessment for formative purposes, the extent to which students and other education stakeholders receive high-quality feedback was also described as limited. Policy influencer 20 (Trade union representative) described potential tensions in how assessment data are used:

> But they [teachers] are gathering a wealth of rich data about each individual child's learning... That does not easily shorthand for people who wish to measure and judge how well teachers are doing individually, how well schools are doing individually, how well Local Authorities are doing. (Scotland, trade union representative)

This participant's contribution shows that data from assessment processes are increasingly used for accountability judgements, and, at least in this participant's view, these judgements are not valid. Tensions from these uses of assessment data were also described as something that teachers attempt to navigate around. Dundee Teacher 4 described this in action as:

> I think teachers do their best for pupils regardless. I think it would be dishonest to say that I'm thinking about league tables when I'm teaching my pupils… But I think as a pivotal part of our job, ensuring that pupils achieve the best that they can before they leave secondary school is something that we all try to do. (Dundee, experienced teacher)

Together, these excerpts from participants highlight potential missed opportunities for using assessment for formative purposes. They also show that there appears to be more encouragement from local and national governments to use assessment to generate data, which serves the purpose of holding individuals and institutions accountable.

Assessment data, whilst reported to parents, does not seem to play much of a role in parents' school decisions. Many policy influencers, such as Policy Influencer 24, compared the situation in Scotland with that in England:

> To a certain extent, in Scotland, it does give parents a choice, although I think to a much lesser extent than in England. We don't have the same postcode lottery system in Scotland - we have it to an extent, but not to nearly the same extent. (Scotland, local policy leader)

This was confirmed by teachers at the case study school in Dundee. However, they detailed that this may be related to socio-economic reasons. Dundee Teacher 2 suggested that school choice based on assessment data were less of a concern:

> So at my school, I don't think parental involvement is quite... we are in quite a kind of, I suppose a socially economic deprived area. So, I don't think that's necessarily within the kind of demographic of what the parents are necessarily looking for. I think they're looking for a school that provides the pupils with an education but also that community and all that kind of stuff. (Dundee, experienced teacher)

The evidence from both policy leaders and teachers indicates that assessment data have limited influence on parental school choice in Scotland's educational system, particularly when compared to England. This appears to be especially true in areas of socio-economic deprivation, where parents prioritise community engagement over performance metrics.

While school accountability in Scotland operates largely independently of parental choice mechanisms, teachers continue to play a significant role in enacting and overseeing critical assessment processes. Teachers described taking an active role in preparing students for qualifications while managing various pressures. Dundee Teacher 2 explained:

> We obviously have unit assessments as well just to track progress... particularly when it comes towards SQA exams and estimates because our estimates need to be based on physical evidence). Here, teachers operate the dual roles of tracking their students (Dundee, experienced teacher)

learning and preparing estimates of future exam performance, especially following the COVID-19 pandemic, where they had to submit centre grades. There are tensions from this, however, or at times when the preparation for future exams causes tensions between these roles. Dundee Teacher 1 highlighted such tensions:

> Again, high-stakes exams. So, I'm thinking I need to assess a critical essay in 2nd year, since they'll have to write one for their fifth-year examination... And so those, because they feed into the high-stakes assessments, become subconsciously something that we tend to focus on a lot more. (Dundee, experienced teacher)

The phrase ‘a lot more’ from this participant emphasises the strong influence of end-of-school examinations over teacher assessment practice and its dominance, even if subconsciously.

Policy influencers also discuss the influence of the high-stakes assessment at the end of school. Policy Influencer 5 observed:

> It isn't the assessments per se, which are the problem... It is to do with the way in which assessment data are used to measure the success of schools and teachers, which then has significant backwash effects on how teachers behave in their decisions and actions around curriculum making. (Scotland, senior academic)

This participant’s description of backwash effects highlights the challenges teachers face in balancing different assessment purposes and how getting this balance wrong can influence not only their assessment practice but also the curriculum and its delivery.

While teachers acknowledged concerns about the washback effects from final assessments, they identified various benefits of using this assessment formatively in their practice. Dundee Teacher 2 described how assessment data from assessments can be used in their practice:

> What I tend to do with these spreadsheets as well is I break them down into individual topics or questions, and then I'll traffic light these, which means that I know there's stuff that perhaps I haven't taught and I need to go back because I can see that the majority in the class haven't understood that or I can help kids identify their individual points that they need to work on. (Dundee, experienced teacher)

However, this raises significant validity concerns about the practice described by this teacher, which may inform the assessments that produce this data. If the teacher is using material from external exams, these questions may not be the most valid way to assess what the student knows, as these tools are designed for summative rather than formative purposes. This will affect the quality of judgements the teacher can make, creating a washback effect in which the teacher effectively teaches to the test material rather than delivering the intended curriculum outcomes.

4.4.1.5 Reform, trust and standardisation

Throughout the discussion with participants on the purposes for assessment, many demonstrated that significant changes are needed. Policy Influencer 5 argued for reform in the moderation processes used to generate assessment data:

> I'd also like to see them being less bureaucratic. I know that having talked to secondary teachers is that there's an awful lot of energy and time goes into doing moderation, for example, for what are effectively high- stakes for schools, but low stakes for, for young people. (Scotland, senior academic)

This policy influencer wanted to argue that teachers should be able to focus their time on gaining a holistic picture of where students are and on planning development opportunities. However, if the data from school assessments are being used, and teachers are offered professional autonomy to do this, Dundee Teacher 2 explained the need for these moderation processes

> I want to know it's fair. I want to know that how I am assessing my class is the same as somebody in Glasgow or Renfrewshire or whatever... You want your pupils to get what they deserve. (Dundee, experienced teacher)

Overall, teachers advocated greater trust in teacher professional judgement but agreed with policy influencer 5's concern that these processes sometimes require significant time that is currently not afforded to them.

The evidence presented above reveals that assessment processes are used for multiple purposes in Scotland's schools. While this was described as causing some tension between the high-stakes nature of school examinations and more formative purposes, teachers described a wide range of formative assessment strategies in their practice. Pressures from accountability mechanisms based on assessment data were pervasive and described as causing washback effects on teacher practice and the curriculum. While parents' choice of school does not create competition among schools, teachers still feel pressure from metrics of exam outcomes and the judgements they inform. The experiences described here suggest that further evolution of assessment processes in Scotland must

consider the restrictive pressures of external high-stakes exams on such policy development.

4.4.2 Perceptions of assessment processes in Scotland

This section explores and analyses the perspectives of policy influencers and Teachers in Dundee regarding the impact of assessment processes on educational stakeholders. Three principal areas of concern were raised by participants: the implications of assessment practices for students' mental health and well-being, the variable levels of assessment literacy among teaching professionals, and the challenges associated with external stakeholders' interpretation and understanding of assessment data. By discussing these key themes, this section aims to provide a more comprehensive understanding of how assessment processes are perceived to affect various educational stakeholders.

4.4.2.1 High-stakes examinations and student mental health

Teachers argued for reforms in the examination system due to concerns about fairness and the effects of examinations on young people. Dundee Teacher 5 argued:

> My first reason is that it's national policy, education policy, children services policy to promote the mental health and wellbeing of all of our young people and exams don't do that. They do quite the opposite. (Dundee, school leader)

This teachers displayed concern about the negative mental health effects of exams that were sat by young people, especially at the end of their school lives and the detrimental effects of pressure caused by them. The necessity of assessing student achievement in this way was also questioned by Dundee Teacher 8:

> I mean I don't think anywhere now does exams at the end of the year; in university or workplaces, I think it's more open book ongoing type assessment and if I'm honest, I mean I certainly don't in the world of work, need to sit or remember a lot of information to then regurgitate it. (Dundee, school leader)

The teachers in the Dundee case study described examinations as serving as a conduit for pressures on student mental health. In addition, their concerns about

the validity and whether examinations are the most appropriate or necessary approach in their context are also an important contribution here.

4.4.2.2 Assessment literacy and external stakeholder understanding

In addition to system changes in the examination system, several participants suggested improving how assessment information is communicated. Policy Influencer 19 explained this need:

> We need to help parents understand what their child needs to do to get better... I'm not sure teachers either in primary or secondary are particularly skilled at doing that with parents and understanding for some what the limitations of the parents are. (Scotland, local policy leader)

This policy influencer highlights the difficulty for parents who lack assessment literacy to fully understand the assessment data being discussed between them and schools. Difficulty with the public's assessment literacy is further discussed by this participant as resulting from media influence:

> I think they [the public] pretend to be [assessment literate], and I think the media plays a huge role in that. So it gives the impression of knowledge I mean, I'm not averse to having a go at the media too with some of their statements and using and creating league tables and all the rest of it which gives the impression to parents that assessments are robust and that the sentences that the media write about attainment and about their perceived attainment of schools is accurate, when in actual fact it's not in the slightest. (Scotland, local policy leader)

This statement describes how some of the uses of assessment data in the media can serve to perpetuate false conclusions that are made from assessment data rather than clarifying them.

Assessment literacy was also a concern for teachers and the policy influencers interviewed. While policy influencers expressed concern about limited assessment literacy among teachers, the teachers themselves generally had confidence in their understanding of assessment. Policy Influencer 26 stated:

> I think teachers get very little training in how to assess when they're in their postgraduate training courses when they first start out. Early teachers have very little experience in assessment, and it's very unfair

> of anyone to ask them to assess. (Scotland, former CEO of SQA, policy critic)

This contribution identifies that the brief nature of teacher training courses may not adequately prepare teachers to be competent in conducting assessments. Whilst teachers, such as Dundee teacher 1, described how their understanding of assessment processes developed over time:

> I think I'm a lot more aware of the formative observational types of assessment which exist during the course of a lesson. I think early on in my career, or early on in many teachers' careers, you're essentially just getting to grips with things like the behaviour, the content of your lessons... But then the assessment and the understanding of actually 'are my pupils progressing? Is what I've created helping them to improve in their understanding and will this help them to be better at the end of the year in something that we were at the beginning?' That's something which is definitely improved. (Dundee, experienced teacher)

As they matured as a teacher, this teachers clearly identified a number of areas that needed development at the start of their career and described it as almost inevitable that it would take time for their competency to improve in these areas.

4.4.2.3 Professional learning as a response through school-based action research

Mechanisms to improve teacher assessment literacy were discussed throughout the interviews with teachers in the case study school. One area of professional development that had been effective was the recent action research in which teachers participated at the school. Dundee Teacher 2 described:

> I suppose school policy, we do have initiatives, so we are part of... you have to do research projects and stuff like that. So, our research project last year was on AIFL and we had to pick a certain aspect of that. So, I suppose these kinds of things do drive forward what your assessment looks like within your classroom and how you're assessing? (Dundee, experienced teacher)

Through this action research as professional development, many of the teachers described being empowered to construct professional development for themselves that impacted their own performance in conducting assessment processes.

4.4.2.4 Summary

This analysis reveals multiple stakeholder concerns regarding assessment literacy, alongside evidence that high-stakes examinations intensify pressure on students. What emerged from the interview data were multiple concerns over people's ability to understand and interpret assessment data through adequate assessment literacy. The pressures created by the high-stakes examination system on students were also raised as a concern. Despite this, the case-study school did reveal a recent school policy of action research that had positively impacted teachers' practice and their assessment literacy. For this change to be widespread, national-level policy changes would need to be considered to further impact assessment practice.

4.4.3 Assessment policy enactment in Scotland and the gap between intentions and practice

Interview data revealed persistent tensions between assessment policy intentions and classroom realities in Scotland, with participants identifying three interconnected factors and cultural resistance to reform. Several major themes emerged from these conversations, including the difficulties of policy enactment in Scotland, the promotion and realisation of fairness in assessments and the need for cultural change to realise the aims of assessment policy.

The enactment of policies in the Dundee case study school faced challenges in resource allocation, with teachers specifically highlighting constraints that hindered effective delivery. These resource limitations created barriers to enacting the policy as intended. Dundee Teacher 1 points to these limitations:

> So from that perspective, I do think I'm understanding and aware of the benefits and the cons, and that's different to necessarily having the resources and time to truly work it out and correctly implement it. (Dundee, experienced teacher)

This teacher is acknowledging that a paucity of resources, such as time, can limit the extent to which policy changes can be fully understood and enacted. With regards to assessment policy, they were able to provide one specific example:

> At the end of each lesson, there's often just a very quick plenary, obviously, that's dependent on time restraints and resources etcetera however, I'll often ask them for something as simple as if you could put your right hand up if… or your left hand up if… or if everybody could just nod their heads if… or 123 finger arrangement for how well you've understood a concept and that allows me to have a bird's eye view understanding of how that particular session has gone within that lesson, and then to tailor next sessions accordingly and that's sort of how I would, day to day. (Dundee, experienced teacher)

The assessment practices described here, such as hand-raising and finger-counting, are rapid, low-fidelity proxies for understanding based on formative assessment theory. Their prevalence in this account suggests that the policy aspiration for responsive, individualised formative assessment has been compressed into gestural shortcuts by unaddressed resource constraints. This is a source of frustration for this teacher, who is demonstrating their knowledge around the use of assessment to inform the next learning opportunity as per the school's policy, but is held back through a lack of resources. They later followed this with:

> So from that perspective, I do think I'm understanding and aware of the benefits and the cons and that's different to necessarily having the resources and time to truly work it out and correctly implement it. (Dundee, experienced teacher)

In their opinion, they feel that they are confident with the various purposes of assessment, but are limited in the enactment of their school's and national assessment policy.

Policy influencers also recognised the extent to which the enactment of policies was limited by the timescales with which they were expected to be enacted in schools. Policy Influencer 20 highlighted the enactment issues with national assessment processes:

> Those qualifications were rushed in rather than schools being given proper time to consider them and to consider how they would need to adapt timetables and pedagogy, and professional learning in order to be able to implement them. The reason why they were rushed was because the independence referendum was happening in the same year as the qualifications for being introduced. (Scotland, trade union representative)

This participant is clearly relating the impact of competing factors in political activity, having a detrimental influence on the enactment of education policy, such as reforming qualifications.

4.4.3.1 Student comprehension and formative assessment practices

Another factor affecting the successful enactment of assessment policy in schools was the extent to which students understood the purpose of assessment and the feedback they received after it, making it a formative experience. Dundee Teacher 5 highlighted some of these challenges:

> we did a CLPL CPD for staff on the use of feedback and we had conducted some pupil voice focus groups on that and I do remember doing these groups myself and at that point, about 2018-2019, young people said ''well we get feedback but sometimes we don't understand it''. (Dundee, school leader)

If the students that teachers are working with cannot comprehend the feedback, this will, of course, inhibit the impact of any guidance provided and not fulfil the formative purpose of the assessment.

4.4.3.2 Meso-level support structures and systemic mediation

Although individual teachers and the school leaders in the case study school spoke optimistically about their assessment literacy, policy influencers reported a lack of support for further development due to systemic issues. Policy Influencer 5 captures one systemic issue:

> The other thing that we lack very much in Scotland is any sense of the importance of the meso. And by that, I mean the processes and structures which are in place to support schools, if you like, to connect schools with policy. (Scotland, senior academic)

This participant is advocating for more support to be provided to schools, to help with navigating the intentions of policy whilst mediating school practicalities. Without this mediation process, tensions between policy intent and classroom reality may persist, with teachers lacking the necessary resources, time, and support to effectively implement policy changes.

4.4.3.3 Teacher agency and the equity-standardisation tension

Given the time and resource limitations available to Dundee teachers, teacher agency in assessment processes for qualifications can also be challenging. National policy influencers tended to advocate for teacher autonomy in assessment, whilst many teachers in the case study school expressed concerns about equity and fairness without proper standardisation processes. Policy Influencer 19 highlighted the undermining of potential strengths of teacher judgement by examinations:

> that balance between the teacher's judgement who must know so much more about that child than any exam will ever know. How do you get that balance between the teachers view and an external exam. (Scotland, local policy leader)

Policy influencers felt that teachers' professional experience could be trusted and encouraged to provide a more comprehensive overview of student performance. However, the teacher participants in the case study were cautious about reducing or removing standardised assessments. Dundee Teacher 2 expressed such concerns:

> My massive concern with that is that if they do remove exams, how can I be confident that how I'm assessing the pupils in front of me is the same as someone else in Glasgow, for example, is assessing pupils? How do I know they're the same level? Whereas having a national exam board, does kind of ensure parity. (Dundee experienced teacher)

This teacher does seem to have concerns around the clarity of assessment standards. However, the issues they raise might also be mediated by moderation processes that were also advocated by many of the other teachers.

4.4.3.4 Professional preparation for assessment roles

Both groups raised concerns about the assessment literacy of teachers and teacher preparation for their role in assessment processes. Policy Influencer 26 raised concerns about the preparedness of teachers for their roles as assessors

> I think teachers get very little training in how to assess when they're in their postgraduate training courses when they first start out. Early teachers have very little experience in assessment and it's very unfair

> of anyone to ask them to assess. (Scotland, former CEO of SQA, policy critic)

This contribution highlights the concerns of this participant for the level of preparation that teachers receive as part of their training. Even early teachers are highlighted as having little experience, which would be of particular concern to many if inexperienced teachers were asked to undertake high-stakes assessment processes. Teacher participants themselves raised further concerns. Dundee Teacher 7 raised particular concerns with newly qualified teachers and their assessment knowledge:

> I would suggest that in our case that is probably not the case, because I think that NQT's on the whole, I would say tend to be, you know there's subject base is good... However, being closer to their own degree learning in terms of time, does not necessarily equip the better for supporting young people to pass SQA exams. Those are two entirely different skills. (Dundee, school middle –tier leader)

These perspectives reflect broader concerns among teachers for the training received for conducting assessment processes. Without the standard and depth of training required, there appears to be consensus of concern among teachers and policy influencers that they will not be able to undertake high-stakes assessment processes.

4.4.3.5 Formative aspirations and high-stakes realities

While participants expressed concerns regarding inadequate teacher training and development for high-stakes summative assessment processes, there was a contrast in perspectives on the extent to which assessment policy reforms had enhanced assessment literacy for formative practices. Overall, however, participants felt that a significant tension persists in Scottish education between the intentions of formative assessment policy and the realities of high-stakes assessment. Both groups of participants acknowledged that high-stakes examinations dominate educational practice, particularly in secondary schools. Policy Influencer 20 highlights the gap between policy aspirations and realities:

> This potentially transformational development... we are perhaps not seeing the realisation of that aspiration to the extent that we wish to see it because of the very, very kind of like augmented status that

> formal qualifications and exams and tests continue to have in Scotland. (Scotland, trade union representative)

This participant is highlighting the tensions between developments in formative assessment practice advocated through assessment is for learning policies, described as ‘transformative’, and the pressures from the national exam system. These tensions were reiterated by teacher participants who also felt high-stakes exams had limited the realisation of policy intentions.

Teachers described how they value formative assessment but report an increase in exam-focused approaches as students progress through the system, with formative practices more evident in the Broad General Education phase than in the examination-centric senior phase. Dundee Teacher 1 explains this tension in secondary age students:

> We are still feeding into a senior stage where high stakes examinations mean very strict outcomes, almost for pupils that achieve or don't achieve a certain something at the end of it. We are then forced back into that box at the BGE stage, and it's not a nice thing to say. It's not how I would like it to work. (Dundee, experienced teacher)

This teacher is explaining how external exams can influence their assessment practice, forcing them to adopt similar styles of assessment process for junior secondary-aged students as well as examination year students. Dundee Teacher 2 also described their similar experience of this:

> We'll try and use that kind of level of assessment as that kind of extension to try and kind of push your pupils on. If you've got a particularly good second year class, sometimes I'll say that I'll write a question on the board and they'll do it in the class, I'll say, well, this is actually a National 5 question. (Dundee, experienced teacher)

This teacher highlights how they extend greater challenges to students in lower secondary school by using past examination materials. This teacher participant’s idea of using these to push their pupils on highlights how they feel pressured to extend their students to meet the demands of the external examinations as soon as possible. This might be an attempt to ensure their students are as primed as possible for the final examinations by starting the preparations early. The teacher participants did not explicitly state what motivated them to use senior

phase exam assessment materials in this stage of the interviews, but did describe the pressure of accountability leading them to begin preparing students for senior examinations as early as possible. The contributions from participants highlight how the desire to prepare students for external examinations has reduced the opportunities for teachers to enact the training they have received for formative purposes of assessment.

4.4.3.6 Professional learning and assessment capacity development

Despite tensions between formative and summative assessment, many participants were complimentary of the training and development that teachers at a local level in Scotland have received in formative assessment practice. Policy Influencer 19 noted improvements in assessment literacy over time:

> I think it's got much, much better over the years... I think the world has moved and teachers now, the young teachers coming in have a really sharp knowledge and understanding of assessment. Much, much more so than when I came through. (Scotland, local policy leaders)

This contribution describes a perceived improvement in assessment practice that appears to contradict previous assertions about teachers' capacity to effectively implement assessment processes. It is noteworthy, however, that this participant specifically referred to continuous classroom assessment for formative purposes, highlighting their perspective that teachers had demonstrated improvement in their deployment of formative assessment strategies in Scottish classrooms.

In the case study school, Dundee Teacher 5 described a school development project that had particularly helped formative assessment practice develop:

> We did a CLPL/CPD for staff on the use of feedback, and we had conducted some pupil voice focus groups on that... young people said ''well we get feedback, but sometimes we don't understand it.'' And the big thing was ''we don't act on it because we're supposed to act on it in our own time at home, and we don't do that.''... We've made that approach a little bit more robust over the past four or so years; we've made the focus of our individual teacher practitioner inquiry, feedback. (Dundee, school leader)

This school leader described a mode of professional learning that was also mentioned positively by many individual teachers, in which classroom teachers

were asked to set action research targets to improve their formative assessment practice. Supported by learner feedback regarding their expectations for feedback and motivated by the perceived autonomy of the action research, the teachers described feeling equipped to refine their formative assessment practice. This enabled them to enact a range of strategies advocated by school policy, applying them selectively in ways they deemed most conducive to student learning.

One area of contention among participants regarding assessment policy impacts centred on the distinction between systemic constraints imposed by macro-level policy changes and more localised limitations that emerge through school and classroom-level factors. Many national policy influencers tended to advocate for the redistribution of power for making wholesale changes to assessment approaches, such as policy influencer 5:

> In five years, if the world had transformed, there would not be this view that Scottish government sit and make policy and teachers implement it. There would be an understanding that teachers are the creators of the policy in their classrooms with their children, in their context, they're critical makers, and that means that they are policy makers, and that includes assessment for me. (Scotland, senior academic)

This participant shared the view with many other contributors that a more localised approach to policy creation and enactment was needed. Policy Influencer 4 was an advocate for systemic restructuring

> My view is that what we should have is a system that doesn't work from curriculum, change the curriculum, change every seven or so years, whether it be Highers still, or whether it be the new National Qualifications, or whether it be intermediate summary, or whether it be 5-14 years. We need a system in Scotland, that is much more progressive system that keeps the curriculum under constant review. (Scotland, general teaching council CEO, policy advisor)

From their perspective, there needed to be a radical change to how policy was designed and how educational reform was enacted. They also advocated for a teacher and school-level approach:

> That can only be done by using the expertise, in my mind it can only be done by using the expertise of the practitioners who are on the

> ground, who can feed through a national agency, what's working well, and what's not working well. And then when it comes to examination, the examination body, then basically being instructed through ministers who, at the end of the day, will always have strategic responsibility. (Scotland, general teaching council CEO, policy advisor)

This participant is critical of periodic policy overhauls where the Scottish educational system undergoes repeated structural reforms without necessarily addressing foundational issues or allowing sufficient time for evaluating its enactment.

While national policy influencers advocated for system reform to policy enactment, participants from the case study school directed their suggestions toward classroom-level constraints and the localised adaptations they perceived as necessary. Dundee Teacher 1 described trying to practice their more progressive pedagogical values within what they perceive to be restrictive high-stakes assessment regimes. Further, they position themselves in alignment with the Curriculum for Excellence, which aims to ‘validate experiences and values’ but feel that this is ultimately undermined by the dominance of high-stakes examination systems at the senior phase.

When considering potential changes to the current system, teacher participants were concerned about how this might be enacted if they were given greater autonomy. Dundee Teacher 2 highlights their concern over workload from changes to assessment policy:

> And I don't want to have to produce all the materials for it because the burden in terms of work code is massive as it is. So, if they want us to do this, they have to say ''here are the materials for you to do it''. (Dundee, experienced teacher)

This contribution reveals a material limit to reforms of teacher practice, where policy aspirations may not be met by the provision of adequate resources, including teacher time. The teacher participants' solution to the intensification of their workload warrants attention; in their view, they need ready-made materials to assess their students, not because they lack the expertise to create them, but because they lack the time resource. Without consideration of the resources needed to realise policy aspirations, many participants were concerned that they could not be enacted successfully.

4.4.3.7 Cultural assumptions and institutional inertia

The contributions from several national and case study participants suggest that while some policy intentions have had some impact, their full realisation is constrained by deep-seated cultural assumptions. Policy Influencer 26 relayed experiences that revealed these assumptions:

> The number of people around that table at the Education Committee who specified that external assessment was the only way that you could actually get through assessment, that written exams were the only valid assessments... the politicians across the spectrum, and parents and teachers are not agreed that different types of assessment are equally valid and as a result of that you have perverse incentives on all types of assessment. (Scotland, former CEO of SQA, policy critic)

This participant appears to be describing an embedded cultural attitude to assessment whereby exams and external assessment are seen as the only robust way to judge learning. This underlying assumption would make changes to assessment processes difficult, because whilst some in Scottish education may appear open to alternative assessment formats, there may be an underlying bias towards exams as a superior method. If new policies were launched that did not address these preconceptions about what makes valid and effective assessment processes, then there would be constant pressure for the intended changes to be unsuccessful.

Further pressures acting against policy reform on a national level were mentioned by Policy Influencer 21:

> There can't be anything wrong with the system and so you've got the unions, the local authorities, Head Teachers and parent bodies are actually in some sort of unholy alliance, basically to prevent any change from taking place in the system. And you might find that ironic, but I actually think that that's the biggest problem as soon as you try and implement change; those four groups, and parents in a sense, are often manipulated by other groups, come together to actually prevent any substantive change from taking place. (Scotland, education leadership consultant)

This policy influencer's account reveals cultural resistance through coalitions of multiple stakeholders in Scottish education, which act with a tendency toward institutional self-protection to prevent change. These thoughts were repeated by

other participants, such as Policy Influencer 4, who located resistance specifically within civil service structures:

> The barriers are very simple. The barriers are the civil service and the extent to which they are responsible for implementing government strategy. And I know just through my engagement with the recommendations that I've made, the expertise that exists within the civil service and the open-mindedness to understand what an education system can be like, is not there. (Scotland, general teaching council CEO, policy advisor)

This participant's reference to 'open-mindedness' suggests potential cultural constraints that operate through institutional thought processes rather than just procedural contradictions. Finally, Policy Influencer 5 identified how assessment might become culturally entangled with performance measurement, creating resistance through professional identity formation:

> This is a story of education systems worldwide actually, that the methodology around performance is not working. So the answer is always to do more of it or to do it in more nuanced ways, or to do it in more specific ways... the methodology is the problem because it engenders performativity in the system. (Scotland, senior academic)

This participant's reference to 'doing more of it' points to a cultural tendency toward intensification of workload rather than transformation when faced with the limitations of policy enactment.

4.4.3.8 From enactment challenges to political dimensions

The participant dialogue above reveals fundamental tensions between transformative policy rhetoric and persistent institutional practices. These tensions do not operate as technical enactment challenges, but as cultural phenomena involving stakeholder interests, professional identities and systemic structures. As participants described the political dimensions of assessment policymaking, it became apparent that many of these same tensions reflected broader political tensions, and participants often made these links explicitly. Understanding these political dimensions provides essential context for understanding why assessment policy reforms repeatedly encounter legislative challenges, even when there may be general consensus about their pedagogical merits.

4.4.4 Political pressures and assessment policymaking in Scotland

Political pressures can exert considerable influence over assessment policymaking in Scotland, with perspectives from both policy influencers and teacher participants revealing how assessment data functions as a contested political resource. The evidence from interviews suggests that data are deployed in various ways to advance constitutional arguments, demonstrate governmental performance, and drive accountability mechanisms.

4.4.4.1 Assessment data as political currency

When considering the impact of political pressures on assessment policymaking, one key point of synergy between policy influencers and teachers was the use of assessment processes as a political tool. Policy influencers made many statements regarding how assessment data might be used to generate data that might serve as a political tool. Policy Influencer 15 describes such instances:

> Yeah, I think education is a political hot potato in Scotland. So, these types of things get debated at a political level. And I'm not suggesting they shouldn't, but they don't always have the educational understanding of what they're just, the data can be used as a stick to beat a different part of government about depending on, you know what, somebody's got a particular emphasis on. (Scotland, general teaching council CEO, professional organisations and teacher professional development)

The phrase ‘education is a political hot potato’ acknowledges the contentious nature of educational policy in Scotland’s political landscape, particularly when assessment data are used to make political arguments.

Political arguments around devolution and the Scottish constitution can also draw on data from assessment processes. Policy Influencer 20 provided an insight into how this might occur:

> The media also, can fixate on the outcomes of that kind of assessment and so do politicians since Devolution, so politicians and our Parliament more so over the last 15 years or so that the SNP government has been in power because, well constitutional question has fuelled political debate that may otherwise not be that heated or that intense, but has fuelled political debate around all of that because the First Minister, when she came into office, declared

> education to be her number one priority, and closing the poverty-related attainment Gap. (Scotland, trade union representative)

This contribution explains how assessment data can be used as a political current in public discourse, especially in the media. In this case, the participant appears to be suggesting that there may be a mechanism through which assessment data are used politically to reinforce a sense of identity as an independent nation. The First Minister's strategic use of assessment data to underpin a political commitment might constrain the political landscape, limiting educational policies to those that demonstrably support this established objective.

Teachers in the Dundee case study school also recognised the political dimension of influence over policy creation. Dundee Teacher 2 identifies:

> I also think politically as well, I am fairly confident that some schools, some local authorities will inflate their results and figures and all that because politically, it looks good for them. (Dundee, experienced teacher)

This contribution captures a view of the ways in which assessment data are used in schools and local authorities to present a positive view of their performance. In this way, from the participant's perspective, data are used to demonstrate success outwardly and shapes institutional behaviour and policy. These behaviours may be attributed to accountability pressures that drive the need to demonstrate educational improvement through apparent improvements in assessment data.

4.4.4.2 Competing accountability paradigms

There were multiple tensions discussed between accountability and the policy goals for improving education. Policy Influencer 13 referenced such tensions:

> Data information evidence is used in very, very crude ways to identify. Let's say, teachers who are successful in teachers who are not successful. and that these very crude ways of using that... I think it's better to move the focus on to improvement. (Scotland, policy advisor and civil servant)

This participant is trying to distinguish the current use of data to drive accountability mechanisms, differentiating between 'successful' and

‘unsuccessful’ teachers, and more authentic ideas of what improvement may look like. In this way, they are trying to highlight how current accountability mechanisms may undermine the improvement processes they are trying to promote. Policy Influencer 15 also articulated a similar view of these different modes of accountability:

> In Scottish education, we float between two meanings. So, we float between a general sense of accountability that nobody would argue with, you know, we're all accountable to other people. Teachers are accountable to learners, to their Community, that's in a general sense, but we flip between using it in a general sense and using it in a technical managerial sense in Scottish education. So accountability then becomes very… I hate this phrase… ‘’data-driven’’. We've got this, these numbers on a bit of paper, and that must mean this, that must mean this about your ability or competence as a teacher. I don't think we've got our definition straight about what we mean when we say accountability. (Scotland, general teaching council CEO, professional organisations and teacher professional development)

This participant expands on the previous observation of ‘floating between two meanings’ of accountability through the use of assessment data and a more relational sense of accountability defined through professional relationships and commitment to learners and the wider school community. These participants appear to assert that if political policy drivers emphasise the use of data metrics to measure educational improvement, then this will demote the emphasis on other forms of measuring advancement.

In the Case-study school, Dundee Teacher 5 identified their attempt to distance themselves from traditional punitive accountability approaches:

> I know, that whereas years ago, if you're the principal teacher, you went into an attainment review with your headteacher, you might have expected to kind of be battered about the head and body about the data because it's just not good enough. Not by me! But I would say my approach is very much an attainment review. The point of it is to change the lives of young people through looking at the data, seeing what changes we can make. (Dundee, school leader)

They explicitly reject punitive approaches, ‘Not by me’, acknowledging that they reject practices of using data to drive improvement through punitive mechanisms. Despite this apparent rejection, the fundamental use of data to drive performativity in the school remains intact. Although they discuss shifting

away from disciplinary control of teacher behaviours in response to policy, there remains a form of surveillance through assessment data.

Teacher participants in the case study also recognise how accountability can prevent them from adopting practices that would be preferable to them or other policies. Policy Influencer 21 describes the focus on narrowing the achievement gap as potentially limiting more ambitious improvements:

> I do think the whole system has been overtaken by the whole idea of narrowing the achievement gap and nothing else matters. Nothing else matters. And it's a relatively safe space to play in. I mean, what would happen to a school that really accelerated the top end by, let's take a crazy number - 30% - and had really done well, and you only accelerated the bottom 20%. To me, that would have been fantastic. It would have been absolutely wonderful. But people would have said well, they're not doing as well. (Scotland, education leadership consultant)

The repetition of ‘’nothing else matters’’ reveals profound frustration with what the speaker appears to perceive as a singular policy focus. They are critical of how the achievement gap policy discourse seems to have monopolised the policy space to the exclusion of other educational aims. The particular focus of politicians also seemed to be in tension with local priorities, according to Dundee Teacher 4:

> When I started teaching in the 80s, it wasn't a high stakes exam system, but I suppose it's a downside of devolution and the politicians being so much closer to what we do in the classroom that has resulted in the high stakes. (Dundee, experienced teacher)

This teacher participant in the school in Dundee is describing the intensification of assessment regimes and is critical of how devolution has increased rather than what they may have expected, decreased accountability pressures in education.

4.4.4.3 Norm-referencing and transparency concerns within the SQA

The issue of norm-referencing within Scottish qualifications represents another contentious dimension of accountability mechanisms. Policy influencers expressed concerns that the Scottish Qualifications Authority (SQA) despite reported claims to the contrary. Policy Influencer 19 recalled:

> I mean I went out publicly about in the norm referenced approach with the exams and [name omitted] from SQA was so cross with me and she even named me in the media, which is actually, in my view, that's bad form, bad form. But it's bloody true! So I've got data that shows that at National 5, regardless of the number of adults and children being presented, 2/3 get A to C, and that's what it was; it was static for all those years. (Scotland, local policy leaders)

This participant identifies a discrepancy between the publicly stated grading methodologies and the SQA's operational practices, raising concerns about transparency in the exam system. If norm-referencing is indeed being applied, despite official denials, this might raise fundamental concerns about what national examinations such as National 5 genuinely signify and what their purpose might be. This contribution does not assert what the motivations might be for SQA. However, the static pass rates may suggest that norm-referencing is occurring.

Teachers in the Scottish case study school also speak to the experience of working across competing assessment paradigms within their everyday practice. One experienced teacher describes the difficulty of judging student progress within the Broad General Education (BGE):

> I sometimes feel that's a bit more kind of wishy washy. It's difficult to actually know that you're definitely Level 3 on target, or you know, I feel these are a bit more kind of interpretive and sometimes you're basing pupils on the kind of overall cohort of that year... Whereas I suppose within the senior phase, you've got 80% you're an A, you've got 60% you're a B, 50% you're a C. That's a bit more kind of tangible, whereas in the junior BGE, is about more interpretive, I suppose, and you're using you're professional judgement, which I think can be a bit more challenging. (Dundee, experienced teacher)

The teacher draws a sharp contrast between two evaluative registers, casting Broad General Education judgements as 'wishy washy' and 'interpretive' while characterising senior phase percentage marking as 'tangible'. This framing surfaces a tension between competing policy logics within Scottish assessment, each carrying different assumptions about what counts as legitimate evidence of learning. The phrase 'basing pupils on the kind of overall cohort of that year' indicates an attachment to comparative or norm-referenced judgement, a logic long associated with accountability-driven systems and the international assessment movement. Sitting against this is a discomfort with Broad General

Education assessment, where Curriculum for Excellence (CfE) sought to displace prescriptive measurement in favour of holistic development, personalisation and professional judgement. The teacher's frustration captures the position of a practitioner caught between two policy logics rather than working comfortably within either.

4.4.4.4 International assessments and policy responsiveness

The political influence of international assessments such as PISA (Programme for International Student Assessment) was discussed not only from a public perspective but also as a factor in the creation of education policy. In particular, some policy influencers referenced the ‘PISA Shock’, a reactive policy response to politically disappointing results, as a process that had influenced assessment practice in Scotland.

Policy Influencer 19 highlights how they see the impact of PISA assessment results differing between different cultures:

> I find PISA fascinating from the kind of stakeholder view. So, I was once, many years ago, on an aeroplane when I was in HMI, coming back from Germany, sitting next to a lady chatting to her, mentioned that I was a schools inspector and she then went on about how awful it was that Germany had done so poorly in their PISA and how traumatic it was for them as a nation. And I thought, wow, that is amazing if I was sitting next to a Scottish person and Pisa had bombed; they would have no conception or understanding. (Scotland, local policy leaders)

In their view, Scottish stakeholders may not be as aware of their nation’s PISA score or find it 'traumatic' should the results be disappointing. Their account shows how international assessments like PISA may serve as technical assessment tools and as cultural phenomena that interact with national identity, public discourse, and political accountability to varying degrees.

It was noted that politicians can spotlight the results of national and international assessments to shape public perception. Policy Influencer 26 suggests politicians engage in a tactical interpretation of PISA assessment outcomes:

> I think politicians, they're held to account by the public, held to account by their opposition, and therefore they want to make the best news that they possibly can, so if there's been a marginal increase in one specific thing, they will blow that as big as possible and if there has been a decline in something else, they will keep it really quiet. (Scotland, former CEO of SQA, policy critic)

This quote illustrates the potential for international assessment results to be strategically selected to support political positions, and the potential for policy to be based on a deliberate selection and utilisation of assessment data. The technical understanding of what the assessment results may reveal about educational quality is said here to be disregarded in favour of conclusions that may be more politically expedient, especially when ‘’held to account by the public, held to account by their opposition’’ as highlighted by this participant.

The sense that international assessment data and evidence from the OECD generally might be used as political currency within national discourse was discussed by other interviewees. Policy Influencer 4 noted:

> I think OECD is, the reaction to OECD is probably overplayed by politicians. In fact, it is, it is overplayed, let's be honest. Because it's a very narrow set of measures that the OECD assessments give us. (Scotland, general teaching council CEO, policy advisor)

This participant appears to challenge the disproportionate use of assessment metrics by the OECD, describing them as ‘overplayed’. Their critique further adds to previous contributions, highlighting an apparent disparity between what international tests measure and the implications for education that might be drawn from them in a political theatre.

The influence of international assessment regimes and organisations such as the OECD was identified as explicitly intended to shape assessment practice at the school level. However, participants consistently highlighted disjunctions between the assessment frameworks operating at international and national levels and their actual enactment in local policy contexts. Dundee Teacher 8 observed:

> Well, yeah. I mean, they're part of the National Improvement Framework...aims as well as improving attainment for all. So yeah, improvements in literacy and numeracy are obviously particularly big.

> But yeah, I mean, if we run, for example, National 5 History, we will be compared through insight to other National 5 History courses, but there is nothing that says that you have to do this, it's just, I suppose, what departments are comfortable in and what pupils are keen to opt into. (Dundee, school leader)

This quote illustrates a complex relationship between national policy, influenced by OECD reports for Scotland, and policy as enacted. They describe navigating multiple, sometimes competing policy imperatives, such as the National Improvement Framework discussed here and specific foci on ‘’improvements in literacy and numeracy’. Perhaps most revealing is the statement that ‘’there is nothing that says you have to do this’’. This reveals a space for professional judgement at the enactment level, although it follows a discussion of national performance metrics that may limit the extent to which any freedom is felt.

Another example provided of direct policy responses to international assessment outcomes was provided by Policy Influencer 5:

> And we saw in terms of Scotland, for example, I think some very ill-judged reforms to CFE introducing the benchmarks, for example, which happened as a result of the perceived performance in the 2015 PISA. (Scotland, senior academic)

The participant describes how the Curriculum for Excellence, originally conceived as a flexible, holistic curriculum framework, was subsequently supplemented with more prescriptive benchmarks following Scotland’s PISA results. This contribution highlights the influence of international large-scale assessments functioning as policy triggers, with quantitative indicators of an education system’s performance driving curricular and assessment reforms. The characterisation of these benchmark introductions as ‘’ill-judged’’ suggests a critique of reactive olicymaking that potentially undermines original curricular intentions. Both teacher participants and policy influencers identified tensions between the olicymaking process and between centralised government and teacher professional autonomy. Policy Influencer 4 argued for a fundamental restructuring of policy governance:

> We need a system in Scotland that is much more progressive system that keeps the curriculum under constant review. And that can only be done by using the expertise … of the practitioners who are on the ground, who can feed through a national agency, what's working well,

> and what's not working well. (Scotland, general teaching council CEO, policy advisor)

This participant appears to be proposing a critique of cyclical policy reforms, which may often fail to build upon previous iterations and create enactment fatigue among practitioners who are often excluded from reviews of policy success. More significantly, they centre practitioner expertise as a basis for informing assessment reform. These ideas were reinforced and then expanded upon by Dundee Teacher 2:

> They bring out stuff all the time, don't they? It's always changing. So, you've got CFC, SFEA, E's and O's and then you've got your benchmarks and all that. But again, nothing is really concrete. It's all just quite wishy washy. Progression frameworks and stuff like that. Yeah, they need to have actual teachers that are in the classroom informing these kinds of things and saying, ''well, can you tell me how I assess that pupil, so I know what level there are?''. (Dundee, experienced teacher)

The teacher participant presented a long list of policy acronyms, highlighting the proliferation of recent assessment policies in Scotland, potentially causing confusion and preventing successful policy enactment. They also call for teacher-led governance of assessment policy reform to enhance the design and enactment of policy with pedagogical realities.

In addition to a lack of understanding of the realities of assessment policy enactment, Scottish participants also critiqued policymakers' assessment literacy in general. Policy Influencer 5 identified the gaps in policymakers' knowledge:

> …a major challenge is actually, I think a poor understanding of what curriculum and assessment are amongst the people who actually make the decisions in government…. (Scotland, senior academic)

This lack of knowledge suggests that policy may be created by those without relevant expertise, leading to poor decision-making. It adds further support to calls for policymaking to be led or at least informed by practitioners. Policy Influencer 3 agreed with the need for practitioner influence in assessment design:

> It's not policy that changes classroom practice. I think it is, it's the individual effectiveness of leaders in schools… for example, the

> comparative judgement movement, and how effective that is, that is down to individual head teachers deciding they want to be more research informed. (Scotland, professional organisations and teacher professional development)

This suggests that policy should emerge from and be responsive to school-level innovation rather than be directed by national policymakers.

4.4.4.5 Professional judgement and trust

One area of contrast between teacher participants and policy influencer perspectives was in their views of trust in professional judgement. Policy influencers generally expressed greater confidence in teacher professional judgement. Policy Influencer 20 emphasised their belief in teacher assessment capabilities:

> So at the very least, to build teachers' self-confidence that actually they do this stuff, they can do this stuff…and it will stand up to, you know, I don't want to use the word scrutiny, but I will, moderation. Moderation and sharing thinking with others will validate what teachers already know and can do, and are doing, and what they will learn in the future. (Scotland, trade union representative)

This participant wanted to emphasise that teacher professional judgement could be relied upon and used with confidence. Teacher participants in the case study school, however, had an opposing view, preferring to defer responsibility for assessment to external assessments. Dundee Teacher 2 demonstrated such a view:

> My massive concern with that is that if they do remove exams, how can I be confident that how I'm assessing the pupils in front of me is the same as someone else in Glasgow, for example, is assessing pupils? How do I know they're the same level? Whereas having a national exam board does kind of ensure parity. (Dundee, experienced teacher)

Teacher participants focused on their ‘concern’ and the ‘burden’ of their role as assessors and seemed reluctant to accept responsibility for the design of assessment processes, opting instead to defer this to external standardised tests. There were signs from teacher participants that one influence for this was their negative experience during the COVID-19 pandemic, and how this had

caused them to be cautious of the potential impacts of this role, which will be discussed more in the next section.

Through these interviews with policy influencers and teachers, an emerging consensus emerged among policy influencers and teacher participants in the case study school around a balanced approach to involving teachers in the design of assessment policy. International assessment frameworks such as PISA appear to shape public and political perceptions of how assessment should be used, especially in the national context in Scotland. At a school level, teacher participants in the case-study school primarily described assessment pressures operating through national qualification frameworks. In these contributions, participants reflected that their perceptions had been influenced by experiences both before and during the COVID-19 pandemic, during which many changes to assessment policy led them to question the current state of assessment policy in Scotland.

4.4.5 Pandemic disruption and systemic vulnerabilities in Scottish Assessment Policy

As with the interviews conducted with England participants, the interviews with Scottish policy influencers and teacher participants in the Dundee case study school took place between January and June 2023. At this time, schools were open to students, although the impact of the pandemic was still being felt through staff and student illness, and assessment modifications were in place for national assessments.

The COVID-19 pandemic disrupted many aspects of Scotland's national assessment system, creating policy challenges and exposing tensions between traditional exam practices and teacher professional judgement. The interviews with Scottish participants revealed a divergence between policy influencers and teacher participants regarding the pandemic's impact, lessons learned and implications for future assessment policy.

4.4.5.1 Recognition of systemic inequalities

One impact of the COVID-19 pandemic caused both teacher participants and policy influencers to recognise and question systemic inequalities in the Scottish

national assessment systems. Policy Influencer 19 observed gender disparities caused by how national assessments were organised during the pandemic:

> When we closed the schools in March 2020, we'd had them all that time, but then they missed that March, April where they come in to support study and they really push in April before their May exams. And so, teachers' professional judgement depressed the boys' attainment. (Scotland, local policy leaders)

This contribution highlights the abrupt shift in teacher judgement of student assessment, enacted before widespread school closures, which could exacerbate existing inequalities between student populations. Some students, identified as boys by this participant, may be more motivated to make an effort toward their studies in the time immediately before their exams, and their perception is that these students would therefore not have had their full potential reflected in their work before school closures. Another conclusion from this contribution might also be that the pandemic highlighted how exams may suit students who are more likely to succeed in cramming their revision into short periods of time.

Policy Influencer 4 also highlighted potential issues with the traditional exam-based assessment models, highlighted through the COVID-19 pandemic. They recalled:

> I've also seen through pretty much all of my time in education, not so much assessment being well used but being misused. And I think that's one of the reasons why we have the system that we have just now, which struggled so badly during COVID, where it's very much attainment driven, examination driven, standardised assessment driven to the detriment I think of the learning experience that children and young people get, both in primary and in secondary. (Scotland, general teaching council CEO, policy advisor)

The participant's observation that the education system in Scotland struggled highlights systemic vulnerabilities in the ways in which all national assessments depend on the setting of synchronous examinations.

The sudden removal of exams that the pandemic necessitated halted the ability to assess students. The impact was also noted as something still predicted to affect students, as reflected by Dundee Teacher 8:

> So, for example, our current 3rd years into 4th year; they've been the ones who have been massively affected by COVID. I think more so than other years because they didn't have a primary seven, they didn't really a transition and they didn't have a full first year. You know, we kind of assessed where they all are and as a result, have planned more vocational qualifications in 4th year to allow them to achieve that level 5. (Dundee, school leader)

This participant describes a strategic deployment of vocational qualifications in their school, representing a significant curriculum adaptation in response to missed schooling during the pandemic. They justify this as a means to 'allow them to achieve that level 5', which raises questions about equity and the extent to which vocational qualifications might be equivalent alternatives or compensatory measures for perceived deficits in students' educational experience. It also highlights potential tensions in maintaining standards. The teacher participant's phrase of allowing students to 'achieve that level 5' suggests a commitment to achieving accountability markers for their students, utilising flexible credentialing routes.

4.4.5.2 Contested accountability and public response to emergency measures

The various methods used during the pandemic to measure student attainment were part of the commentary from policy influencers in Scotland. Policy influencers discussed the contingency measures put in place following the cancellation of national examinations, including their reception by the public and the media. Policy Influencer 21 discussed the reaction from the public following the ill-received and subsequent retraction of the national grading algorithms:

> I thought it was outrageous that the SQA got the blame during the COVID arrangements. Absolutely outrageous that that should happen, but it's always something else other than the system. There can't be anything wrong with the system. (Scotland, education leadership consultant)

This statement captures a contradiction in emergency educational responses. While COVID-19 necessitated rapid policy changes, the underlying assumptions about national assessment requiring examinations remained largely unquestioned. The participant recognises that the SQA functioned within such

restraints and was blamed for the public's rejection of the revised system, suggesting that meaningful reform requires an exploration of foundational assumptions about what is required for national assessments. In addition, the participant's defensive position on the SQA, whilst simultaneously critiquing systemic inertia, suggests a recognition that such operational bodies may become proxies for broader policy tensions, such as the debates over the use of certain types of assessment for grading student attainment during the COVID-19 pandemic.

Policy adaptation during the COVID-19 pandemic appeared crisis-driven. Policy Influencer 15 suggested a sense of frustration with governmental decision-making at this time:

> So we had COVID and all the sort of rigmarole about how we're going to measure achievement of kids. What changed the government's mind about how they had planned to do it. Young people. Not the profession of teachers, not any guidance from anybody that would see themselves to be important, but it was actually young people that influenced that change through protest and taking to the streets. (Scotland, general teaching council CEO, professional organisations and teacher professional development)

This quote highlights youth activism in Scotland protesting the algorithmic process used to assign centre grades. Their emphasis that 'young people' rather than 'the profession of teachers' or other established authorities, underscores these events as a significant moment of departure from top-down educational governance of national assessment. This statement's reference to 'rigmarole' suggests a sense of frustration with the ways in which policy responses are felt to be arbitrary and insufficiently imaginative to manage the crises at the time, without departing from the traditional exam-based ranking of student performance.

4.4.5.3 Emergent professional questioning of assessment orthodoxies

Despite various parts of the education system in Scotland working to restore the exam system during and following the COVID-19 pandemic, some professionals continue to question the suitability of the exam-based system. Dundee Teacher 4 in the Dundee case study exemplified this:

> Is an exam the best way to assess our subject? In many ways, I feel that given the whole kind of juxtaposition that COVID threw into the mix, it did allow us to question this. In my opinion, just having had our own exam on Monday of last week -- it's a two-and-a-half-hour exam which makes it one of the longest exams on the exam diet - for many of our pupils, it's an exercise in memory rather than knowledge. (Dundee, experienced teacher)

This participant's characterisation of the pandemic as a 'juxtaposition' appears to signal a perceived contrast between established assessment practices and the alternatives that emergency conditions made visible. Their observation that a two-and-a-half-hour exam constitutes 'an exercise in memory rather than knowledge' reveals a practitioner questioning whether assessment can measure genuine understanding, a distinction that the pre-pandemic policy environment had rendered largely academic. This participant felt there was no opportunity to explore previously unquestioned assumptions about exam-based assessment in olicymaking. Dundee Teacher 8 also highlighted how they felt a sense of optimism following student feedback on the changes to assessment during the pandemic:

> I mean pretty much every pupil you speak who has gone through COVID and has done their ongoing assessment rather than their exams at the end of the year, far prefer that model, so I would suspect there'll be a lot of happier pupils if we get that right! (Dundee, school leader)

This quote suggests that, for some students and school professionals, there was a sense of optimism in the wake of the pandemic that changes to national assessments might consider the positive impacts of moving away from exam-based assessment and towards the use of broader assessment modalities.

The COVID-19 pandemic exposes systemic vulnerabilities within Scotland's external assessment processes whilst simultaneously creating contested spaces for reimagining national assessment practices. Whilst policy influencers and teacher participants converged in recognising pre-existing inequalities, particularly gender disparities and structural reliance on synchronous examinations, their divergence emerged over the implications for future reform. The algorithmic grading controversy and subsequent youth activism demonstrated how emergency measures catalysed broader questions about assessment validity and equity. However, despite practitioners' expressed

optimism about alternative assessment modalities and student preferences for ongoing assessment approaches, the evidence suggests limited systemic transformation, with institutional momentum favouring restoration of traditional exam-based protocols. This pattern reveals a characteristic tension within assessment policy, the recognition of structural limitations coexisting with institutional resistance to fundamental reform, even when crises create opportunities for meaningful change.

4.4.6 Assessment reform as contested policy space

Despite the policy pressures and accountability mechanisms documented thus far, participants suggested that the COVID-19 pandemic offered an opportunity to question conventional ideas about national assessment, representing what some characterised as a critical juncture for policy reconsideration.

4.4.6.1 Calls for diversified assessment pathways

One type of suggestion centred around the need for a future consideration of diverse assessment methods, such as Policy Influencer 26 who critiqued the narrow focus:

> The number of people around that table at the Education Committee who specified that external assessment was the only way that you could actually get through assessment, that written exams were the only valid assessments. (Scotland, Former CEO of SQA, policy critic)

This participant was frustrated by the lack of support for considering alternative forms of assessment and wanted a more open approach to how achievement might be measured at the end of secondary education in Scotland.

4.4.6.2 Professional judgement and external moderation models

This approach to change was not consistent amongst policy-influencers, however, and Policy Influencer 19 advocated for a balanced approach to reform:

> I think the answer lies somewhere between the teacher's professional judgement and the final summative assessment… So, I think that's why it's a very appropriate moment for us to revisit and if you look at the assessment in a university - it's internal. It's internal assessment,

> externally moderated. So why can't we do that in schools? (Scotland, local policy leader)

This participant is advocating for a type of assessment reform that encapsulates the professional judgement of teachers alongside some method of moderation to ensure consistency in their judgement.

Teacher participants in the Dundee case study were also optimistic about the capacity for changes to assessment procedures. Dundee Teacher 8 articulated their desire to remove the notion of judgements based on single-day examinations:

> young people's futures, their careers, their lifestyles are all based - if they're all going to be based on one day, that's something I don't want to be... I think we should get rid of that, that's old-fashioned and really not appropriate for the aims of the Scottish Government. (Dundee school leader)

The participant's characterisation of single-day examinations as 'old-fashioned' and 'not appropriate' reveals an assumed alignment between their reform preferences and the Scottish Government's modernisation agenda. However, this assumption warrants scrutiny given the evidence presented earlier regarding the persistence of accountability-driven assessment practices.

4.4.6.3 Structural impediments to reform

Despite many calls for change, not all participants were optimistic that this was possible. Policy Influencer 3 observed the difficulty with current connections between assessment and accountability:

> Can you actually uncouple assessment from accountability? I'm not sure that you actually can really effectively. (Scotland, professional organisations and teacher professional development)

This participant succinctly highlights many cultural and systemic features of the current education system in Scotland, discussed throughout this chapter, that may frustrate ambitions for change in how assessment processes are used to measure student attainment and progress. While the pandemic has been cited by participants as a potential hinge-point for decisions over how assessment processes are purposed in Scotland.

These divergent perspectives reveal assessment reform as a fundamentally contested policy space, in which the pandemic served as what participants termed a potential hinge point for reconsidering established approaches. Whilst some policy influencers and teachers advocate for diversified assessment methods or hybrid models incorporating professional judgement, others question whether the structural coupling of assessment and accountability can be effectively disrupted. The evidence suggests that, despite widespread acknowledgement of the pandemic as an opportune moment for reform, participants recognised substantive systemic and cultural barriers that might frustrate ambitions for transformative change. This tension between reform possibility and institutional constraint frames the subsequent analysis of how policy actors work through these competing pressures in practice.

4.5 Chapter summary

This chapter has explored the interplay between research, policy and practice through a comparative analysis of assessment policy in Scotland and England. Drawing on semi-structured interviews with 26 policy influencers (13 from each jurisdiction) and 20 teacher participants from the Liverpool and Dundee case studies, four significant themes were generated through reflexive thematic analysis, illuminating divergent approaches to evidence-based policy formation.

First, policy intentions were grounded in different evidence hierarchies. Whilst both jurisdictions privileged high-stakes assessment data as robust justification for intervention, England and Scotland demonstrated markedly different approaches to incorporating socioeconomic indices and alternative forms of evidence into policy discussions.

Second, cultural factors emerged as powerful influences of policy enactment. England's competitive educational culture incentivised behaviours that participants identified as potentially counterproductive to policy intentions. Conversely, Scotland's collaborative ethos, whilst generally viewed positively, was perceived by some participants as creating accountability challenges that impeded systematic improvement efforts.

Third, international assessments, particularly PISA, functioned as malleable policy instruments rather than objective performance indicators. Both jurisdictions selectively interpreted and cited international evidence to support pre-existing policy agendas, with notable differences in which comparator countries were deemed exemplary and how national performance was contextualised.

Finally, the relationship between policymakers and school professionals revealed contrasting models of policy engagement. Scottish participants described experiencing both collaborative invitation and professional marginalisation, suggesting tensions between aspirational inclusive policy rhetoric and operational enactment realities.

Taken together, the findings indicate that the relationships between research, policy and practice in assessment are mediated by national context in ways that resist a simple, linear model of evidence-based policy formation. The next chapter applies the theoretical resources introduced in the literature review to interpret the findings. It considers what these findings suggest about wider patterns of policy enactment and about the influences over what counts as evidence in contemporary assessment policy.

5 Discussion

5.1.1 Introduction

The previous chapter set out what teachers, school leaders and policy influencers had said about assessment in England and Scotland. It deliberately stopped short of interpretation. The task of this chapter is to take those accounts and ask what they tell us, together, about how assessment policy is being made, enacted and experienced across the two jurisdictions.

Three patterns recur strongly enough across the interviews to warrant separate treatment. The first is that policy influencers in both countries acknowledged the costs of accountability-intensive assessment policies and chose to act regardless. The second is that the cultures of accountability in which these decisions were made differ markedly between England and Scotland, and these differences shape which actors are trusted to mediate policy. The third is that teachers in both jurisdictions filter contemporary reforms through assessment identities formed under earlier policy regimes, generating responses that policymakers tend to read as resistance rather than as inheritance.

This chapter takes those three patterns in turn. It draws on the policy influencer interviews and the case study material together, treating each strand as a check on the other rather than as a separate body of evidence. The accounts of teacher participants in Liverpool and Dundee are read as illustrative of work within their respective national policy environments and not as representative of all schools in either country.

5.1.2 Policy actors deploy accountability mechanisms, with awareness of detrimental consequences, for political expediency

The evidence from interviews with policy influencers suggests that policymakers selectively deploy research evidence to legitimise predetermined policy trajectories for both Scotland and England, while being aware of the potential detrimental consequences of their chosen approaches. This section examines how research findings and practitioner voices are appropriated to justify accountability-driven reforms that policy influencers themselves recognise as

problematic. The analysis of interview data reveals that political expediency, rather than pedagogical evidence, drives the continuation of assessment policies that systematically constrain teacher autonomy and erode professional well-being. This selective deployment has changed how assessment is used in both jurisdictions. Accountability functions have come to override the formative and learning-focused ones. The evidence in this chapter illuminates how these policy decisions materialise in school settings. Understanding their consequences requires an analysis of the complex relationship between academic research and education policy, as well as the mechanisms through which both countries have constructed accountability frameworks that diverge from research recommendations.

Recognising the complex relationship between academic research and education policy, policy influencers observed the selective deployment of research findings to legitimise pre-determined policy trajectories. In England, external examination data and school performance metrics such as Progress 8 serve as mechanisms for establishing institutional accountability and at the same time fostering school competition. The policy foundations for this approach were established through the Education Reform Act 1988 (c. 40). During the period of review in this study, think tanks such as Policy Exchange provided empirical support for competitive school models. Their 2015 report ‘A rising tide’ (Porter and Simons, 2015) presented evidence suggesting that the establishment of new Free Schools generated competitive pressures resulting in improved academic performance amongst surrounding institutions identified as underperforming. However, despite these conclusions from such free-market think tanks, research by Ball (2003, 2015, 2021) has repeatedly challenged the view that competition between schools, which is promoted through external exams, leads to genuine improvement in education. Similarly, Scotland demonstrates expanding utilisation of metrics by policymakers designed to monitor and compare attainment outcomes between secondary cohorts and evaluate the relative performance of local authorities (Cowie and Croxford, 2007). This expanding data collection regime was consolidated into national policy in 2015 when the Scottish Government introduced national standardised assessments for four year groups (Primary 1,4,5, and Secondary 3), stating that the aim is to track

attainment gaps for students from low- and high-socioeconomic backgrounds (Scottish Government, 2015).

Scotland's accountability mechanisms have attracted criticism from national politicians, who argue that the delegation of authority to local authorities, professional bodies, and national agencies constrains meaningful scrutiny of both national policy effectiveness and individual performance (The Scottish Parliament, 2019). Yet the evidence suggests that this same devolved structure reflects and sustains what Baumfield et al. (2010) characterised as a distinctive consultation culture, in which extensive professional engagement is embedded within the processes of policy formulation. Scotland's "horizontal logic" of professional networks thus operates simultaneously as a source of political frustration and as a foundational feature of its approach to educational governance.

The findings presented here from participants add further empirical evidence to support a recurring theme in the literature that assessment processes in the secondary stage of education are often used for multiple purposes (Newton, 2007). These themes are discussed in the literature review (p. 19). Whilst the idea of using assessment for multiple purposes has been defended (Newton, 2010)other authors have critiqued this approach. For example, the use of external examinations to create qualifications for individual students, whilst simultaneously using examination data as a proxy for judging the quality of education (Biesta, 2009). School participants highlighted a number of ways in which assessment data, especially external exam data, is used for accountability and changing behaviours:

> I think where schools, over the last number of years have gone wrong with that is they've used assessment as a purpose to create data and this kind of vast volume of data, don't get wrong, we're as guilty as any school to create this vast volume of data; five tracking points across a year to create all these data drops which, the whole purpose in the assessment was just to create data rather than actually the sole purpose, which is what have the students learnt? What haven't they learnt? And what misconceptions they have picked up on the way so we can intervene correctly with those students. (Head Teacher, Liverpool)

> You know, a few years ago, we weren't allowed to put in [for examination] young people who were doing National 5; we were told not to gather evidence for them in case they failed the National 5, so that we could submit for National 4. We were told that's not allowed. Strangely, it became clear that Dundee was the only authority that went down that route. We questioned it and queried it through SQA coordinators. (Head Teacher, Dundee)

This evidence indicates that policymakers' dual use of national qualifications for accountability and certification purposes can incentivise school professionals to adopt unexpected and potentially counterproductive practices to meet attainment-based metrics (Stobart, 2008). The desirability of achieving certain levels of attainment can lead to a blurring of what is valuable and valued in learning (Ball and Olmedo, 2013), creating multidirectional pressures on education professionals (Black, 1997).

In Scotland, academics and government advisors were pessimistic about the current extent to which data could be used to ensure improvement in education, as well as the lack of data infrastructure:

> First and foremost, Scotland is quite poor in terms of its data. It is very difficult to find out whether the education system is improving or getting worse. The only measures that are historically reviewed have been Highers, National Fives and the things that are announced by the SQA in the summer and that is a small portion of what happens in Scotland's education system, and it is a skewed portion, and it should be changed. (Scotland, policy advisor and civil servant)

Scottish policy influencers identified both methodological and practical challenges within existing frameworks, observations that, taken together, point towards the need for substantive reform.

In England, the freedom for schools to remove themselves from local authority control and attain academy status has encouraged localised competition for student numbers as politicians from New Labour and Conservative governments alike encourage schools to seek this 'autonomy' (Gibb, 2015b). With this autonomy, however, has come an increased focus on accountability, with both Michael Gove, the Secretary of State for Education and Nick Gibb, the Schools Minister, agreeing that a mantra of 'high autonomy, high accountability' would

be essential to deliver their policy aims against potential resistance from ‘orthodoxies’ in the education system (Gibb and Pearl, 2025, p. 35).

Despite Scotland traditionally resisting England’s model of academies and free schools, a critical turning point occurred with the 2015 OECD policy review, which explicitly recommended increased accountability mechanisms in Scotland, whilst simultaneously warning of unintended consequences from such processes (Hayward, 2018). Following this, the Scottish Government announced the introduction of national standardised assessments for school students in Scotland, justifying this with:

> We need more data on children’s progress in primary school and early secondary to ‘know the gap in order to close the gap’ and pinpoint where more intensive improvement is required. (Scottish Government, 2016, p. 4)

The findings from this work suggest that Scotland and England have engaged in similar actions to promote the use of data produced by external tests and examinations as a motivational tool and method for tracking improvement in education. The National Improvement Framework, in Scotland, published in 2016, provided further opportunity to create systematic accountability through the launch of data collection systems to compare the performance of groups of students.

The interview data presented here provide evidence consistent with earlier literature documenting increases in the use of external examination data for accountability and national quality assurance purposes in both England and Scotland. Leaders in both jurisdictions have asserted this data collection as a driver for educational improvement, despite research warning of the potential for negative consequences when assessment data are used for multiple purposes in this way (Stobart, 2008).

In Scotland, responding to OECD recommendations for improved data collection and use in schools (McIlroy, 2018), national policy was developed to expand the extent to which practitioners were required to generate assessment data for accountability purposes. The evidence from interview participants in this study suggests that this policy direction has produced a paradoxical outcome. Despite

substantial increases in both the volume of data created and the time invested in its production, the interview data indicate that this information has not served its intended purpose of adequately informing headteachers and local authority leaders about the quality of learning in Scottish schools. Furthermore, the predicted negative consequences cautioned against in the research literature (Childs and Baird, 2020) appear to have been realised. These include increased teacher workload and an intensified focus on examination performance in the senior secondary year groups that, according to participants, has occurred to the detriment of curriculum coverage and pedagogy responsive to learning needs.

In England, there also remains criticism of the ability of political leaders to adequately synthesise and understand assessment data at a local or national level to allow for evaluation of educational standards.

> Data literacy in schools and in local authorities is decidedly a mixed picture. There is, I mean, it's not just the danger, but we know it happens, people do take data and run with it and say, look at say, for example, value-added data or something and don't know that that's a model, and there is a confidence interval around that, and that's not an absolute number, etcetera. (Senior academic, Former national leader governmental)

Although participants in this work, both in school and at a policy influencer level, suggest that there are attempts to refocus attention away from exam data and towards a renewed focus on curriculum, the threat of accountability based upon exam results remains pervasive. This indicates that once practitioners and leaders become fixated with collecting exam data to measure the quality of education, it can be difficult to shift this focus onto other indicators that may be more qualitatively judged and more complex to analyse. School report cards in England were proposed as a potential replacement for league tables in 2010; however, a parliamentary inquiry raised concerns that indicators based on surveys had methodological issues (House of Commons Children, Schools and Families Committee, 2010). Standardised assessment currently serves as a necessary mechanism for educational quality in England, a function that emerges from the market-oriented structure within which schools operate. The evidence suggests that this quasi-market environment, characterised by inter-institutional competition for pupil enrolment, generates considerable demand for comparable

performance metrics to evidence school improvement (Gewirtz *et al.*, 2021). Such standardised measures are intended to enable parents to evaluate and differentiate among educational institutions, thereby facilitating informed school-choice decisions (Ball, 2003).

On this evidence, policy influencers in both Scotland and England describe explicit awareness of the consequences of accountability-driven assessment reforms, and yet these policies continue to be deployed and intensified. The accounts considered here show a consistent pattern. Research evidence and practitioner testimony are drawn on selectively to legitimise politically expedient reforms, even where the same accounts document the constraints those reforms impose on professional practice. Whilst the specific mechanisms differ between jurisdictions, with England's marketised system contrasting with Scotland's consultation-based approach, both contexts demonstrate convergent patterns of policy persistence despite acknowledged awareness of negative outcomes. This finding challenges simplistic explanations that attribute policy-practice gaps to insufficient evidence or lack of awareness, suggesting instead that political imperatives override pedagogical concerns. To understand why enactment continues despite this awareness, it is necessary to examine the cultural and institutional constraints that shape assessment policy reform in both national contexts.

5.1.3 Divergent accountability cultures in England and Scotland shape differential trust in policy mediators

Cultural contexts and institutional structures differ markedly between Scotland and England, and these differences shape which actors are trusted, held accountable, and empowered to mediate assessment policy. The accounts that follow consider how teachers' professional formation under earlier policy regimes interacts with the contemporary accountability cultures of each country.

In England, there were explicit political references to the oppositional powers exerted against policy change, with some even adopting a military metaphor. In 2012, the Conservative Education Secretary Michael Gove gave a speech on the intended academy changes, where he labelled potential resistors of this change

the ‘enemies of promise’ and ‘enemies of reform’ (Gove, 2012). Since then, Nick Gibb, the former Conservative leader, has identified ‘redoubts of the progressivist ideology prior to 2010’ as local authorities, government quangos, and universities (Gibb and Pearl, 2025, p. 33). By using such combative language, both Gove and Gibb seek to assert that educational professionals with a passion for reforming education and securing educational improvement should embrace their reform ideas and take pride in being part of them, rather than being members of the opposition groups identified.

In Scotland, attempts to enact education policy have been complicated due to the complex interface between research, policy and practice (Spencer and Hayward, 2016). National politicians in Scotland have complained about the difficulties in enacting national policy that could be set at a national level, but then had to be ‘left in the hands of local authorities, who didn’t always want to play ball’ (Sturgeon, 2025, p. 346). Given that local authorities hold statutory responsibility for educational provision (The Scottish Parliament, 2019), national governments face the imperative of influencing local professional networks to facilitate school-level transformation, reflecting broader tensions in multi-level governance structures.

Both Scotland and England appear to have cultural and systemic features that frustrate the enactment of education policy. These features were also highlighted by policy influencers, providing data that revealed similarities and differences between the constraints for both countries. Scottish participants described a more cohesive resistance to change, whilst English participants described ideologically driven tensions between political parties and professional organisations such as unions.

Cultural resistance to change in Scotland was characterised by an alliance among unions, professional organisations, local authorities, head teachers, and parent bodies. (Former leader, middle tier, Consultant) termed this an ‘unholy alliance’. This same participant outlined how they saw this operating in Scotland:

> The unions will come in; when there are concerns, there is actually anything that starts to potentially change or might have implications

> longer term, even crazily, in terms and conditions. Let's just stop that, let's close it down. They will then put pressure on to local authorities, local authorities -- you get a few councillors starting to talk and then schools and Head Teachers will also manipulate parent bodies in saying, you know this isn't good for kids... and before you know it you've got this tidal wave of objection of anything that looks remotely radical (Scotland, Former leader middle tier, Consultant).

Scottish policy influencers also identified a cultural constraint within Scotland, whereby any deviation from uniformity across schools is viewed with intense suspicion under a rhetoric of preventing a 'postcode lottery'. This participant described an illustrative incident where someone challenged an educational intervention:

> …if your experiment works, you're disadvantaging young people who have not had the chance to take part in your experiment,' to which I quite frankly exploded - in terms of thank you very much for identifying what is wrong with Scottish education and it's the, if you want a shorthand for it, Mike, the predominant language is postcode lottery and how that translates itself in Scottish education is every school needs to be identical in terms of what it offers. (Scotland, Former leader middle tier).

Beyond these themes, policy influencers characterised Scotland's education culture as one in which teachers exercise professional autonomy with limited accountability mechanisms. These findings are consistent with the research literature, which describes how educational reform initiatives in Scotland encounter resistance from multiple stakeholders, including teachers, local authorities, and professional bodies (Gillies, 2018; McGinley, 2018). Policy influencers in England described a far more negative relationship between political parties and teaching unions, although some viewed this as a healthy tension. They described how ideological differences led some policymakers to perceive teaching unions as adversarial rather than collaborative partners in an education system. However, this perspective appeared to coexist with recognition that professional organisations provided a necessary counterweight to consolidate professional opinion:

> There was, on the Conservative Tory side, a suspicion of the kind of left-wing leanings of the teaching unions and so on. So, some of them regarded teaching unions in particular as natural enemies rather than allies in the system… I found that to be really helpful as a counterbalance to the kind of conglomerate of professional opinion,

> which are sometimes divided, but very often there was quite a strong consensus amongst teachers, for example, teacher organisations... I think it was a healthy counterbalance to the kind of takeover of professionals in teaching, who got into unions or into subject disciplines. (Former leader middle-tier Consultant).

While this tension was evident across both jurisdictions, the data suggests that English participants tended to characterise it as combative more frequently than Scottish participants. The evidence indicates that education operated as a politically sensitive policy instrument in the Scottish Parliament, with politicians appearing acutely conscious of the potential reputational risks associated with unsuccessful reform attempts:

> We were invited into the DfE to advise them about what the key element would be in improving student learning, and we just said, you should give grades of schools less often, and the reaction of the Civil Service was, we can't do that. The Daily Mail wouldn't like it (England, Senior academic Consultant).

In response to pressure to improve the national data on school performance, another participant described their encounter with a Prime Minister who was desperate to see external assessment data show an improvement:

> The plateau occurred in my first year in office, and one of my jobs was to have to sort of go and present the data for the performance of the school system, every September, to the Prime Minister and the Cabinet. So, you rock up there, this is the first time I've done it. You rocked up there with your PowerPoint slides and you wander into the cabinet room at Number 10, with him and his people arranged against you and, although I got on quite well with the Blair, he wasn't very happy because of the plateau effect and I don't think, if you're going to record this, I won't actually use the words he used. Well, he said, with a degree of exasperation, and he sort of leaned across the table, began banging his fist on the cabinet table and then pointing at me, '*Name*, *Name*, you've got to f'ing micromanage those primary schools, alright?'. And I said it was one of those moments in your life that you have to think well, how do you respond? How do you respond to this? (England, Senior academic, former national leader, governmental).

These quotations illustrate how anticipated media reaction in England functions as a direct veto on policy reform, even when expert advisors recommend changes that would benefit student learning. In both England and Scotland, although takes form differently, media influence creates powerful constraints on

educational policy change by making policymakers fearful of public and media reaction, even when convincing recommendations from research evidence are present.

In particular, with assessment changes, Policy influencers described how public attitudes, particularly regarding traditional qualifications and accountability, constrain innovation. National examinations in both countries were seen to inform parental expectations. They described deeply embedded cultural preferences for traditional qualifications as markers of educational success. These expectations were portrayed as both culturally embedded within societal values and reinforced through institutional accountability structures such as school inspections. There was also a suggestion that in England, parental attitudes may not be shaped solely by educational philosophy, but by a pragmatic self-interest, sometimes even to maintain house prices:

> So if your [the school's] data are good, you're likely to be judged favourably in inspection, and if it's bad, you're judged negatively. The repercussions from that are that you become a school that, if you're successful, your school is one to which parents want their children to come, and if you're a struggling school, in many cases, parents don't want their child to come to your school. So, the league tables and the inspection outcomes become a good indicator of house prices… (England, Former national leader governmental, Policy advisor).

The differences between Scotland and England suggest that cultural and institutional contexts shape not only what constraints exist, but also how different stakeholder groups interact to enable or frustrate educational policy change. These findings contribute to the empirical literature by demonstrating how educational change is differentially constrained across Scotland and England through multiple interrelated cultural factors. A particularly notable finding was that teachers did not identify national policy as a source of constraint, suggesting that whilst they experience the consequences of policy decisions, they may lack visibility of, or direct engagement with, higher-level political and organisational dynamics that shape educational practice. The evidence is consistent with an interpretation that teachers' responses are primarily oriented towards the localised systems that create and maintain these constraints. This interpretation may account for why participants in policy-influencing roles predominantly described the voice of the teaching profession as being

channelled through unions and subject associations rather than through individual teacher agency.

This analysis of the differential ways in which national contexts shape educational policy enactment provides an essential foundation for examining how international comparisons themselves are mobilised within policy discourse. The following section explores how international assessments are selectively deployed by politicians to legitimise particular ideological positions concerning the purposes and functions of assessment processes.

5.1.4 Politicians selectively appropriate international assessment data to legitimise ideologically driven policy reforms

International assessment data, in particular PISA results, occupy a strange position in the policy accounts examined here. Teachers in Liverpool and Dundee did not mention them. Policy influencers in both jurisdictions did so frequently. This raises the question of what work PISA does for politicians that it does not do for teachers. The accounts considered in this section indicate that PISA is used as a resource for political argument rather than as a guide to assessment design. Politicians in England and Scotland have drawn on the same PISA results to justify opposite policy directions, with England using them to argue for more inter-school competition and Scotland using them to argue for greater teacher autonomy. Both readings cannot be correct as interpretations of the data. Both can, however, be useful as political claims, and the section that follows considers how that usefulness operates.

International assessments such as the PISA tests produced by the OECD were discussed by national policy influencers as having a direct impact on how assessment processes were designed in both England and Scotland. In contrast, teachers in the case study schools did not mention international assessments as an influencer over school assessment processes; rather, they cited school inspection frameworks and national policy changes as having an impact on their practice. However, the changes and influences they did note, such as increases in inter-school competitiveness, are discussed by policymakers as a policy associated with PISA rankings. These examples, specific to each country, are discussed later in this section.

OECD argue that PISA assessments were originally designed to generate insights into how countries were preparing students: 'to become lifelong learners and to play constructive roles as citizens in society.' (OECD, 1999, p. 7). The OECD designed PISA with the intention that these insights would be used by governments and schools to inform educational policy reform and monitor the effectiveness of their education systems at a national level. However, a number of authors argue that PISA data are used in ways that are inconsistent with the original OECD aspirations. For example, Crehan (2016) and Ball (2021) contend that the data are frequently employed to borrow policy and adopt practices from one country and deploy them in another without adequate consideration of cultural differences that might limit the success of policy transfer. Humes *et al.* (2018) further demonstrate that PISA data can be exploited by opposing politicians as an opportunity to critique government performance. The evidence suggests that England and Scotland have enacted divergent policies in response to their PISA performance through politically motivated interpretations (Grek, 2009).

The differing policy responses across these neighbouring systems show how political ideology and evidence-informed assessment reform pull in different directions when international performance data are translated into national policy. For example, in England, political arguments for improving league table position through the increase of accountability measures to make teachers and schools more accountable for their students' results (Sturrock, 2024). In Scotland, the same drivers are used to justify greater teacher autonomy and professional judgement (McMahon, 2018).

Moving beyond these general patterns of policy borrowing and political interpretation, interviews with policy influencers in Scotland provided multiple examples of how international assessment data, especially PISA data, have been instrumentally deployed in the service of broader political aspirations. Consistent with examples from existing literature on the uses of international assessment (Gale and Parker, 2018; Milner and Ydesen, 2025), interview participants identified where senior policymakers within Scotland's education system strategically positioned PISA data as a key mechanism for demonstrating educational improvement and national education effectiveness:

> I think for me, what the PISA assessments have done and the sort of broader discursive sort of field around me, the OECD and similar organisations is that it has, I suppose what it's done is it's created an environment where countries have to compete against each other. You know, the PISA shock is a real phenomenon that countries experience. (Scotland, Former leader non-governmental policy advisor)

This provides evidence for Scotland's motivation to engage with PISA, serving the dual purpose of legitimising educational policy interventions while simultaneously asserting educational autonomy within the broader constitutional framework.

Within England's policy environment, interviews with policy influencers demonstrated alignment with contemporary literature on the borrowing of policies from countries identified as high performing in PISA assessments. However, this approach to policy borrowing raises critical questions about whether educational practices can be successfully transferred across different institutional and cultural contexts. The PISA ranking of England was proposed by Policy Influencers to have led to policy borrowing from countries where competition was encouraged. They proposed that politicians wanted to borrow policies from countries such as Hong Kong and Singapore to increase competition between schools, in a bid to improve education in England. The themes of how data from international assessments were raised often by policy influencers, but were surprisingly absent from interviews with school practitioners in both Dundee and Liverpool.

Although teachers in Dundee did not explicitly reference PISA's direct policy influence, their perspectives aligned with those of policy influencers regarding the role of accountability frameworks in shaping classroom assessment practices. Teachers identified systematic pressure from local and national authorities to intensify data-driven assessment processes, recognising these mechanisms as tools for comparisons between students, schools and local authorities. Examples of how this pressure operated in the classroom included teachers creating spreadsheets of student performance in individual questions and larger end-of-year or mock external assessments:

> So, we tend to have a large spreadsheet that we start at the beginning of the year, and we track all the assessments we conduct. What I tend to do with these spreadsheets as well is I break them down into individual topics or questions and then I'll traffic light these, which means that I know there's stuff that perhaps I haven't taught and I need to go back because I can see that the majority in the class haven't understood that or I can help kids identify their individual points that they need to work on. So, if most of their assessments are green, but they've got a red in this one and a red and that one it means that I can then have that learning conversation with them to say ‘you really need to work on this aspect of the course' and we'll do that throughout the year to track progress and to see how they're getting on. (Dundee, experienced teacher)

This is then later used in attainment meetings between the Head Teacher and the Education Officer for the local authority to evaluate the school's performance. This data-centric approach creates tensions between external accountability demands and established professional cultures, as participants noted that collaborative networks among schools mitigate competitive pressures. While schools generate extensive performance data to meet accountability requirements, participants described how parents demonstrated limited engagement with these metrics, suggesting a disconnect between the prioritisation of assessment data and what is culturally valued in Scotland.

Teacher Participants in the Liverpool case study school expressed increasing institutional pressure to generate data demonstrating positive student progress aligned with performance targets. Although they acknowledge that recent reforms to national accountability frameworks have reduced the emphasis on such data, most notably through revised Ofsted inspection protocols, they described persistent institutional practices rooted in a legacy of data-driven protection strategies. These practices, originally developed to shield schools from adverse judgements through inspections and local communities, demonstrate the enduring influence of historical accountability pressures on contemporary school leadership decisions. These pressures can be traced to political discourse by English politicians advocating competition between institutions as a means to drive performance improvements (Stobart, 2008).

In political discourse, international assessment data are frequently recontextualised to legitimate education policies rooted in ideology rather than empirical evidence. This is exemplified by Gove’s tenure as conservative

Education Secretary in England, where international comparative data served as justification for school autonomy and competition (Gove, 2012). Significantly, policy discourse deliberately maintained ambiguity regarding the precise dimensions of educational 'performance' and their measurement criteria, while also positioning these undefined metrics as pathways to enhance PISA performance and educational improvement. This enabled their legitimisation while circumventing rigorous debate about the ideological assumptions of what is meant by improvements in performance (Milner and Ydesen, 2025).

In announcing curricular changes in 2011, Education Secretary Michael Gove leveraged PISA performance data to construct a narrative of educational decline, positioning England's international ranking as both symptom and justification for reform. Gove's discourse claims that:

> 'We have sunk in international league tables, and the National Curriculum is substandard. Meanwhile, the pace of economic and technological change is accelerating, and our children are being left behind'

exemplifying the use of comparative assessment data to legitimise radical policy interventions (Department for Education, 2011).

The existing literature on political appropriation of international assessment data suggests comparable patterns in the Scottish context. A notable illustration of this occurred when the 2009 PISA findings, initially cited by Michael Gove, were subsequently appropriated by Mike Russell, the SNP Cabinet Secretary for Education and Lifelong Learning, who argued that the PISA evidence demonstrated substantial improvements in Scottish education (Hayward, Macbride and Spencer, 2016). In addition to demonstrating improvement, participation in PISA for Scotland aims to legitimise itself as a member nation in its own right, joining international assessment as what has been conceptualised as a 'global ritual of belonging' (Addey, 2015, p. 158).

Scotland and England construct distinct national identities through strategic engagement with international comparative frameworks, selecting peer nations that align with their respective policy narrative and aspirational positions. Despite many divergences in their discourse and governance structures, both

demonstrate convergent reliance on intensified assessment regimes to generate performance data that they aim to validate educational quality. Although discussing organisations, DiMaggio and Powell (1983, p. 152) define this behaviour as mimetic isomorphism, whereby: ‘Organizations tend to model themselves after similar organizations in their field that they perceive to be more legitimate or successful.’ Through providing a ranking system for national education systems, the PISA assessment acts as a tool and data source whereby countries operationalise this isomorphism to adopt similar practices with assessment, despite their different interpretations of the data.

This work contributes empirical data to support theoretical understandings of how politicians adapt ideas from different contexts and time periods to match their rhetoric in their local circumstances (Milner and Ydesen, 2025). Whilst politicians in Scotland and England profess their divergent solutions to perceived shortcomings in the PISA data, this work has provided empirical evidence of how politicians use data from PISA assessments to justify policy ideology. This may operate through the creation of concern over a country’s relative performance and the selective borrowing of politically expedient policies from countries that happen to be highly rated in PISA rankings. The case study schools demonstrate how this national preoccupation with international rankings, passed down through the system from government to schools, compels schools to prioritise data production for external validation over pedagogically responsive assessment practices. However, the differential success of policy enactment in these two contexts cannot be attributed solely to political will or data availability; rather, understanding why certain reforms take hold whilst others encounter resistance requires examination of the cultural and institutional constraints that shape assessment policy enactment within each national context.

5.1.5 Teacher professional agency is constrained by temporal policy effects and professional formation under previous policy frameworks

Teachers do not encounter assessment policy as something new. They encounter it through the assessment identity they have already developed across their careers. That identity has been formed under the policy regimes that preceded the current one. This temporal dimension is worth stating before considering the

empirical material, because it changes what counts as resistance and what counts as inheritance. When teachers respond to a current reform by continuing to do what an earlier reform asked of them, this is not necessarily opposition to the current reform. It may be the working-through of an earlier one. The accounts that follow consider how this temporal dimension shapes the responses of teachers in Liverpool and Dundee to recent assessment policy in their respective systems.

These tensions intensify where assessment serves competing purposes. Within Scotland's Curriculum for Excellence, professionals reported experiencing 'double vision' when managing multiple demands (Baumfield *et al.*, 2010), whilst Hayward et al.'s (2012) Assessment at Transition report documented explicit conflicts between formative assessment purposes and accountability imperatives. Policy pressure exerts tangible influence on practice, exemplified by teachers who feel compelled to prioritise examination techniques over substantive learning, and by English schools generating substantial assessment data to second-guess how to improve their performance, following the removal of national curriculum levels (Millard, Small and Menzies, 2017). Biesta, Priestley and Robinson (2015) demonstrate that whilst teachers 'operate with their own belief systems as agents within a policy landscape', this agency remains constrained by cultures of performativity and accountability. Nevertheless, professionals exercise resistance through institutional mechanisms such as trade unions and subject associations. Looney *et al.'s* (2018) conception of assessment identity as 'dynamic and interactive… constituted by beliefs, feelings, knowledge, and skills' reinforces that professional practice represents active construction rather than passive policy reception, positioning agency as generative even within contested and constrained spaces.

The evidence from participants provided support on many levels for school professional agency operating under policy pressure. However, the evidence also revealed important nuances and tensions between the development of agency over assessment decisions, particularly during the COVID-19 pandemic. Both policy influencers and teachers described how accountability pressures cascade through the system from government to schools to departments to individual teachers, creating an environment where assessment serves primarily to

maintain institutional position rather than advance pupil learning. The evidence from Scottish participants reveals the mechanics through which national accountability systems transmit pressure through hierarchical educational structures, ultimately shaping classroom-level pedagogical decisions.

One English policy influencer described how this pressure passes down through the system from government to schools and into individual classrooms:

> because we have a situation where it's all about school choice... schools have their results reported, parents look at those results, and they choose their school based partially on those results... what happens is when you report results, Is you put pressure on the schools to maintain their position or improve their position within the family of schools that they sit in... which therefore, puts pressure on senior leaders to raise standards, but it puts pressure on individual teachers to raise standards (England, Former national leader Non-governmental Policy advisor).

This account demonstrates how performativity operates as a multi-layered government mechanism, consistent with Ball's (2012) theorisation of how high-stakes accountability produces particular forms of institutional and professional subjectivity. The participant's language reveals a conception of pressure as something applied at the system level that flows inexorably downwards, compelling responses at each institutional stratum.

However, this vertically oriented understanding of accountability represents only one dimension of the regulatory environment. Teacher 7's perspective expands this top-down narrative by emphasising the simultaneous existence of horizontal accountability structures within schools:

In the first instance, because all teachers are responsible for their own learning and their own classes, the class teacher delivering the curriculum area is responsible for a young person's progression. In a wider sense, they are operating within a departmental team or curriculum team. (Liverpool, School middle tier leader).

This perspective suggests a dual nature of teacher accountability, in which individual practitioners are responsible for specific learners while simultaneously functioning as members of collaborative teams.

Whilst the horizontal accountability frameworks appear to afford teachers a degree of professional autonomy in collaborative decision-making, the evidence from interviews suggests that such agency is fundamentally constrained by the external accountability requirements described earlier. These system-level pressures, operating through the vertical accountability mechanisms above, ultimately limit teachers' capacity to use assessment for formative, learning-focused purposes, which they often described as more of a personal priority.

Within Scotland, Policy Influencer 3 articulated how accountability structures constrain classroom autonomy, positioning external reporting requirements as barriers to responsive pedagogy:

> If I were in my classroom, I'm carrying out an assessment, I am assessing my pupils and thinking about what's next. And I don't have to report that anywhere; I can do what's right for the pupils. And I can see that progress has been made, and I can assess accordingly. But if I am being held accountable for the percentage of pupils that have reached a certain level, or for the amount of coverage of the curriculum, which is a common thing as well, then that influences what I am able to do in my classroom. So, having the pressure of accountability on a head teacher filters down to the classroom, and is a barrier to allowing a teacher to be able to see and have the confidence to do what the child needs. (Scotland, Non-governmental Policy advisor).

Meanwhile, teachers in the Dundee case study emphasised teacher responsibility and curriculum team collaboration without acknowledging the constraints described by some policy influencers. The divergence raises questions about whether Scotland's commitment to professional autonomy masks practical constraints or whether genuine variation exists between schools in how accountability pressures limit teacher decision-making. Both accounts, however, implicitly confirm that autonomy operates within predetermined curricular and accountability frameworks rather than representing genuine professional discretion about how assessment processes are used within the case study school in Dundee.

The English Policy Influencers and teachers demonstrated complete convergence in identifying severely constrained autonomy, whilst revealing different mechanisms of constraint. One Policy Influencer described how outcome-focused accountability transforms teaching into test preparation:

> Too often, accountability turns into answerability, so you're answerable for the results that you get and therefore, the danger is that you want the results to demonstrate whatever you're proposing and therefore, you distort the assessment in order to make that work. So, if you're a teacher in a classroom that is going to be measured on your outcomes at GCSE or SAT. They become the be-all and end-all of the assessment. Therefore, they become the be-all and end-all of the teaching, and you teach what's going to be tested rather than teaching children the science or the art or the mathematics or history or whatever it is, in order for them to become historians, scientists, mathematicians... You lose the integrity -- that is the key point. You lose the integrity of what you're trying to teach. (England, Former national leader governmental Policy advisor)

This policy influencer's contribution provides precise operational confirmation with Teacher 6's description of school-mandated 'blind' assessments administered on prescribed dates:

> Largely, in my sort of assessment-for-learning, I have a lot more autonomy over day-to-day lesson planning. When we're talking more, we call them rightly or wrongly, hinge assessments and then summative assessments, they're directed. So we are directed what we do, when we do it and on what day and they're given to us in the Academy, in English, at least. And they're blind to the teacher, so the teacher won't see them either, to try and ensure that they're fair. So, I'm not getting that assessment a week before and then just teaching to the question. So that's what happens; I'm not autonomous in that element (Liverpool, School middle-tier leader).

Teacher 6's account reveals how their school enacted an internal control mechanism that prevents teachers from even knowing assessment content in advance to prevent 'teaching to the question' but eliminates professional judgement about assessment appropriateness, timing, or pedagogical integration. The distinction Teacher 6 draws between 'assessment for learning', where they have autonomy, and summative assessment, where they do not, precisely exemplifies Policy influencer 11's argument about pedagogical integrity loss: teachers retain autonomy only in aspects deemed inconsequential to accountability, whilst losing it entirely where assessment serves system-level purposes. The convergence between policy influencer critique and teacher testimony evidences concerns that English teachers experience professional autonomy as illusory, existing only within parameters predetermined by

accountability demands that prioritise outcome standardisation over formative responsiveness.

The pressure to serve accountability purposes through assessment significantly affects teacher workload, professional satisfaction and career decisions, contributing to workforce retention challenges (Stobart and Eggen, 2012; Sturrock, 2024; Brady and Richardson, 2025). The perspectives of participants in Scotland showed strong convergence in identifying the psychological burden of assessment pressure, with notable convergence in what each reveals. Policy Influencer 24 provided an observation that senior phase pressure in secondary schools provides an accountability pressure:

> for secondary teachers, specifically in the senior phase, there is a huge amount of pressure on the teachers. There is also a huge amount of pressure on the young people. That is really, really challenging on a whole number of levels and the fallout of that… those high-pressure points that exist in the senior phase, that drives what happens in a secondary school. There's no point seeing anything else, it just does. (Scotland, Former leader, middle tier)

This observation that the upper years of secondary school 'drives what happens' across entire schools is also supported by Teacher 2 in the Dundee Case Study:

> There was a real fear because our results haven't necessarily been very positive in the past, I think there was a real fear of digging deep and actually seeing what the students did know or didn't know because there was a drive for just better data and I can see why that happens, because it is really difficult sometimes to have those conversations to be like 'but this is the point that they're at' and I think that's really important to not shy away from the fact that they don't know because you're not helping anybody there and you're not helping the students. (Dundee Experienced teacher)

The contributions from teachers like Teacher 2 above give an account of 'real fear' surrounding accurate student assessment. They provide an emotional testimony and reveal how fear manifests not merely as excessive workload but as ethical distress when professional honesty is in conflict with demands to improve school data. Both accounts confirm that whilst Scotland may lack some of England's explicit linkage to teacher retention problems discussed below, assessment-driven pressure is creating similar psychological burdens. This convergence is consistent with concerns that Scotland's senior phase examination

system produces wellbeing effects comparable to those in England, despite the differences in assessment and accountability arrangements between the two systems.

There was also convergence between the English Policy Influencer and teachers in the case study school. Policy Influencer 9 made strong connections between assessment policy, professional autonomy and teacher retention:

> Assessment and all of the trappings of assessment linked to accountability are the biggest drivers for change within the school system, conformity within the school system compliance. Lack of capacity for teachers to respond to individual needs. And I strongly think, although I haven't got any evidence of it, but I strongly think that links to teachers' tenure, you know, if I feel like I can't do what I want to do to make a difference, maybe it's not the job for me. And it does lead to workload. I mean, all of the marking and everything else that schools require is partly to do with assessing but mainly it's to do with being able to show that you're assessing. It's not actually necessarily anything to do with making a difference to children. It's all showing that you're doing what you're supposed to do. (England, professional organisation and teacher professional development)

This account makes a bleak link between professional efficacy and retention. Policy Influencer 9 reads workload pressures as a contributor to teachers leaving the profession. Liverpool Teacher 1 made the same connection from the classroom side:

> I think it's a national thing, it's not this school. Nobody wants to go into teaching anymore, and teachers feel extreme and immense pressure to achieve, you know, take students from ridiculous starting points and these made-up target grades to an extent... teachers are working god knows how many hours under the sun, that the students are almost expected to do the same, and I think teachers lose that... I think teachers lose that empathy and that kind of understanding of what these students have to do, what you're doing for one subject for 9 subjects. (Liverpool, school middle-tier leader)

Teacher 1’s contribution confirms Policy Influencer 9’s workload analysis that marking serves primarily performative rather than educational purposes, showing that you’re assessing rather than making a difference to children. Most importantly, they elevate Participant 9’s concern by identifying moral injury beyond mere professional dissatisfaction, noting that teachers lose empathy and understanding for pupils under excessive pressure. This suggests that

assessment-driven accountability degrades not just teachers' capacity to respond to individual needs but also fundamentally alters their professional identity and relational capacities. The way that Teacher 1 characterises target grades as ‘made up’ supports Policy Inluencer 9’s recognition of accountability metrics’ artificiality, whilst being compelled to pursue them. Further adding to the cognitive dissonance of what matters for teachers and the hopelessness that is suggested to be driving workforce departure.

This analysis provides evidence that accountability-driven assessment reforms systematically constrain teacher agency and erode professional morale despite explicit awareness of these detrimental outcomes amongst policy influencers themselves. The triangulation of national-level policy perspectives with case-study evidence revealed a troubling consistency in how policy shifts directly translate into material consequences for teacher workload, wellbeing, and professional identity. Critically, the contested spaces of assessment decision-making are fundamentally shaped by temporal effects, such as teachers' formation of their assessment identity under previous policy frameworks, which generate systematic resistance to contemporary reforms. As practitioners filter new initiatives through established beliefs and practices developed across decades of conflicting policy directives, this resistance persists. This professional inheritance operates differently in the two jurisdictions, in line with their differing accountability cultures. Scottish teachers respond to reform through collective professional networks. English teachers respond within vertical accountability structures that position them as subjects of oversight. The evidence suggests that continued enactment of accountability-intensive reforms cannot be attributed to insufficient awareness of consequences. Rather, it reflects political expediency overriding pedagogical evidence within contexts where professional agency is systematically constrained. Both contemporary policy pressures and historical positioning within evolving assessment regimes contribute to this constraint. Craske's (2021) account of populist logic in conservative reform, developed further in the Conclusion (p. 286), helps to explain how this constraint was achieved rhetorically. The conservative reform project did not simply implement particular accountability measures, but constructed a political framing in which professional dissent was characterised as obstruction. The findings of this study indicate that this rhetorical move was

effective in displacing professional knowledge from English policy formation across the study window.

5.1.6 Chapter summary

This chapter has set out four readings of the policy influencer and case study material, taken together. Each is summarised here, with the closing observation that the three readings are not independent of one another.

First, the evidence suggests that many assessment policies yield unintended consequences that policy influencers do not entirely foresee. The triangulation of policy influencer perspectives with teacher testimony provided empirical validation that explicit awareness of detrimental outcomes exists at policy levels, yet accountability mechanisms continue to be deployed and intensified across both jurisdictions. Policy influencers articulated a detailed understanding of how accountability pressures systematically constrain teacher autonomy, erode professional morale, and contribute to workforce retention challenges. Despite this awareness, reforms designed to intensify external accountability continued to be enacted, suggesting that political expediency overrides pedagogical evidence when assessment policies serve broader governmental objectives. This finding extends beyond documenting policy-practice gaps to examining why such gaps persist despite acknowledged awareness of their consequences. The evidence revealed that research findings and practitioner voices are selectively appropriated to legitimise predetermined policy trajectories, with international assessment data instrumentally deployed to justify ideologically driven reforms. Reckhow, Tompkins-Stange and Galey-Horn's (2021) account of the political economy of knowledge production, introduced in Chapter 2 (p. 35), helps to interpret this finding. Their argument that the provenance of evidence often outweighs its methodological quality is consistent with the patterns observed in this study. The selective use of PISA and OECD data across both jurisdictions illustrates the same dynamic operating in the assessment domain. Politicians in both Scotland and England recognised that their chosen approaches generate negative consequences for pupils and teachers; yet these policies continue because they serve politically expedient purposes, enabling performance measurement, facilitating institutional

comparisons, and satisfying demands for accountability for public funds, regardless of the pedagogical cost.

The selective use of research evidence described in the Discussion can be read more closely through two further interpretive resources, both of which were introduced earlier in this thesis. Butler's (2025) account of motivated reasoning offers a description of the cognitive process. Decision-makers do not simply ignore the available evidence. They engage with it, and they discount it where it conflicts with their existing policy preferences. Parkhurst's (2017b) account of evidence governance offers a description of the institutional process. What Parkhurst terms issue bias, in which a political preference for accountability evidence displaces evidence about pedagogical consequence, is consistent with the patterns described in the Discussion. Together, these two readings move the contribution beyond the empirical observation that policy influencers acted with awareness of consequence. They support an account of why that pattern is durable across changes of government and across the differing accountability cultures of the two jurisdictions examined here.

Barber's (2015) defence of accountability mechanisms, introduced in Chapter 2 (p. 7), offers a competing reading of the findings presented in this section. From his perspective, the structures that this thesis treats critically are precisely those needed to translate political intent into improved outcomes for citizens. The evidence in this thesis qualifies that account by suggesting that the pedagogical costs of accountability intensification in school assessment are systematically underweighted in delivery-oriented framings.

Second, this analysis reveals how temporal effects and professional inertia fundamentally shape policy enactment in ways that policymakers often fail to adequately account for. Assessment policy reforms must contend not only with teachers' immediate pedagogical concerns, but also with deeply embedded assessment identities forged through decades of evolving, often conflicting, policy frameworks. Teachers' professional formation under previous assessment regimes imposes systematic constraints on contemporary reform efforts, as practitioners filter new initiatives through established beliefs, values, and practices developed over the course of their careers. The evidence from both jurisdictions demonstrated how persistent institutional practices, rooted in

historical accountability pressures, continued to shape school leadership decisions even after formal policy frameworks changed. In England, schools have maintained data-intensive assessment practices despite revised Ofsted protocols reducing emphasis on performance metrics, whilst in Scotland, teachers' responses to reform initiatives reflected professional cultures developed under earlier policy eras. This professional inertia operates not through passive resistance but through active mediation: teachers selectively enact policy elements they consider pedagogically sound, while offering superficial compliance to aspects that conflict with their assessment identities. Policy changes, therefore, encounter resistance not merely from contemporary opposition but from the accumulated weight of previous reforms that have shaped how practitioners understand the purposes and appropriate uses of assessment. Policymakers' failure to account for where teachers are situated in their professional journeys, in terms of their beliefs, values, and training developed under previous frameworks, fundamentally undermines the enactment of reform, regardless of policy rhetoric or stated intentions.

Third, the comparative analysis shows differences between Scotland and England in which actors are trusted, held accountable, and empowered as policy mediators within assessment reform processes. Scotland's horizontal accountability culture, characterised by consultation with professional networks and substantial union influence over policy enactment, contrasts markedly with England's vertical accountability structures, wherein teachers are positioned as subjects of external oversight rather than collaborative partners in reform. These differential patterns of trust and institutional power create distinct capacities and constraints for enactment. In Scotland, coordinated professional opposition, termed by policy influencer participants as 'unholy alliances' between unions, professional organisations, local authorities and parent bodies, operates through collective resistance to reforms perceived as threatening professional autonomy or educational quality. Cultural constraints include suspicion of any deviation from uniformity across schools and suspicion of policies imported from England's market-oriented system. Conversely, in England, ideologically driven tensions between political parties and professional organisations position teaching unions as adversarial rather than collaborative, with policymakers deploying combative rhetoric characterising reform opponents

as “enemies of promise”. The quasi-market environment generates demand for comparable performance metrics enabling parental choice, making standardised accountability mechanisms politically necessary regardless of pedagogical concerns. These divergent accountability cultures fundamentally shape not only which reforms can be enacted but also how teachers within contrasting professional contexts interpret and enact policy initiatives. Trust in policy mediators operates differently across jurisdictions; Scottish reforms require negotiation, whilst English reforms proceed through governmental authority despite professional opposition.

The three readings are not independent. The political calculation described in the first reading meets the temporal effects described in the second through the differences in accountability culture described in the third. Together they help to explain why accountability-intensive assessment reforms have persisted in both jurisdictions across the period examined here, and why these reforms have produced comparable outcomes for teachers despite the divergent structures of the two systems. The concluding chapter takes these readings forward into the response to the research questions and to the wider literature on policy enactment.

6 Conclusion

6.1 Response to research questions

This thesis has argued that two decades of divergent assessment reforms in England and Scotland have produced convergent outcomes for teachers who enact policy in classrooms. Accountability structures, although organised differently across countries, generated comparable pressures on teacher assessment identity and on the conditions under which teachers could use assessment to support learning.

The argument rests on cross-referenced accounts from policy influencers in both jurisdictions and from teachers and school leaders in two case study schools, one in England and one in Scotland, treated as illustrative within their respective national policy environments. The claims advanced in this chapter are therefore analytically rather than statistically generalisable (Yin, 2018).

The chapter answers the three research questions in turn, before considering what these answers contribute to existing scholarship and what they mean for the contemporary reform trajectories in both countries.

6.1.1 The purposes of assessment policy

The interviews with elite policy influencers revealed that assessment processes were designed for multiple purposes, especially accountability, teacher assessment identity, and the production of assessment data for self-promotion by teachers, individual schools, and individual governments. These findings are consistent with the literature regarding the attempted use of assessment data to create a performative culture in England (Ball, 2003), as well as to demonstrate government competence in Scotland with devolved powers for education policy (Ozga and Arnott, 2019). In both governmental cases, it is sought to demonstrate the positive outcomes from government action.

Interviews with teachers also supported the view that the national assessment policy was received positively in schools as a driver to improve their relative performance in external examinations in both countries. However, this was felt

most strongly in the Liverpool school due to the added pressure of competition for students driven through the marketisation of school choice (Perryman *et al.*, 2011). Both policy influencers and school participants described negative impacts of these policies on student wellbeing, teacher retention, and how teaching and learning operate within classrooms.

6.1.2 Alignment between policy and research

Assessment policies in Scotland and England demonstrate substantial misalignment in their research-informed creation and enactment, despite both jurisdictions' rhetorical commitments to research-informed practice. The evidence from this study reveals that policy influencers in both countries were aware of research warnings about the detrimental consequences of accountability-driven reforms to assessment processes, including constraints on teacher autonomy, increased workload, and the use of assessment for multiple competing purposes. Despite this, these policies continued to be deployed and intensified. In Scotland, the 2015 OECD review explicitly cautioned against unintended consequences from increased accountability mechanisms, yet the Scottish Government proceeded to introduce national standardised assessments. In England, political leaders advocated for competitive school models and high-stakes accountability, despite documented research evidence indicating that such approaches systematically constrain professional practice. The analysis demonstrates that research findings were selectively appropriated to legitimise predetermined policy trajectories rather than genuinely informing policy direction, with political expediency overriding pedagogical evidence when assessment served broader governmental objectives such as performance measurement and institutional comparison.

Teachers' lived experiences in both jurisdictions corroborated research concerns whilst contradicting policy rhetoric about professional autonomy. Scottish teacher participants described generating extensive assessment data that failed to inform learning, whilst intensifying workload and examination focus, precisely the negative outcomes predicted by research reviewed in Chapter 2. English teacher participants reported severely constrained autonomy, with assessment practices dictated by accountability demands that transformed teaching into test preparation, resulting in what one policy influencer acknowledged as a loss

of “the integrity of what you’re trying to teach”. Both groups of participants experienced the psychological burden and ethical distress of working within systems where accountability imperatives overrode formative, learning-focused purposes, as documented in Chapter 2. This evidence reveals that the policy-research gap persisted not through ignorance or insufficient evidence, but through deliberate political choices that prioritised accountability mechanisms over research-informed pedagogical practice.

6.1.3 Teacher, school leader and policy influencer perceptions of assessment policy

Teachers in both jurisdictions described assessment decisions as a site of contestation, in which external accountability demands pulled against their professional judgements about student learning. What differed between England and Scotland was less the presence of this tension than its source, namely the form of accountability through which it was exerted.

Scottish teachers work within what Baumfield et al. (2010) term ‘double vision’, simultaneously managing professional pedagogical judgements whilst generating extensive performance data through detailed tracking systems to satisfy hierarchical accountability requirements. Senior phase examination pressure emerges as the dominant force that “drives what happens” across secondary schools, creating what participants describe as “real fear” surrounding accurate assessment. This generates ethical distress when professional honesty about student attainment conflicts with institutional pressure for improved data outcomes. Despite materially greater professional autonomy compared to England, including stronger workplace organisation, greater curricular flexibility, and retention of professional trust within policy structures, Scottish teachers nonetheless perceive their decision-making as operating within predetermined frameworks where external reporting requirements constrain responsive pedagogy.

The Liverpool teacher participants described their autonomy as largely illusory. They retained discretion in areas considered inconsequential to accountability and lost it where assessment served system-level purposes. Teaching becomes reoriented towards test preparation, with mandated “blind” assessments on

prescribed dates precluding professional judgement about pedagogical appropriateness or curricular integration. Participants distinguish sharply between retaining limited autonomy in formative assessment whilst possessing none in summative processes that generate accountability data. Teachers link intensive assessment pressure to workforce retention challenges and describe losing empathy for pupils under these conditions.

This convergence appears explicitly in both contexts. In both jurisdictions, teachers described assessment as serving accountability rather than learning. Significantly, teachers in both jurisdictions filter contemporary reforms through assessment identities forged under previous policy regimes, generating systematic resistance rooted not in opposition to change itself, but in long-established professional beliefs about the purposes of assessment, formed over decades of conflicting policy directives.

6.2 Original contributions to knowledge

The misalignment between assessment research, policy and classroom practice has more often been read as a problem of communication than as a problem of politics. The standard account is that policy goes wrong because evidence does not reach decision-makers in a usable form, or because researchers and policymakers occupy different professional worlds. On this reading, a better translation of research findings, more accessible briefings, plainer language, closer relationships between researchers and civil servants, would in time close the gap. The standard account is attractive because it preserves the assumption that decision-makers want what the research recommends, and would act on it if only they could see it clearly.

The evidence assembled in this thesis points to a different reading. The policy influencers interviewed for this study were not under-informed. Many of them held senior positions in the very organisations that produce the research evidence under discussion. They could speak fluently about the consequences of the reforms they were enacting. Yet those reforms continued, and in several cases were intensified.

The first original contribution of this thesis follows from that observation. The gap between research and policy in assessment is better understood as the product of political calculation than as a failure of communication. The decision is made with awareness of the likely consequences, not in the absence of that awareness. The first original contribution of this thesis is to demonstrate that misalignment between research evidence and assessment policy cannot be adequately explained by deficit models of knowledge transfer. Conventional accounts of the policy-practice gap attribute such misalignment to insufficient communication between researchers and policymakers, or to inadequate translation of research evidence into accessible forms. The evidence presented in Chapters 4 and 5 indicates a more complex reality, in which many policy consequences are unintended but not unknown. Policy influencers in both England and Scotland were explicitly aware of the detrimental outcomes associated with accountability-intensive reforms, yet proceeded with enactment when those reforms served politically expedient purposes. The policy-practice gap in assessment is sustained, this thesis argues, by political calculation rather than by communication failure.

This finding is consistent with Butler's (2025, p. 5) analysis of UK governmental decision-making. They argue that 'motivated reasoning' enables political actors to dismiss contradictory evidence in favour of signals that reinforce predetermined policy preferences. Butler's work examines this process in relation to electoral consequences; this study documents it in relation to pedagogical ones. The convergence of these findings across two distinct countries suggests that damaging policies persist as a structural feature of political decision-making. This is not primarily a deficit of information, nor a failure of knowledge translation.

This has significant implications for how the policy-practice gap in educational assessment is understood. Where political purposes are well served by existing policy trajectories, improvements in evidence communication alone are unlikely to close that gap. The selective processing of evidence may operate independently of the quality or accessibility of available research.

Beyond this enactment-awareness paradox, policy change encounters resistance rooted in accumulated professional identities rather than merely reflecting

opposition to specific initiatives. Teachers' assessment identities, described through the framework developed by Looney *et al.* (2018) and explored in Chapter 2 are formed through professional socialisation under previous policy regimes, creating temporal constraints that fundamentally shape the enactment of assessment reform. In England, teachers trained during data-intensive accountability eras maintain data production practices despite revised inspection frameworks that de-emphasise performance metrics. In Scotland, teachers' collaborative professional networks, whilst offering advantages for incremental adaptation, simultaneously reflect professional cultures developed under earlier policy frameworks that constrain capacity for radical reform towards standardised assessment and data production.

These temporal effects intersect with divergent accountability structures across jurisdictions. In Scotland, horizontal accountability structures position teachers as trusted policy mediators, with unions and professional organisations exercising substantive influence over policy development and enactment. England's accountability systems, by contrast, reflects a longer neoliberal trajectory whose origins precede the study window. Established under Conservative governments from the late 1980s through the 1988 Education Reform Act and continued under New Labour after 1997 (Sammons, 2008), this trajectory was intensified by the Coalition and successive Conservative governments from 2010. Across the study period, professional and union voices were progressively dismantled, with trust in teacher judgement replaced by vertical oversight and quasi-market accountability frameworks. These fundamentally different assumptions about professional motivation create distinct capacities and constraints for policy enactment across the two systems.

6.2.1 The unintended but not unknown pattern

This thesis uses the phrase 'unintended but not unknown' deliberately, and in preference to 'unanticipated'. A consequence may be unintended, in the sense that policymakers did not design the policy in order to produce it, while at the same time being known to be likely. The evidence in this study points to the second condition holding even where the first does. Research, professional warning and prior policy evaluation were available to those making the decisions described in Chapters 4 and 5. Where those decisions produced consequences

that the literature had predicted, the consequences cannot reasonably be described as unanticipated. The language of unanticipated has often been deployed to manage the political reception of consequences that were, on the available evidence, foreseen. This difference is important because the deficit-model account of the policy-practice gap relies on the stronger claim, that policymakers genuinely did not see the consequences coming. The evidence assembled here does not support that stronger claim. What it supports is the weaker and more troubling claim that the consequences were known and that they were tolerated as a price worth paying for the policy's political utility.

The data from policy influencers suggested that many policy outcomes with negative consequences may be unintended but not unanticipated. Where politicians anticipate harm, the public might reasonably expect a policy to be revised or redirected so that the identified risks are addressed. The evidence from this study points to a different pattern. Policies whose consequences had been flagged by research and by professional bodies were nevertheless enacted, because they served the political argument of the day rather than the pedagogical purpose claimed for them. The evidence from both countries suggests that policy documents and ministerial speeches may strategically omit anticipated negative consequences. Opposition to reform may also be reframed as the partisan product of political adversaries rather than as a legitimate pedagogical concern (Ball, 1993). This pattern is taken up again in the final reflections, where it is connected to Parkhurst's account of issue bias in evidence governance (see Conclusion p. 273).

In England, policy influencers demonstrated explicit awareness that accountability linked to assessment processes could distort these processes and undermine the integrity of the information produced. As a result, the data may not accurately reflect students' learning progress. Union leaders linked the increases in pressure from assessment-related accountability to negative workload factors that contributed to teacher retention difficulties in England.

In Scotland, the OECD issued recommendations in 2015 for increased accountability mechanisms, accompanied by warnings about unintended consequences of those processes (Hayward, 2018). Participants in this study identified several of those consequences as now present, including increased

teacher workload and an intensified focus on examination performance at the expense of curriculum coverage and pedagogy responsive to learning need.

The Scottish response to the OECD review and to the 2015 PISA results, which McIlroy (2018, p. 2) describes as attracting 'alarmed media headlines and vociferous criticism', took an instructive form. Rather than maintain the sample-based instrument that had provided system-level evidence, the Scottish Government discontinued the Scottish Survey of Literacy and Numeracy in 2016 and introduced the Scottish National Standardised Assessments under the National Improvement Framework (Kreijkes and Johnson, 2023). McIlroy (2018, p. 11) argues that the new arrangements raised the stakes of assessment without adding value for learners, by mixing 'more and less reliable and more and less valid data into a single data set'.

What is striking about the Scottish episode is that the consequences McIlroy describes were not unforeseen. Research warning of comparable effects from high-stakes data production had been in circulation well before the National Improvement Framework was adopted (Hayward, 2018; Childs and Baird, 2020). The decision therefore sits inside a wider pattern, observed across both jurisdictions in this study, in which consequences known to be likely at the point of decision were nevertheless tolerated as the price of a politically advantageous course. Locating that pattern precisely calls for a conceptual scaffold.

The relationship between policy intent and policy consequence can usefully be mapped along two axes. The first axis distinguishes consequences that the policy designer intended from those they did not. The second axis distinguishes consequences that were anticipated at the point of decision from those that were not. The resulting four-cell typology, set out in Figure 6, allows the consequences identified in Chapters 4 and 5 to be placed and compared.

Reading the cells in turn allows the contribution of this thesis to be located precisely. The lower-left cell, unintended but not unknown, is the cell most heavily populated by the evidence assembled in Chapters 4 and 5. Policy influencers in both jurisdictions could speak fluently about the consequences that fell within this cell. Research warnings, professional testimony and prior

policy evaluations had been available to them at the point of decision. The deficit-model account of the policy-practice gap relies on the lower-right cell, where decisions are taken without the information needed to anticipate harm. The evidence presented here does not support that account. It supports the weaker but more troubling claim that the harms were known and tolerated as a price worth paying for the policy's political utility.

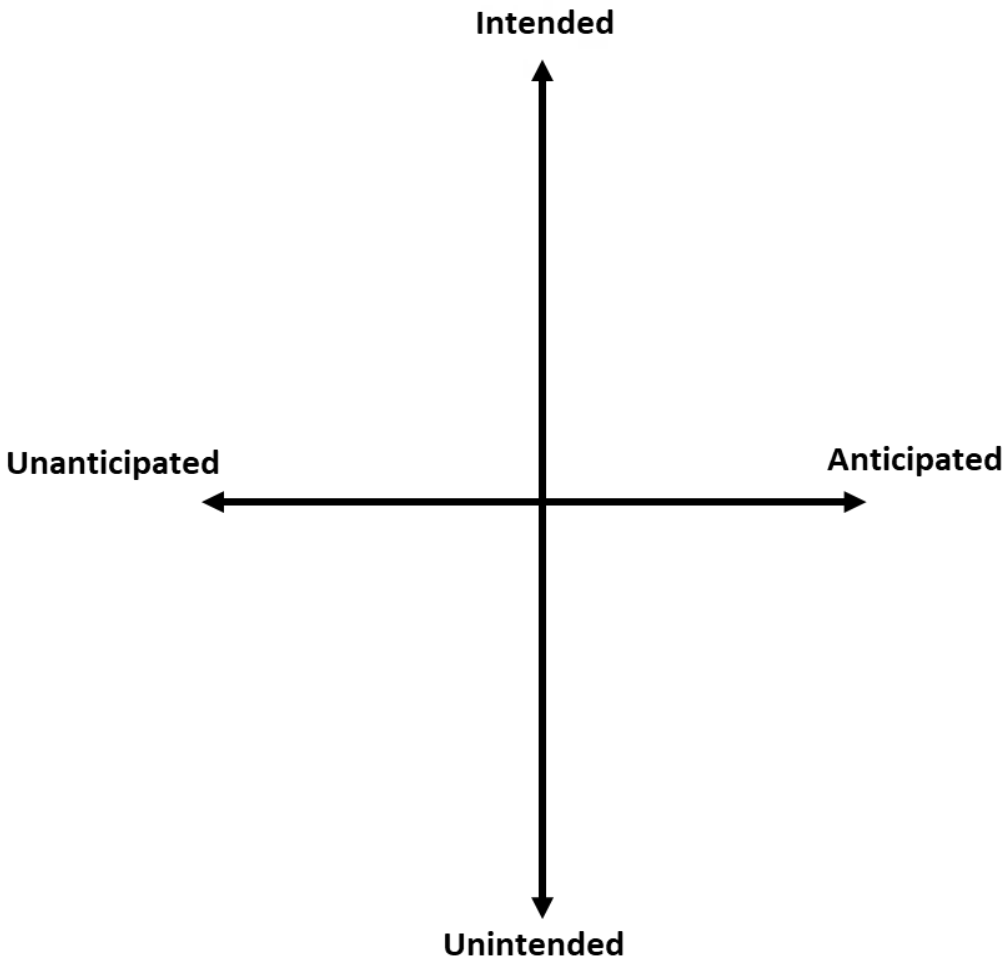


Figure 6: Framework exploring the intersection of policy intentions and anticipated consequences.

When politicians address anticipated negative consequences, they deliberately make observable attempts to undermine and minimise them. This might take the form of delegitimising counterarguments to their political rhetoric or of being mindful of which consequences the electorate might tolerate during a parliamentary period.

Butler (2025, p. 4) identifies a pattern he terms ‘rational unresponsiveness’, in which political actors proceed with policies they recognise may generate negative consequences, having calculated that other priorities, party ideology, longer-term strategic advantage, or policy commitments, outweigh the anticipated costs. This concept offers a productive theoretical framing for the pattern identified in this study. The evidence from national policy influencers in both England and Scotland suggests a comparable process. Warnings from practitioners, researchers, and international review bodies were acknowledged by those with responsibility for policy yet were subsequently reframed as enactment failures or partisan resistance, rather than as evidence warranting revision of the policy.

Butler's (2025) account does not claim that decision-makers ignore evidence. It argues that they process it selectively, weighing signals consistent with their preferred policy direction over those that contradict it. The mechanism is one of motivated reasoning rather than the absence of information. Butler demonstrates that decision-makers with access to information that alerts them to likely negative consequences still choose to dismiss it in favour of evidence consistent with their preferred trajectories. This mechanism offers a plausible account of the cognitive process underlying the unintended but not unknown pattern documented in this thesis.

In England, what the media, particularly newspapers such as the Daily Mail, accepted appeared to exert significant pressure on policy development. Politicians were characterised as being fearful of public and media reaction to the wrong type of reporting on their policy decisions and the measured outcomes. Modern political rhetoric, as seen in speeches by ministers such as Michael Gove, draws attention to their political base through the characterisation of political opponents as ‘enemies of promise’ and ‘enemies of reform’. These combative statements are used to undermine the warnings of negative policy consequences issued by opponents, such as unions and education researchers. This particularly helped keep political rhetoric aligned with think tanks such as Policy Exchange, which provided empirical data to support the case for competitive school models. Through this process, the negative impacts are perhaps considered unanticipated and unintended in the media and public

perception, as they are rarely fully addressed during the policy design process and are often portrayed by politicians of the day as emerging from groups characterised as illegitimate.

During the New Labour government, there was awareness of the pressures from accountability processes and their negative effects, but there were demands for further intensification. Quotes from policy influencers in the previous chapter (see page 245) discussing the demands of the then Prime Minister, Tony Blair, highlighted the political desperation and the difficulty of demonstrating improvement in assessment data despite negative consequences from such processes.

In Scotland, unions, professional organisations, local authorities and parent bodies hold substantial influence over policy development. Politicians remain aware of the reputational cost of unsuccessful reform, and policies that promise political advantage proceed even where the research literature warns of harm. Scotland's consultation culture embeds professional engagement in policy formulation, yet it also limits the scope for rapid reform. Policy influencers described this dynamic through the image of an "unholy alliance" of stakeholders mobilising against changes that threatened established practice.

This resistance operates through culturally embedded expectations, particularly the rhetoric of preventing a ‘postcode lottery’, which demands uniformity across schools and views any differentiation with intense suspicion. Politicians work through these constraints by selectively deploying international assessment data, particularly international PISA results, to legitimise policy interventions and at the same time asserting Scotland's educational distinctiveness. A point of similarity between England and Scotland is that both utilise PISA to outline competitiveness. For Scotland, the strategic appropriation of PISA data serves dual purposes: demonstrating educational effectiveness to domestic audiences whilst positioning Scotland as a capable, self-governing nation within international frameworks. This was exemplified when Cabinet Secretary Mike Russell appropriated the same 2009 PISA findings that Michael Gove had cited as evidence of decline, instead arguing they demonstrated substantial improvements in Scottish education (Hayward, Macbride and Spencer, 2016). Through this process, politicians publicly promote reforms aligned with

international benchmarks while privately acknowledging the barriers to enactment created by multi-stakeholder opposition.

The consultation culture itself becomes a mechanism through which negative policy consequences are deflected or reframed. When faced with warnings from practitioners and researchers about detrimental impacts, politicians can point to extensive consultation processes as evidence of democratic legitimacy, even whilst policy influencers privately acknowledge that predetermined trajectories remain unchanged. National governments, constrained by local authorities' statutory responsibility for educational provision, must influence professional networks to facilitate reform, creating what Nicola Sturgeon characterised as difficulties in enacting national policy that 'had to be left in the hands of local authorities, who didn't always want to play ball' (Sturgeon, 2025, p. 346). This multi-layered governance structure allows politicians to pursue policies with known negative consequences, such as the intensification of data collection following OECD recommendations, by framing opposition as resistance to improvement and parochial protection of vested interests, rather than legitimate pedagogical concerns supported by research evidence. The predicted negative consequences, including increased teacher workload and intensified focus on examination performance to the detriment of curriculum coverage, are subsequently positioned as enactment failures rather than inherent policy flaws, thereby insulating political decision-makers from accountability for outcomes they were explicitly warned about during policy development.

This analysis demonstrates that the negative consequences of accountability-intensive assessment policies emerge not from politicians' ignorance but from calculated political decision-making. The evidence from both England and Scotland reveals a consistent pattern in which national policy influencers proceed with reforms despite documented warnings, implicitly calculating that negative outcomes, increased teacher workload, intensified examination focus, and compromised data integrity are acceptable costs in pursuit of politically advantageous reforms.

The constraints of short parliamentary terms, typically four to five years, create conditions whereby politicians must demonstrate measurable improvement to their electorate within limited timeframes. This temporal pressure prioritises

the production of politically useful data over long-term pedagogical coherence, with negative consequences tolerated as collateral damage provided they remain invisible to the media and voting public or can be rhetorically deflected onto critics characterised as unreliable and partisan opponents. This structural dynamic was identified by Stronach and Morris (1994, p. 5) who described ‘policy hysteria’ as producing conditions in which reforms were ‘never in an uncontested place long enough’ to succeed. The political calculus documented in this thesis is therefore not a recent development but a persistent feature of accountability-intensive policy formation.

The convergence of outcomes across these contrasting governance systems suggests that the underlying political calculus, in which short-term political gains outweigh suspected pedagogical harm, operates regardless of the specific enactment mechanisms. However, whilst politicians operate within brief parliamentary cycles that incentivise rapid, measurable change, the teaching profession operates according to fundamentally different temporal dynamics. The policies enacted during these short political windows create lasting effects that accumulate and interact with teachers’ professional formation in ways that policy-makers fail to anticipate. The following section examines how teachers from different policy eras respond to reforms based on their professional socialisation under previous frameworks, revealing temporal effects and professional inertia that further complicate the enactment of accountability-intensive assessment policies.

6.2.2 Temporal policy effects and professional identity

The first contribution means that policy intentions, however well grounded, do not arrive in classrooms unmediated. They arrive into a professional context that has its own history. Teachers in both jurisdictions described their current assessment practice through reference to the policy environments that had shaped their training and early careers. Read this way, the resistance that policymakers often identify in reform enactment is less a position taken in opposition to a current reform than a working-through of an earlier one. The second contribution of this thesis therefore concerns the temporal dimension of assessment policy enactment. Looney et al.'s (2018) account of teacher assessment identity provides a useful vocabulary for describing this dimension,

although it is used here as a sensitising concept rather than as an operational framework. What the empirical material adds is evidence that assessment identities are inherited from prior policy regimes, and that this inheritance constrains the enactment of subsequent reforms in ways that policy designers have not consistently anticipated.

Teachers' professional identities as assessors have been explored in this thesis, drawing on a model built on previous research on assessment literacy and then theorised within a framework of teacher assessment identity proposed by Looney *et al.* (2018). Teachers in England and Scotland have formed assessment identities that appear resistant to policy change, retaining their understanding and perceptions of their roles as assessors from previous policy environments. Teachers demonstrated resistance to policy changes in both Scotland and England.

In England, recent policy changes have aimed to create a shift away from the expectation that schools generate internal assessment data, focusing instead on curriculum intent, enactment, and impact within the Ofsted inspection framework. Despite these intended changes in practice, teachers in the case study school described much of their assessment practice as fulfilling a need to generate assessment data to feed accountability mechanisms. Teacher participants in the English policy environment described being trained and developed in a previous policy environment in which test data were used defensively to avoid accountability frameworks that might otherwise judge them as underperforming. The production of such data through tests and examinations, referred to as data drops, became an essential part of teachers' identities and roles, ensuring that external metrics were met and that potential negative consequences, such as an unfavourable inspection rating, were avoided. Despite attempts to reform the focus of school inspections away from test data, participants described a continued focus on such metrics to serve models of internal and external accountability.

In Scotland, policy shifts have aimed to increase the use and production of test data and encourage teachers to utilise this data for tracking and monitoring student cohorts. The teachers interviewed in the case study school, however, described the tests as having limited educational value, focusing much of their

discussion instead on formative purposes of assessment rather than tracking student progress through test data. Teachers were described by policy influencers as being formed during different policy eras (Pre-Curriculum for Excellence experienced teachers compared to younger Curriculum for Excellence-era teachers), bringing different identities in their approaches to assessment. Teachers demonstrated resistance to standardised national assessment and a preference for classroom-based formative assessments, providing a variety of in-the-moment information for teachers to respond to. Assessment is for Learning was described as a successful and popular policy by both policy influencers and teachers, with a lasting legacy evident in discussions with classroom teachers and policy influencers, highlighting the involvement of students as co-creators in assessment.

In this section, the evidence for differential trust in political rhetoric and accountability frameworks across England and Scotland is reviewed, along with the groups of professionals who are trusted. In England, market mechanisms, external oversight, and performance data are upheld as reliable and trustworthy, whereas teacher professional judgement is discussed with distrust. In Scotland, greater trust is placed in professional networks, consultation processes and collective decision-making, though both ultimately maintain accountability pressures.

Political rhetoric in the English context, outlined in the previous chapter, deployed combative characterisations that positioned reform critics, particularly teaching unions and education researchers, as obstacles to the government's reform agenda. This framing appears to cast such groups as obstructive rather than constructive, implying they should defer to the changes advanced by politicians such as the former Education Secretary Michael Gove. Teachers in England were described by participants as under forms of surveillance to ensure policy compliance rather than as trusted participants in policy enactment. In the English case study school, teachers were asked to conduct assessments “blind”, meaning they could not see the content in advance, a manifestation of distrust in teachers’ professional judgement.

Teaching unions were characterised as adversarial rather than collaborative partners with a contribution to make towards policy design and enactment. The

government, especially but not exclusively Conservative ministers, characterised the unions as left-leaning and therefore in opposition. In England, trust is placed in a quasi-market environment where school competition is facilitated through inspection and exam performance data, rather than relying on teachers, to improve educational standards.

By contrast, the Scottish system was characterised by policy influencers as fostering professional networks that enabled consultation with teachers and teaching unions as a routine aspect of policy development. Whilst teacher participants from Dundee did not always perceive their individual perspectives as being heard, policy influencers identified professional networks and union partnerships as mechanisms through which the profession's collective views were frequently represented during both the design and enactment phases of national policy. Unlike in the English context, Local Authorities in Scotland were reported to exercise delegated authority over schools, working in partnership with unions and professional organisations to mediate the implementation of national education policies. This arrangement was interpreted divergently by policy influencers: some characterised it as an obstructive 'unholy alliance' (p. 215) that constrained improvements in educational standards, whilst others credited this collaborative structure with enabling resistance to test-based accountability measures and supporting the preservation of formative assessment practices in Scottish classrooms, practices which had been described as subdued in England as a result of accountability pressures.

Le Grand's (2006) theorisation of fundamental ideological positions about human motivation provides a basis in the literature for these findings. His conceptualisation of Knights, Knaves, Pawns, and Queens helps describe how policymakers might identify people and design systems based on assumptions about their behaviour. These ideas and their application to the conclusions in this work are captured in Table 6.4.

Table 6.4: The categories and descriptions of public actors, including empirical evidence. Categories and behaviours adapted from Le Grand (2006).

Quadrant	Described behaviours	Empirical evidence from this work.
Knights	Altruistic, public-spirited motivated to help others.	Scottish evidence reveals contrasting assumptions that position teachers simultaneously as "knights" and "queens" within Le Grand's (2006) typology. Horizontal accountability structures and a consultation culture indicate assumptions consistent with the "knight" conception, positioning teachers as altruistic professionals deserving trust, manifesting through delegated authority and policy development processes that incorporate professional expertise. Yet teachers are simultaneously positioned as "queens", powerful collective actors capable of coordinating resistance across unions, local authorities, and professional bodies, thereby enabling or frustrating reform initiatives depending on professional consensus. This dual positioning creates enactment environments wherein national governments must negotiate with rather than direct professional networks, generating particular constraints when professional consensus opposes policy directions. This positioning reflects assumptions consistent with social democratic traditions that value professional expertise whilst acknowledging the legitimate power of organised bodies, though it produces markedly different enactment dynamics than England's model. Where England's approach assumes distrust necessitates constraint, Scotland's assumes trust enables professional empowerment, revealing how fundamental assumptions about motivation and agency shape accountability structures.
Queens	Active agents Autonomous capable of independent action	
Knaves	Self-interested motivated by personal gain.	The empirical evidence from England reveals that policymakers position teachers simultaneously as "knaves" and "pawns" within Le Grand's (2006) framework. Policy influencers described political rhetoric that characterised professional resisters as adversaries, indicating fundamental distrust consistent with the "knave" assumption, positioning teachers as self-interested actors requiring external constraints. This distrust takes practical form in quasi-market accountability frameworks designed to harness self-interest through competition. However, enactment structures simultaneously position teachers as "pawns": passive recipients of policy with limited autonomous agency. Teacher participants report highly directed assessment processes with predetermined timing and content that remain inaccessible to professional influence, with external assessment increasingly imposed upon teachers rather than allowing them substantive roles in the awarding of qualifications. Professional discretion is largely retained within formative assessment practices, whilst external summative processes position teachers as enactors rather than co-constructors of assessment judgements. This dual positioning, assuming both untrustworthiness and limited capacity for autonomous action, reflects assumptions consistent with neoliberal ideology, in which distrust necessitates constraint through prescribed content and external oversight, whilst market mechanisms align professional self-interest with policy objectives.
Pawns	Passive victims of circumstance with limited autonomous action.	

The theoretical categorisation of how teachers may be framed by policymakers must be understood as describing predominant policy assumptions rather than

totalising characterisations that apply uniformly across all contexts and temporal moments within Scotland and England. It is essential to acknowledge that these overarching theoretical distinctions may obscure some convergence in classroom practice. The day-to-day experiences of teachers in both Scotland and England reveal more subtle patterns of constraint and agency than the binary knight/knave characterisation might suggest. English teachers operate within vertical accountability structures premised on distrust. Despite this, they retain discretion in formative assessment and attempt to influence policy through subject associations and union networks. Aspects of knight-like motivation therefore persist in a system designed around the knave assumption.

Conversely, Scottish teachers, while benefiting from horizontal accountability cultures that emphasise professional trust, frequently reported feeling excluded from national policy decisions and systematically constrained by secondary examinations. The evidence thus suggests that whilst policymaker assumptions fundamentally differ between Scotland and England, the material consequences for teacher autonomy may not always be so disparate, particularly regarding high-stakes assessment processes that constrain pedagogical responsiveness in both contexts, despite the legacy of their divergent accountability philosophies.

On the reading offered here, effective policy enactment depends less on the alignment between research evidence and policy intent than on two further conditions. The first is the mismatch in time-horizon between political cycles and professional formation. The second is the differing assumptions about professional motivation built into each accountability system. The convergence of negative outcomes across England and Scotland, despite their contrasting approaches to professional trust and governance structures, suggests that accountability-intensive assessment policies impose comparable constraints on pedagogical responsiveness regardless of the mechanisms of their enactment. However, the differential positioning of teachers as policy mediators, trusted partners in Scotland, versus surveyed enactors in England, creates distinct capacities for incremental reform and professional resistance. Policy designers who fail to account for teachers' accumulated assessment identities and the temporal dynamics of professional socialisation will continue to encounter enactment barriers independent of policy content or evidence base, whilst

researchers examining the policy-practice gap must move beyond deficit models that privilege communication failures over political calculations that knowingly tolerate detrimental consequences.

6.3 Implications and recommendations

6.3.1 Professional formation and reform enactment

The temporal dimension considered in the Conclusion has implications for how policy is designed, not only for how it is interpreted. Ball, Maguire and Braun's (2012) account of teachers as policy mediators describes what teachers do when policy reaches them. The evidence in this thesis adds a further point about what policy designers should expect when their policy arrives. Teachers do not respond to a new assessment policy in the abstract. They respond to it through the assessment identity they have already developed, which is the working-through of one or more earlier policies. The implication for design is that the staging of any new policy needs to take seriously the previous policy regime each cohort of teachers was formed.

A policy that addresses teachers as if they had no professional history will be enacted, in practice, by teachers who do have one. The result, observed in both jurisdictions in this study, is that elements of the previous policy continue to be enacted within the new one. Where this is anticipated and accommodated in the policy design, the inheritance is a resource. Where it is not, the inheritance is recoded as resistance. The recommendation that follows from this part of the analysis is therefore that policy designers and enactment specialists should treat teachers' assessment identities as an empirical starting point for design, not as a set of attitudes to be changed in the course of enactment.

Policy designers should recognise that reform enactment strategies require explicit attention to where practitioners are situated in their professional journeys and what assessment paradigms shaped their professional formation. Effective policy change cannot assume that practitioners begin from neutral positions. Enactment timelines and support structures need to account for the work of reshaping professional identities formed under previous policy regimes. This suggests that successful reform works with, rather than against, established

professional frameworks, providing structured transition pathways that acknowledge and build upon existing expertise whilst enabling paradigm shifts. For researchers examining policy enactment, this temporal dimension indicates that enactment studies must investigate not only current policy contexts but also the historical policy trajectories that shaped practitioners' professional identities. Understanding policy resistance requires attending to the accumulated effects of prior reform cycles on professional cultures and institutional practices.

6.3.2 Political expediency and pedagogical evidence

Deficit models of the policy-practice gap locate enactment failures in insufficient evidence or in poor communication between researchers and practitioners. The evidence assembled here points to a different explanation. Policy influencers in both jurisdictions articulated explicit knowledge of the detrimental consequences of accountability-intensive reforms, yet they continued to deploy such mechanisms when these served politically expedient purposes. Triangulating policy influencer testimony with the lived experiences of teacher participants strengthens the interpretation that negative outcomes are not unforeseen but politically tolerated. Scottish teacher participants described a 'real fear' surrounding assessment accountability, and English teacher participants reported a loss of 'integrity' in their classroom practice. Policy influencers acknowledged that teacher morale was eroding and that workforce retention was under strain. Successive English governments nevertheless maintained high-stakes performance metrics to satisfy perceived market demands and to manage media expectations. The policy-practice gap in this domain is sustained, in other words, less by what is not known than by what is politically inconvenient to act upon.

These findings challenge conventional assumptions about evidence-informed policymaking. The evidence suggests that the selective appropriation of research findings to legitimise predetermined policy trajectories generates professional cynicism and systematic resistance among teachers, ultimately undermining the effectiveness of reform despite rhetorical commitments to evidence-based practice. This pattern suggests that research on policy enactment should treat political expediency as a primary driver of reform decisions. It should move

beyond frameworks that attribute policy-practice gaps to knowledge deficits or to communication failures alone.

6.3.3 Divergent accountability and policy mediation

The comparative analysis in this thesis reveals fundamentally different assumptions about professional motivation embedded within England's and Scotland's assessment policy frameworks. England's system, characterised by distrust of professional judgement, necessitates vertical oversight structures, whilst Scotland's approach, grounded in trust of teacher professionalism, enables horizontal collaborative mechanisms. These differing accountability structures create distinct capacities and constraints for policy enactment, providing empirical evidence for Le Grand's (2006) conceptual distinction between “knights” and “knaves” motivational assumptions when applied to educational assessment policy.

In England, vertical accountability structures position teachers as “knaves”, actors requiring external oversight through market-based competition and standardised performance metrics. This positioning, enacted by successive governments, manifests through combative political rhetoric and accountability mechanisms designed to serve quasi-market parental choice rather than formative pedagogical purposes. Conversely, Scotland's horizontal accountability framework assumes teachers are “knights” deserving professional autonomy, operationalised through extensive consultation processes and formal union partnership in policy development.

The evidence from this research suggests that policymakers in both jurisdictions should recognise that neither accountability culture operates as a panacea for eliminating variation in standards. Vertical accountability structures, whilst potentially enabling consistent policy enactment across diverse contexts, simultaneously constrain professional expertise and adaptive responses to local needs. Horizontal accountability structures protect professional agency and enable contextualised policy adaptation for specific school and student circumstances, yet diffuse responsibility and create enactment friction when radical systemic reform is pursued. These findings indicate that effective policy design and enactment require explicit attention to how trust assumptions

fundamentally shape which policy influencers possess the capacity to enable or frustrate reform initiatives.

6.3.4 Conceptualising policy divergence

There are divergent outcomes and resistance to policy reform in both England and Scotland, deriving from their distinctive administrative structures and the legacy of pre-devolution policy networks (Arnott, 2005). Adding to this Arnott and Menter (2007) demonstrate how the Scottish education community successfully resisted Thatcherite reforms in the late 1980s. The authors illustrate how Scotland is enacting diluted alternatives, Curriculum Guidance 5-14 rather than the National Curriculum, and teacher assessment instead of published test results, whilst quietly subverting imposed policies like Devolved Management of Schools. These events have created two distinct policy environments where similar reform rhetoric encounters markedly different institutional landscapes and professional cultures, which this work considers through the lens of educational assessment policy.

Professional organisations have been shown to exert considerably greater influence in Scotland than England. For example, Kennedy (2018, p. 827) highlights the Educational Institute of Scotland (Established 1847) and General Teaching Council for Scotland as providing teachers with ‘comparatively stable professional existence’ and ensuring the teacher profession is heard. Arnott (2005) demonstrates that whilst England introduced market-based accountability through parent and governor participation to circumvent professional autonomy, Scotland maintained professional and local authority influence in policymaking, emphasising professional rather than market-based accountability modes. However, Bryce and Humes (2018) and Humes (2023) identify potential limitations of Scotland’s close-knit policy community, suggesting it may promote “groupthink” and conformity rather than critical interrogation of orthodoxies. This thesis explores how different actors in the process of policy creation and enactment shape how policy is enacted or resisted in each country.

Political culture and trust in professionals may diverge markedly between jurisdictions. In Scotland, education constitutes one of three pillars of cultural identity alongside law and the Kirk (Arnott and Menter, 2007), with the ‘Scottish

myth' positioning education as inherently democratic. Anderson (1983, p. 1) characterises this as a belief that 'Scottish education was peculiarly "democratic"' and argues it 'influences history by … ruling out some developments as inconsistent with the national tradition," creating conservatism within educational establishments and rendering incompatible policies vulnerable to being labelled 'anglicising' (Arnott and Menter, 2007). Conversely, England experienced Callaghan's 1976 Ruskin speech, which initiated systematic questioning of educational standards and established a foundational distrust that legitimised market-based accountability reforms prioritising parental and governor participation over professional autonomy (Arnott and Menter, 2007).

This adversarial dynamic intensified under Michael Gove's reforms, with education professionals characterised as "the blob", imported from Ronald Reagan's Education Secretary, and positioned as "enemies of promise" who were "politically motivated" and prioritised 'valuing Marxism' over 'fighting ignorance' (Bennett, 2019). Gove explicitly framed reform as requiring combative positioning, stating 'it is a battle in which you have to take sides', with the Department for Education reportedly placed on a "war footing" for the introduction of a reformed National Curriculum (Bennett, 2019). Craske (2021) reads this rhetorical strategy through the framework of populist logic, in which a reforming project constructs its legitimacy by naming an establishment adversary. On his account, the recurrent invocation of 'the blob' was not incidental rhetoric but a structural feature of the conservative reform project. Within this framing, the rhetorical move not only critiques specific policies but enables the displacement of professional knowledge from policy formation more broadly. Stronach and Morris (1994, p. 5) characterised these conditions as 'policy hysteria', identifying a structural pattern of shortening reform cycles, overlapping innovations and the scapegoating of professional groups. They argued that this pattern generated an 'endemic crisis of legitimation' (p. 4), requiring each reform to be replaced before its effects could be known. This work provides empirical evidence of how these political environments operate in the modern creation of education policies and reveals the extent to which they have succeeded in influencing classroom assessment practices.

6.4 Research limitations and delimitations

The scope of this study reflects both deliberate design choices and constraints arising from the research context. The research focuses on two secondary schools, one in Liverpool and one in Dundee, selected for their comparable socio-economic profiles and contrasting national policy environments. This selection supported the in-depth, contextually grounded case study analysis required by the research questions, but it does not represent the full diversity of school contexts within either jurisdiction. The findings should therefore be regarded as analytically generalisable to the theoretical frameworks developed here, rather than statistically generalisable to the broader population of schools in England or Scotland (Yin, 2018).

The reliance on semi-structured interview data captures policy influencers', teachers' and school leaders' perceptions and professional understandings of policy enactment, rather than offering observational evidence of classroom practice. Classroom observation formed part of the original research design, but consent for school visits was not granted owing to COVID-19 restrictions at the point of data collection (see Methodology, p.126). Online interviews replaced this element, enabling sustained inquiry into how teachers and school leaders understood and articulated their assessment practice. The exclusion of Wales and Northern Ireland was a deliberate scoping choice, allowing close comparison between two jurisdictions that share a legislative history yet have followed divergent post-devolution trajectories. Future comparative research extending the analysis to these jurisdictions could usefully test the analytical frameworks developed here.

6.5 Final reflections

Starting with a question about the alignment between policy, research and practice, this thesis concludes with an uncomfortable finding about the true nature of that misalignment. The evidence presented in this work suggests that the policy-practice gap in the purposing of educational assessment processes is not primarily a problem of knowledge translation or communication failure, but one of political purpose. National policy influencers in both Scotland and England demonstrated awareness of research evidence warning of the impacts of

reforms they enacted; teachers in both contexts experienced precisely the negative consequences that the literature had long predicted. The gap persists not because policymakers lack access to the evidence, but because they have chosen, repeatedly and with documented awareness, to prioritise accountability mechanisms serving governmental interests over pedagogical approaches serving students and teachers. Parkhurst's (2017a) account of evidence governance, introduced in Chapter 2 (p. 35), offers a useful lens for characterising what is happening here. The dominant bias evident in both jurisdictions is what Parkhurst terms issue bias, in which a political preference for accountability evidence displaces evidence about pedagogical consequence.

This “unintended but not unknown” pattern represents a fundamental challenge to optimistic models of evidence-informed policymaking.

Both England and Scotland have, since the time of this thesis’s primary data collection, indicated significant assessment reform programmes. In Scotland, the Hayward Review (Scottish Government, 2023) recommended a Scottish Diploma of Achievement to broaden recognition of learner accomplishments, whilst the Education (Scotland) Act 2025 (Scottish Government, 2025) established Qualifications Scotland to replace the SQA, with curriculum reform scheduled through to 2032. In England, the Curriculum and Assessment Review has led to the removal of the EBacc performance measure, proposed reforms to the Progress 8 accountability mechanisms, and commitments to a broader, more skills-oriented curriculum. Both reform programmes employ rhetoric that echoes longstanding research recommendations, reducing high-stakes examination pressure, recognising diverse achievements, and restoring professional agency. Yet the findings of this thesis counsel caution about whether such rhetorical commitments will translate into meaningful, diverse achievements and the restoration of professional agency. Recent international analysis reinforces this concern; the OECD (2026, p. 175) observed that when countries reform their upper secondary certificates, ‘matters of credibility, relevance and fairness tend to dominate policy debates and stakeholder consultation’ whilst manageability, defined as the workload implications for students, teachers, and assessment systems, emerges as ‘the “forgotten principle” and yet as the factor that makes or breaks certification reform’. The Scottish teachers in this study already

operated within a system that rhetorically valued professional trust and formative assessment, yet nonetheless experienced “real fear” surrounding accountability and data generation that constrained their practice. English teachers, despite working within explicitly prescriptive frameworks, articulated how autonomy was confined to spaces deemed inconsequential to accountability purposes. If the negative consequences documented in this research emerged from policy frameworks that already claimed alignment with research evidence and professional voice, there is little to secure real change in these policies unless the processes of assessment accountability and the conditions of teacher workload receive the sustained attention that the OECD analysis suggests they have historically been denied. Without explicit consideration of how reforms will be enacted within the context of existing professional identities and institutional constraints, the pattern identified in this thesis, in which awareness of negative consequences is insufficient to prevent their recurrence, is likely to persist across successive reform cycles.

The discontent with policy environments across this study should not be read as a grievance about workload alone. Teachers objected to accountability-driven assessment because it obstructed something they valued, namely an honest account of what their students had learned. A teacher in the Dundee school described an institutional reluctance to examine attainment closely, then rejected that reluctance on explicitly ethical grounds, because concealing what students did not know helped nobody and least of all the students themselves. A school leader in Liverpool identified the same objection in the empathy that examination pressure was eroding, treating that erosion as damage to the relationship they had with their students. These motivations are driven by the moral and ethical concerns teachers have about the value of assessment and its purpose to serve their students’ learning.

The comparative analysis suggests that the instrumentalisation of assessment carries costs for both teachers and students. When assessment processes are repurposed for objectives beyond supporting learning, whether for market competition or governmental legitimation, these consequences emerge regardless of the trust assumptions or reform rhetoric embedded in policy frameworks. Teachers in both jurisdictions demonstrated persistent commitment

to their students' learning, often finding ways to exercise professional judgement within or against constraining structures; yet this resilience should not be mistaken for evidence that current systems are functioning adequately. The Scottish teachers who maintained formative practices despite accountability pressures, and the English teachers who found spaces for pedagogical integrity within prescriptive systems, remind us that policy enactment is never simply policy imposition. However, understanding how teachers navigate, adapt and resist policy pressures, and recognising the considerable personal and professional costs they bear in doing so, may ultimately prove more consequential for improving educational assessment processes than any further refinement of research evidence that policymakers have already demonstrated they are willing to disregard when it conflicts with political imperatives. Until policymaking authentically centres on the pedagogical relationships between teachers and learners, rather than the accountability relationships between governments and electorates, the misalignment between research, policy and practice documented in this thesis is likely to persist.

Appendices

Appendix A: Interview schedule for Policy Influencer

Interview schedule for policy influencers

- Thank you for:
  - your time. This interview is expected to last for one hour.
  - willingness to participate in this recorded interview today.
- As you may be aware from the participant information sheet, I am interested in the relationship between national assessment policy and classroom assessment practice. I would be interested to hear about your experiences with educational assessment from about the year 1998.
- To start our discussion, I have a list of questions, but please free to add anything else that comes to mind.

Q1.1 What does the term educational assessment mean to you?

a) Thinking of your own experiences of educational assessment, do you remember any point where assessment had a personal impact on you?
b) What purposes do you believe educational assessment is currently serving in England. then Scotland?
c) In what ways do you think educational assessment impacts on what happens in the classroom? Why this impact?

Probe:

- Might there be any positive or negative effects of this?
- Can you give me a specific example?

Q2.1 I am interested in the extent to which assessment is used to hold people to account. What do you understand by the term accountability? If accountability were 'intelligent' as advocated by Onora O'Neill – in your view, what would it be like?

a) From your perspective, how has accountability shaped assessment since 1998- if appropriate, please feel free to comment on both England and Scotland?
b) To what extent do you believe that accountability has had an effect on school-based assessment and teaching? In England? In Scotland?

Probe:

- Can you offer any examples of a country where accountability is used intelligently?
- Can you give me a specific example of why you think they have managed this?

Q3.1 What do you see as the most significant consequences of using and reporting educational assessment data?

a) Who do you think is most significantly affected as a result of reporting educational assessment?

b) In your experience, have there been any changes to the way in which the results of assessments might be reported to groups of people e.g. parents, local government or national government?

Probe:

- Can you tell me about any positive and negative implications there may be from this?
- Can you give me a specific example?
- Might different teachers in some contexts with students from some backgrounds have different views on this?

Q4.1 How might the assessment process be used for political purposes, including international assessment, national assessments and classroom assessment.

a) To what extent do you believe international assessments such as PISA are used for political purposes?

b) Can you describe your involvement with changes or redesign of assessment processes? Were there political motivations for the changes proposed? What were they?

Probe:

- Can you tell me about any positive and negative implications there may be from this?
- Can you give me a specific example?

Q5.1 To what extent have changes to assessment policy that you have had involvement with been successful at changing classroom practice?

a) What are the most challenging aspects of enacting an assessment policy? Would this be the same nationally as well as locally?

b) What are the biggest enactment challenges that remain for the future of educational assessment?

Perhaps ask for an example?

Probe:

- Can you tell me about any positive and negative implications there may be from this?
- Can you give me a specific example?

6.1 In five years' time, if we could say that classroom assessment has improved, what would have changed in your opinion? What, if anything, might be a barrier/s to this currently?

Probe:

- Can you tell me about any positive and negative implications there may be from this?
- Can you give me a specific example?

- Do you have any further comments or information about what influences assessment in classrooms or our discussion today?

- Thank you for your time. If there are any other comments that you would like to follow up on then please do contact me. I will be in contact soon with your transcript for approval.

Appendix B: Interview schedule for educational professionals

Prompt items for interview



Photograph of classroom scene to elicit ideas of classroom assessment processes.

	Where the Learner is Going	Where the Learner is Right Now	How to Get There
Teacher	Clarifying and sharing learning intentions and criteria for success	Engineering effective discussions and tasks that elicit evidence of learning	Providing feedback that moves learners forward
Peer	Understanding and Sharing learning intentions and criteria for success	Activating students as instructional resources for one another	
Learner	Understanding learning intentions and criteria for success	Activating students as the owners of their own learning	

Table to illustrate the roles of classroom participants in assessment. Adapted from Wiliam and Thompson (2007).

Interview schedule for educational professionals

Preamble:

- Thank you for:
  - your time. This interview is expected to last for one hour.
  - willingness to participate in this recorded interview today.

- As you may be aware from the participant information sheet, I am interested in the relationship between national assessment policy and classroom assessment processes. I would be interested to hear about your experiences with educational assessment processes from about the year 1998.

- To start our discussion, I have two images that are intended to stimulate our discussion on
  - Different ways you use assessment in your classroom
  - Why you might use a variety of assessment approaches.
  - what you would do with information you gather from these processes
  - Who do you believe this information is useful for.
- Give Teachers the images on Page 1 and allow them to look at them unprompted for a short time before starting the discussion.

Probe:
What kinds of assessment processes do you think the images I have provided you with show?
What other types of assessment do you use? (If only a select few are mentioned).
What informs your decision of when and if you and your students should participate in these assessment processes? Can you tell me about anything else, current or in the past, that might influence what you consider to be effective assessment processes?
Who might be held to account for the results of any assessments you have described?
Are your choices influenced by policy? If so, are you more influenced by national policy, local policies or something else? Does anything else affect your decision to engage or not engage in certain assessment processes?
How knowledgeable are you with your use of assessment and assessment data?
Do external exams come to mind when you are designing assessment processes?
Do you consider any of the assessment processes you have described as necessary for successful learning, and could you explain why? Are any not essential for learning and progress?
What do you consider is the most appropriate tool to use for this purpose?
Are exams fair?

Q1 What does the term educational assessment mean to you?

a) Thinking of your own experiences of educational assessment, do you remember any point where assessment had a personal impact on you?

b) What purposes do you believe educational assessment is currently serving in England / Scotland?
c) In what ways do you think educational assessment impacts on other aspects of day-to-day teaching in your classroom?

Probe:

- *Might there be any positive or negative effects of this?*
- *Can you give me a specific example?*
- *Would this be something that you might do as a teacher anyway to increase student learning?*

Q2 To what extent have national policies influenced assessment processes in your classroom?

a) Have national policies increased or decreased the frequency of assessment in your classroom?
b) To what extent have middle-tier (e.g. LA or MAT Trust) groups influenced your school/classroom policies on assessment processes?
c) Have they caused you to question or adjust the types of assessment tools that you use in your classroom?
d) Do you think that you will continue to conduct assessment processes in this way for the foreseeable future?

Probe:

- *Might there be any positive or negative effects of this?*
- *Can you give me a specific example?*
- *Would this be something that you might do as a teacher anyway to increase student learning?*

Q3 Who do you do assessments for? What is your first consideration when you think about doing an assessment?

c) How has your approach to conducting assessments changed during your career?
d) In what ways are you mindful of the results from assessments or the likely future results of the assessment? How does this affect your practice?
e) What do you consider to be important enough to assess? Why is this important and for who?

Probe:

- *Might there be any positive or negative effects of this?*
- *Can you give me a specific example?*
- *Would this be something that you might do as a teacher anyway to increase student learning?*

---

4 In five years' time, if you could say that classroom assessment has improved, what would have changed in how you use assessment processes or the information from them in your opinion? What, if anything, might be a barrier/s to this currently?

Probe:

a) Might there be any positive or negative effects of this?
b) Can you give me a specific example?

c) Would this be something that you might do as a teacher anyway to increase student learning?

- Do you have any further comments or information about what influences assessment in your classroom or our discussion today?
- Thank you for your time. If there are any other comments that you would like to follow up on, then please do contact me. I will be in contact soon with your transcript for approval.

Appendix C: Consent form for policy influencers

University of Glasgow

College of Social Sciences

Consent Form

Title of Project: Assessment policy in England and Scotland

Name of Researcher: Michael Taylor
Louise Hayward

Please tick as appropriate

Yes ☐ No ☐ I confirm that I have read and understood the Participant Information Sheet for the above study and have had the opportunity to ask questions.

Yes ☐ No ☐ I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason.

Yes ☐ No ☐ I consent to interviews being audio-recorded

Yes ☐ No ☐ I acknowledge that copies of transcripts will be returned to participants for verification.

Yes ☐ No ☐ I acknowledge that participants will be referred to by pseudonym.

I agree that:

Yes ☐ No ☐ All names and other material likely to identify individuals will be anonymised.

Yes ☐ No ☐ The material will be treated as confidential and kept in secure storage at all times.

Yes ☐ No ☐ The material will be destroyed once the project is complete.

Yes ☐ No ☐ The material may be used in future publications, both print and online.

Yes ☐ No ☐ I waive my copyright to any data collected as part of this project.

Yes ☐ No ☐ Other authenticated researchers will have access to this data only if they agree to preserve the confidentiality of the information as requested in this form.

Yes ☐ No ☐ Other authenticated researchers may use my words in publications, reports, web pages, and other research outputs, only if they agree to preserve the confidentiality of the information as requested in this form

Yes ☐ No ☐ I acknowledge the provision of a Privacy Notice in relation to this research project.

I agree to take part in this research study ☐

I do not agree to take part in this research study ☐

Name of Participant …… Signature ……

Date ……

Name of Researcher ……Signature ……

Date ……

Appendix D: Consent form for educational professionals

University of Glasgow

College of Social Sciences

Consent Form

Title of Project: Assessment policy in England and Scotland

Name of Researcher: Michael Taylor
Louise Hayward

Please tick as appropriate

Yes ☐ No ☐	I confirm that I have read and understood the Participant Information Sheet for the above study and have had the opportunity to ask questions.
Yes ☐ No ☐	I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason.
Yes ☐ No ☐	I consent to interviews being audio-recorded
Yes ☐ No ☐	I acknowledge that copies of transcripts will be returned to participants for verification.
Yes ☐ No ☐	I acknowledge that participants will be referred to by pseudonym.

I agree that:

Yes ☐ No ☐	All names and other material likely to identify individuals will be anonymised.
Yes ☐ No ☐	The material will be treated as confidential and kept in secure storage at all times.
Yes ☐ No ☐	The material will be destroyed once the project is complete.
Yes ☐ No ☐	The material may be used in future publications, both print and online.
Yes ☐ No ☐	I waive my copyright to any data collected as part of this project.
Yes ☐ No ☐	Other authenticated researchers will have access to this data only if they agree to preserve the confidentiality of the information as requested in this form.
Yes ☐ No ☐	Other authenticated researchers may use my words in publications, reports, web pages, and other research outputs, only if they agree to preserve the confidentiality of the information as requested in this form
Yes ☐ No ☐	I acknowledge the provision of a Privacy Notice in relation to this research project.

I agree to take part in this research study ☐

I do not agree to take part in this research study ☐

Name of Participant …… Signature ……
Date ……
Name of Researcher ……Signature ……
Date ……

Appendix E: Participant information sheet

University of Glasgow

College of Social Sciences

Participant Information Sheet

My name is Michael Taylor, and I am a research student at the University of Glasgow. The title of this research is: Assessment policy impact in England and Scotland. It aims to seek an understanding of the relationship between national policy in two nations of the United Kingdom and the practice which can be observed in classrooms. The context of these policy relationships is assessment and the purposes to which it is used in classrooms in these two nations.
You are being invited to take part in a research study. Before you decide to take part it is important for you to understand why the research is being done and what it will involve. Please read the following information carefully and discuss it with others if you wish. Ask the researcher/s if there is anything that is not clear or if you would like more information. Take some time to decide whether or not you wish to take part.

Thank you for reading this.

The purpose of this research is to better understand the relationship between national assessment policy and the extent to which it influences classroom practice. In participating in this research, you will help gauge the motivations behind policy, its intent, and its impact. Your participation will involve a one-hour interview which will be via online meeting.

Participation in this research is completely voluntary and will allow you to consider your own involvement in assessment practice and consider your thoughts on best practice. You will have the opportunity to consider what the purpose of assessment is and how it is most constructively used in education. This might be of benefit and improve the understanding of anyone who participates through the discussion of thoughts and understanding.

You will have the right to withdraw at any point in this research. The interview will be recorded and transcribed, and you will be provided with a copy. Should you wish to redact any or all of the transcript you can request this at any time. This will be acted upon without prejudice and you do not need to provide a reason for doing so.

Any details that might identify you by name will be removed. You will be assigned a participant number and school or organisation number. Once the research has finished, all recordings from participant interviews will be deleted.

Please note that confidentiality will be maintained as far as it possible, unless during our conversation I hear anything which makes me worried that someone might be in danger of harm, I might have to inform relevant agencies of this.
All data recorded will be used for purposes directly related to the submission of this PhD thesis. It may also be used in relation to the publication of works associated with the aims of this research including conference papers, journals etc. The data may be used for these purposes, but interview recordings will be deleted as soon as approved transcripts have been produced.
This project has been considered and approved by the College Research Ethics Committee.

If you have any questions relating to this immediate research, or any queries following your interview, transcript or otherwise please contact the researcher or research supervisor below.

Researcher:
Michael Taylor
m.taylor.2@research.gla.ac.uk

Research supervisor:
Louise Hayward
Louise.Hayward@glasgow.ac.uk

If you would like to speak to someone about your wellbeing, stress or any issues raised about your health as a result of engaging with this research, please contact the charity: Education Support using the website or Helpline below:
Helpline:08000562561
Website: https://www.educationsupport.org.uk/get-help/help-for-you/

To pursue any complaint about the conduct of the research: contact the College of Social Sciences Ethics Officer, Dr Muir Houston, email: Muir.Houston@glasgow.ac.uk

Appendix F: Ethics approval form

05 May 2022
Dear Michael Taylor

College of Social Sciences Research Ethics Committee
Project Title: The impact of assessment policy on classrooms in Glasgow and Liverpool Click here to enter project title
Application No: 400210111
The College Research Ethics Committee has reviewed your application and has agreed that there is no objection on ethical grounds to the proposed study. It is happy therefore to approve the project, subject to the following conditions:

- Start date of ethical approval: 05/05/2022
- Project end date: 01/08/2023
- Any outstanding permissions needed from third parties in order to recruit research participants or to access facilities or venues for research purposes must be obtained in writing and submitted to the CoSS Research Ethics Administrator before research commences: socsci-ethics@glasgow.ac.uk
- The research should be carried out only on the sites, and/or with the groups and using the methods defined in the application.
- The data should be held securely for a period of ten years after the completion of the research project, or for longer if specified by the research funder or sponsor, in accordance with the University's Code of Good Practice in Research: (https://www.gla.ac.uk/media/media_490311_en.pdf)
- Any proposed changes in the protocol should be submitted for reassessment as an amendment to the original application. The **Request for Amendments to an Approved Application** form should be used:

  https://www.gla.ac.uk/colleges/socialsciences/students/ethics/forms/staffandpostgraduateresearchstudents/

Yours sincerely,

Dr Susan A. Batchelor
College Ethics Lead

List of References

Addey, C. (2015) 'International Literacy Assessments in Lao PDR and Mongolia: A Global Ritual of Belonging', in M. Hamilton, B. Maddox, and C. Addey (eds) *Literacy as Numbers: Researching the Politics and Practices of International Literacy Assessment*. Cambridge: Cambridge University Press.

Ainscow, M., Barrs, D. and Martin, J. (1998) 'Taking school improvement into the classroom', *Improving Schools*, 1(3), pp. 43–48. Available at: https://doi.org/10.1177/136548029800100316.

Ainscow, M., Chapman, C. and Hadfield, M. (2019) 'THE SCOTTISH ATTAINMENT CHALLENGE Addressing poverty-related outcomes', *CHANGING EDUCATION SYSTEMS*. London, pp. 91–111. Available at: http://ebookcentral.proquest.com/lib/gla/detail.action?docID=5964696.

Alderson, J.C. and Wall, D. (1993) 'Does Washback Exist?', *Applied Linguistics*, 14(2), pp. 115–129. Available at: https://doi.org/10.1093/applin/14.2.115.

Ames, C. and Archer, J. (1988) 'Achievement Goals in the Classroom: Students' Learning Strategies and Motivation Processes', *Journal of Education Psychology*, 80(3), pp. 260–267.

Amrein, A. and Berliner, D. (2003) 'The effects of high-stakes testing on student motivation and learning', *Educational Leadership*, 60(5), pp. 32–38.

Anderson, R.D. (1983) *Education and opportunity in Victorian Scotland: schools and universities*. Oxford: Clarendon press.

Armstrong, P.W., Brown, C. and Chapman, C.J. (2021) 'School-to-school collaboration in England: A configurative review of the empirical evidence', *Review of Education*, 9(1), pp. 319–351. Available at: https://doi.org/10.1002/rev3.3248.

Arnott, M. (2005) 'Devolution, territorial politics and the politics of education', in G. Mooney and G. Scott (eds) *Exploring social policy in the 'new' Scotland*. Policy Press, pp. 239–262. Available at: https://doi.org/10.51952/9781447342427.ch010.

Arnott, M. and Menter, I. (2007) 'The Same but Different? Post-Devolution Regulation and Control in Education in Scotland and England', *European Educational Research Journal*, 6(3), pp. 250–265. Available at: https://doi.org/10.2304/eerj.2007.6.3.250.

Atkin, J.M. and Black, P. (1997) 'Policy Perils of International Comparisons: The TIMSS Case', *Phi Delta Kappa*, 79(1), pp. 22–28.

Atkins, L. and Wallace, S. (2012) *Qualitative research in education*. 1. publ. Los Angeles: SAGE.

Au, W. (2007) 'High-Stakes Testing and Curricular Control: A Qualitative Metasynthesis', *Educational Researcher*, 36(5), pp. 258–267. Available at: https://doi.org/10.3102/0013189X07306523.

Bailey, P.L.J. and Ball, S.J. (2016) 'The coalition government, the general election and the policy ratchet in education: a reflection on the "ghosts" of policy past, present and yet to come', in H. Bochell and M. Powell (eds) *The Coalition Government and Social Policy: Restructuring the Welfare State*. Bristol: Policy Press, pp. 127–152. Available at: https://doi.org/10.1332/policypress/9781447324560.003.0006.

Baird, J., Greatorex, J. and Bell, J.F. (2004) 'What makes marking reliable? Experiments with UK examinations', *Assessment in Education: Principles, Policy & Practice*, 11(3), pp. 331–348. Available at: https://doi.org/10.1080/0969594042000304627.

Baird, J.-A. *et al.* (2017) 'Assessment and learning: fields apart?', *Assessment in Education: Principles, Policy & Practice*, 24(3), pp. 317–350. Available at: https://doi.org/10.1080/0969594X.2017.1319337.

Ball, S. (1993) 'What is Policy? Texts, Trajectories and Toolboxes', *The Australian Journal of Education Studies*, 13(2), pp. 10–17. Available at: https://doi.org/10.1080/0159630930130203.

Ball, S. *et al.* (2012) 'Assessment technologies in schools: "deliverology" and the "play of dominations"', *Research Papers in Education*, 27(5), pp. 513–533. Available at: https://doi.org/10.1080/02671522.2010.550012.

Ball, S. (2018) 'The tragedy of state education in England: Reluctance, compromise and muddle— a system in disarray', *Journal of the British Academy*, 6(September), pp. 207–238. Available at: https://doi.org/10.5871/jba/006.207.

Ball, S., Maguire, M. and Braun, A. (2012) *How Schools do Policy: Policy Enactments in Secondary Schools*. London: Routledge.

Ball, S.J. (2003) 'The teacher's soul and the terrors of performativity', *Journal of Education Policy*, 18(2), pp. 215–228. Available at: https://doi.org/10.1080/0268093022000043065.

Ball, S.J. *et al.* (2011) 'Policy actors: doing policy work in schools', *Discourse: Studies in the Cultural Politics of Education*, 32(4), pp. 625–639. Available at: https://doi.org/10.1080/01596306.2011.601565.

Ball, S.J. (2015a) 'Education, governance and the tyranny of numbers', *Journal of Education Policy*, 30(3), pp. 299–301. Available at: https://doi.org/10.1080/02680939.2015.1013271.

Ball, S.J. (2015b) 'What is policy? 21 years later: reflections on the possibilities of policy research', *Discourse: Studies in the Cultural Politics of Education*, 36(3), pp. 306–313. Available at: https://doi.org/10.1080/01596306.2015.1015279.

Ball, S.J. (2017) *The education debate*. 3rd edn. Bristol: Policy Press.

Ball, S.J. (2021) *The education debate*. 4th edn. Bristol: Policy Press.

Ball, S.J., Bowe, R. and Gewirtz, S. (1996) 'School choice, social class and distinction: the realization of social advantage in education', *Journal of*

Education Policy, 11(1), pp. 89–112. Available at: https://doi.org/10.1080/0268093960110105.

Ball, S.J. and Olmedo, A. (2013) 'Care of the self, resistance and subjectivity under neoliberal governmentalities', *Critical Studies in Education*, 54(1), pp. 85–96. Available at: https://doi.org/10.1080/17508487.2013.740678.

Bandura, A. (2002) 'Social Foundations of Thought and Action', in D. Marks (ed.) *The Health Psychology Reader*. London: Sage Publications Ltd.

Barber, M. (2001) 'High expectations and standards for all, no matter what'.

Barber, M. (2015) *How to run a government: so that citizens benefit and taxpayers don't go crazy*. UK: Allen Lane, an imprint of Penguin Books.

Bassey, M. (1999) *Case study research in educational settings*. Edited by P. Sikes. Buckingham: Open University Press.

Baumfield, V. *et al.* (2010) 'Consultation and engagement? The reshaping of teacher professionalism through curriculum reform in 21st Century Scotland', *Scottish Educational Review*, 42(2), pp. 57–73. Available at: https://doi.org/10.1163/27730840-04202005.

Belsey, C. (2002) 'Creatures of difference', *Post-structuralism : a very short introduction*. Oxford University Press, p. 117.

Bennett, O. (2019) *Michael Gove: a man in a hurry*. London: Biteback Publishing Ltd.

Bennett, R.E. (2011) 'Formative assessment: a critical review', *Assessment in Education: Principles, Policy & Practice*, 18(1), pp. 5–25. Available at: https://doi.org/10.1080/0969594X.2010.513678.

Berlak, H. (2000) 'Cultural Politics, the Science of Assessment and Democratic Renewal of Public Education', in A. Filer (ed.) *Assessment: social practice and social product*. London New York: Routledge Falmer, pp. 189–207.

Bew, P. (2011) *Independent Review of Key Stage 2 testing, assessment and accountability. Final Report*. London: Department for Education. Available at: https://core.ac.uk/download/pdf/4153747.pdf (Accessed: 19 June 2024).

Bhattacharya, A. (2021) 'How Much Choice is Enough? Parental Satisfaction with Secondary School Choice in England and Scotland', *Journal of Social Policy* [Preprint]. Available at: https://doi.org/10.1017/S004727942100091X.

Bhattacharya, A. (2023) 'School Choice in Scotland and England: Divergent Paths and Different Destinations', *FORUM: for promoting 3-19 comprehensive education*, 65(1), pp. 55–62. Available at: https://doi.org/10.3898/forum.2023.65.1.08.

Bhopal, K. and Myers, M. (2023) 'The impact of COVID-19 on A Level exams in England: Students as consumers', *British Educational Research Journal*, 49(1), pp. 142–157. Available at: https://doi.org/10.1002/berj.3834.

Biesta, G. (2007) ‘Why “What Works” Won’t Work: Evidence-Based Practice and the Democratic Deficit in Educational Research’, *Educational Theory*, 57(1), pp. 1–22. Available at: https://doi.org/10.1111/j.1741-5446.2006.00241.x.

Biesta, G. (2009) ‘Good education in an age of measurement: on the need to reconnect with the question of purpose in education’, *Educational Assessment, Evaluation and Accountability(formerly: Journal of Personnel Evaluation in Education)*, 21(1), pp. 33–46. Available at: https://doi.org/10.1007/s11092-008-9064-9.

Biesta, G., Priestley, M. and Robinson, S. (2015) ‘The role of beliefs in teacher agency’, *Teachers and Teaching: Theory and Practice*, 21(6), pp. 624–640. Available at: https://doi.org/10.1080/13540602.2015.1044325.

Black, P. (1997) *Testing: Friend or Foe?: Theory and Practice of Assessment and Testing*. 1st edn. Oxon: Routledge.

Black, P. (1999) ‘Assessment, learning theories and testing systems’, in P. Murphy (ed.) *Learners, Learning and Assessment*. London: Paul Chapman Publishing, pp. 120–121.

Black, P. and Wiliam, D. (1998a) ‘Assessment and Classroom Learning’, *Assessment in Education: Principles, Policy & Practice*, 5(1), pp. 7–74. Available at: https://doi.org/10.1080/0969595980050102.

Black, P. and Wiliam, D. (1998b) ‘Inside the Black Box Raising Standards Through Classroom Assessment’, *Phi Delta Kappan*, 80(2), pp. 139–148.

Black, P. and Wiliam, D. (2003) ‘“In praise of educational research”: formative assessment’, *British Educational Research Journal*, 29(5), pp. 623–637. Available at: https://doi.org/10.1080/0141192032000133721.

Black, P. and Wiliam, D. (2012) ‘The Reliability of Assessments’, *Assessment and Learning*. 2nd edn. London: SAGE Publications Ltd. Available at: https://doi.org/10.4135/9781446250808.

Black, P. and Wiliam, D. (2018) ‘Classroom assessment and pedagogy’, *Assessment in Education: Principles, Policy & Practice*, 25(6), pp. 551–575. Available at: https://doi.org/10.1080/0969594X.2018.1441807.

Bliem, C.L. and Davinroy, K.H. (1997) ‘Teachers’ Beliefs About Assessment and Instruction in Literacy’.

Bloxham, S., Hughes, C. and Adie, L. (2016) ‘What’s the point of moderation? A discussion of the purposes achieved through contemporary moderation practices’, *Assessment & Evaluation in Higher Education*, 41(4), pp. 638–653. Available at: https://doi.org/10.1080/02602938.2015.1039932.

Blunkett, D. (2000) ‘Influence or irrelevance: can social science improve government?’, *Research Intelligence*, 71, pp. 12–21.

Bousted, M. (2017) ‘Assessment should be a support system for all’, *The Times Educational Supplement*. Available at:

https://www.tes.com/magazine/archived/assessment-should-be-support-system-all (Accessed: 28 May 2026).

Boyle, B. and Bragg, J. (2005) 'No science today—the demise of primary science', *The Curriculum Journal*, 16(4), pp. 423–437. Available at: https://doi.org/10.1080/09585170500384115.

Boyle, B. and Bragg, J. (2006) 'A curriculum without foundation', *British Educational Research Journal*, 32(4), pp. 569–582. Available at: https://doi.org/10.1080/01411920600775225.

Bradbury, A. (2019) 'Datafied at four: the role of data in the "schoolification" of early childhood education in England', *Learning, Media and Technology*, 44(1), pp. 7–21. Available at: https://doi.org/10.1080/17439884.2018.1511577.

Bradbury, A. *et al.* (2022) 'Crisis policy enactment: primary school leaders' responses to the COVID-19 pandemic in England', *Journal of Education Policy*, pp. 1–21. Available at: https://doi.org/10.1080/02680939.2022.2097316.

Brady, A. and Richardson, M. (2025) 'In Grades We Trust? Challenging Beliefs About Assessment in Education', in D. Wyse et al. (eds) *The BERA-Sage Handbook of Research-Informed Education Practice and Policy*. London: Sage.

Braun, A. *et al.* (2011) 'Taking context seriously: towards explaining policy enactments in the secondary school', *Discourse: Studies in the Cultural Politics of Education*, 32(4), pp. 585–596. Available at: https://doi.org/10.1080/01596306.2011.601555.

Braun, A. and Maguire, M. (2020) 'Doing without believing–enacting policy in the English primary school', *Critical Studies in Education*, 61(4), pp. 433–447. Available at: https://doi.org/10.1080/17508487.2018.1500384.

Braun, A., Maguire, M. and Ball, S.J. (2010) 'Policy enactments in the UK secondary school: Examining policy, practice and school positioning', *Journal of Education Policy*, 25(4), pp. 547–560. Available at: https://doi.org/10.1080/02680931003698544.

Braun, H.I. and Marion, S.F. (2022) 'Accountability and assessment in U.S. education: let's not allow another crisis go to waste!', *Assessment in Education: Principles, Policy & Practice*, pp. 1–20. Available at: https://doi.org/10.1080/0969594X.2022.2068503.

Braun, V. and Clarke, V. (2021) 'One size fits all? What counts as quality practice in (reflexive) thematic analysis?', *Qualitative Research in Psychology*, 18(3), pp. 328–352. Available at: https://doi.org/10.1080/14780887.2020.1769238.

Braun, V. and Clarke, V. (2022) *Thematic Analysis a Practical Guide*. Edited by A. Maher. SAGE Publications.

Brighouse, T. and Waters, M. (2022) *About Our Schools Improving on previous best*. Crown House Publishing.

Brinkmann, S. and Kvale, S. (2015) *Interviews: Learning the Craft of Qualitative Research Interviewing*. Third. Sage.

Broadfoot, P. (1999) ‘Empowerment or performativity? English assessment policy in the late twentieth century.’, in J. Furlong and R. Phillips (eds) *Education, Reform and the State Twenty Five Years of Politics, Policy and Practice*. 1st edn. Paper presented at the Assessment Reform Group Symposium on Assessment Policy at the British Educational Research Association annual conference: Routledge Falmer, pp. 136–155.

Brown, G.T.L. (2004) ‘Teachers’ conceptions of assessment: Implications for policy and professional development’, *Assessment in Education: Principles, Policy and Practice*, 11(3), pp. 301–318. Available at: https://doi.org/10.1080/0969594042000304609.

Bryce, T. (2018) ‘68 Assessment in Scottish Schools’, in T.G.K. Bryce et al. (eds) *Scottish Education*. Edinburgh University Press, pp. 589–604. Available at: https://doi.org/10.1515/9781474437851-069.

Bryce, T. and Humes, W. (2018a) ‘4 Scottish Secondary Education’, in T.G.K. Bryce et al. (eds) *Scottish Education*. Edinburgh University Press, pp. 45–55. Available at: https://doi.org/10.1515/9781474437851-005.

Bryce, T. and Humes, W. (2018b) ‘69 The Scottish Qualifications Authority’, in T.G.K. Bryce et al. (eds) *Scottish Education*. Edinburgh University Press, pp. 605–614. Available at: https://doi.org/10.1515/9781474437851-070.

Bryce, T. and Humes, W. (2018c) ‘Scottish Secondary Education’, *Scottish Education*. 5th edn. Edinburgh: Edinburgh University Press.

Buckler, S. and Moore, H. (2023) ‘Qualitative Considerations and Methods’, *Essentials of Research Methods in Education*. SAGE Publications, pp. 223–243.

Buie, E. (2010) ‘Pisa study reveals social rift, despite Russell’s cheers’, *TeS* [Preprint]. Available at: https://www.tes.com/magazine/archive/pisa-study-reveals-social-rift-despite-russells-cheers.

Buie, E. (2012) ‘Timetable shift for National assessments’, *The Times Educational Supplement*, p. 6.

Butler, C. (2025) ‘Do Governments Always Take Unpopular Decisions Knowingly? How Perceptual Inaccuracy Affects Policy Decisions’, *Representation*, 61(1), pp. 21–38. Available at: https://doi.org/10.1080/00344893.2023.2281312.

Chapman, C. (2000) ‘Inspection and self-review’, *Improving Schools*, 3, pp. 57–63.

Chapman, C. (2011) ‘School Improvement research and practice A case of back to the future?’, in C. Chapman et al. (eds) *School Effectiveness and Improvement Research, Policy and Practice*. London: Routledge, pp. 27–43.

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), pp. 46–60. Available at: https://doi.org/10.1177/1741143213494883.

Chapman, C. (2019) ‘From hierarchies to networks: Possibilities and pitfalls for educational reform of the middle tier’, *Journal of Educational Administration*, 57(5), pp. 554–570. Available at: https://doi.org/10.1108/JEA-12-2018-0222.

Childs, A. and Baird, J.-A. (2020) ‘General Certificate of Secondary Education (GCSE) and the assessment of science practical work: an historical review of assessment policy’, *The Curriculum Journal*, 31(3), pp. 357–378. Available at: https://doi.org/10.1002/curj.20.

Codd, J.A. (1988) ‘The construction and deconstruction of educational policy documents’, *Journal of Education Policy*, 3(3), pp. 235–247. Available at: https://doi.org/10.1080/0268093880030303.

Coe, R. *et al.* (eds) (2017) *Research methods & methodologies in education*. 2nd edition. Los Angeles London New Delhi Singapore Washington DC Melbourne: SAGE.

Coffey, J.E. *et al.* (2011) ‘The missing disciplinary substance of formative assessment’, *Journal of Research in Science Teaching*, 48(10), pp. 1109–1136. Available at: https://doi.org/10.1002/tea.20440.

Cohen, L., Manion, L. and Morrison, K. (2018a) ‘Critical educational research’, *Research methods in education*. eighth edition. London New York: Routledge, pp. 51–67.

Cohen, L., Manion, L. and Morrison, K. (2018b) *Research methods in education*. Eighth edition / Louis Cohen, Lawrence Manion and Keith Morrison. Oxon: Routledge.

Cohen, L., Manion, L. and Morrison, K. (2018c) *Research Methods in Education*. 7th edn. Routledge.

Correia, C.F. and Harrison, C. (2020) ‘Teachers’ beliefs about inquiry-based learning and its impact on formative assessment practice’, *Research in Science & Technological Education*, 38(3), pp. 355–376. Available at: https://doi.org/10.1080/02635143.2019.1634040.

Coughlan, S. (2003) *‘Stop the Sats’ campaign launched*, *BBC News*. Available at: http://news.bbc.co.uk/1/hi/education/3028766.stm (Accessed: 13 May 2026).

Courtney, S.J. (2016) ‘Post-panopticism and school inspection in England’, *British Journal of Sociology of Education*, 37(4), pp. 623–642. Available at: https://doi.org/10.1080/01425692.2014.965806.

Cowie, M. and Croxford, L. (2007) *“Intelligent Accountability: Sound-Bite or Sea-Change?* University of Edinburgh: Edinburgh: Centre for Educational Sociology.

Craske, J. (2021) ‘Logics, rhetoric and “the blob”: Populist logic in the Conservative reforms to English schooling’, *British Educational Research Journal*, 47(2), pp. 279–298. Available at: https://doi.org/10.1002/berj.3682.

Crehan, L. (2016) *Cleverlands: The Secrets Behind the Success of the World’s Education Superpowers*. London: Unbound.

Creswell, J.W. and Poth, C.N. (2018) *Qualitative inquiry & research design: choosing among five approaches*.

Cuban, L. (1993) *How Teachers Taught: Constancy and Change in American Classrooms, 1990-1990*. Teachers College Press.

Curtis, P. (2008) 'Sats for 14-year-olds are scrapped', *The Guardian* [Preprint].

Daly, A.L. *et al*. (2012) 'Assessment reform: students' and teachers' responses to the introduction of stretch and challenge at A-level', *The Curriculum Journal*, 23(2), pp. 139–155. Available at: https://doi.org/10.1080/09585176.2012.678683.

Dayal, H. (2021) 'How Teachers use Formative Assessment Strategies during Teaching: Evidence from the Classroom', *Australian Journal of Teacher Education*, 46(7). Available at: https://doi.org/10.14221/ajte.2021v46n7.1.

DeLuca, C. and Johnson, S. (2017) 'Developing assessment capable teachers in this age of accountability', *Assessment in Education: Principles, Policy & Practice*, 24(2), pp. 121–126. Available at: https://doi.org/10.1080/0969594X.2017.1297010.

DeLuca, C. and Klinger, D.A. (2010) 'Assessment literacy development: identifying gaps in teacher candidates' learning', *Assessment in Education: Principles, Policy & Practice*, 17(4), pp. 419–438. Available at: https://doi.org/10.1080/0969594X.2010.516643.

Department for Children, Schools and Families (2008) *The Assessment for Learning Strategy*. 978-1-84775-147–8. Nottingham.

Department for Education (2010) 'The importance of teaching: Schools White Paper 2010'. Available at: https://doi.org/10.1662/0002-7685(2006)68%5B197:TIOTTN%5D2.0.CO;2.

Department for Education (2011) *Press release: National curriculum review launched.*, *GOV.UK*. Available at: https://www.gov.uk/government/news/national-curriculum-review-launched (Accessed: 23 July 2025).

Department for Education (2025) *Curriculum and Assessment Review final report: Building a world-class curriculum for all*. Available at: https://assets.publishing.service.gov.uk/media/690b96bbc22e4ed8b051854d/Curriculum_and_Assessment_Review_final_report_-_Building_a_world-class_curriculum_for_all.pdf (Accessed: 25 April 2026).

Department for Education and Employment (1999) 'The National Curriculum for England'. Available at: https://doi.org/10.2307/j.ctvh1dhfr.13.

Department for Education and Skills (2001) *Schools Achieving Success*. White Paper Cm 5230. London: The Stationery Office, pp. 76–76. Available at: https://assets.publishing.service.gov.uk/media/5a7edcf8ed915d74e6226efb/Schools_Achieving_Success.pdf (Accessed: 27 March 2024).

Department for Education and Skills (2002) ‘Making the most of booster classes’. Department of Education and Science (DES). Available at: https://www.stem.org.uk/resources/elibrary/resource/31015/booster-lessons-guidance-and-training-materials.

Department for Education and Skills (2007) *Making Good Progress: How can we help every pupil to make good progress at school*. Consultation. London.

Devine, T.M. (2016) *Independence or union: Scotland’s past and Scotland’s present*. UK: Penguin.

Dewey, J. (1997) *How We Think*. Mineola: Dover Publications.

DiMaggio, P.J. and Powell, W.W. (1983) ‘The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields’, *American Sociological Review*, 48(2), pp. 147–160. Available at: https://doi.org/10.2307/2095101.

Donaldson, G. (2013) ‘Assessment must play a greater part in teaching’, *The Times Educational Supplement*, (2317), pp. 37–37.

Dorrell, E. (2018) ‘A little flexibility could go a long way in improving the profession’, *The Times Education Supplement* [Preprint].

Dunbar, S.B., Koretz, D.M. and Hoover, H. d. (1991) ‘Quality Control in the Development and Use of Performance Assessments’, *Applied Measurement in Education*, 4(4), p. 289. Available at: https://doi.org/10.1207/s15324818ame0404_3.

Dundee City Council (2021) ‘Dundee Poverty Profile 2021’, *Dundee City Poverty Profile*. Available at: https://dundeecity.gov.uk/sites/default/files/publications/poverty_profile2021.pdf.

Dundee City Council (2022) *Dundee Poverty Profile 2022*. Dundee: Dundee City Council. Available at: https://www.dundeecity.gov.uk/sites/default/files/publications/poverty_profile_2022.pdf (Accessed: 20 February 2026).

Education Reform Act 1988 (1988) *c. 40*. Statute Law Database. Available at: https://www.legislation.gov.uk/ukpga/1988/40/contents (Accessed: 15 September 2025).

Education Scotland (2023) ‘Scottish Attainment Challenge’. Available at: https://education.gov.scot/learning-in-scotland/programmes/scottish-attainment-challenge/.

Eisenbach, B.B. (2012) ‘Teacher Belief and Practice in a Scripted Curriculum’, *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 85(4), pp. 153–156. Available at: https://doi.org/10.1080/00098655.2012.663816.

Elliott, V.F. (2023) *Foundations of educational research*. Bloomsbury academic.

Evans, G. (2022) ‘Back to the future? Reflections on three phases of education policy reform in Wales and their implications for teachers’, *Journal of*

Educational Change, 23(3), pp. 371–396. Available at: https://doi.org/10.1007/s10833-021-09422-6.

Executive, S. (2004) 'Curriculum for Excellence building the curriculum 3: a framework for learning and teaching'.

Ezzy, D. (2002) *Qualitative analysis: practice and innovation*. London new York: Routledge, Taylor & Francis Group (Social research today).

Fairclough, N. (2013) *Critical discourse analysis: the critical study of language*. 2nd edn. London: Routledge.

Fang, Z. (1996) 'A review of research on teacher beliefs and practices', *Educational Research*, 38(1), pp. 47–65. Available at: https://doi.org/10.1080/0013188960380104.

Forde, C. and Torrance, D. (2021) 'Leadership at all levels: system alignment through empowerment in Scottish education?', *School Leadership & Management*, 41(1–2), pp. 22–40. Available at: https://doi.org/10.1080/13632434.2020.1825372.

Forrester, G. and Garratt, D. (2016) *Education Policy Unravelled*. 2nd edn. London: Bloomsbury Academic.

Foucault, M. (2001) *The Order of Things: An archaeology of the human sciences*. London: Routledge.

Fulcher, G. (2012) 'Assessment literacy for the language classroom', *Language Assessment Quarterly*, 9(2), pp. 113–132. Available at: https://doi.org/10.1080/15434303.2011.642041.

Fullan, M. (1993) *Change Forces*. London: Falmer.

Fullan, M. (2017) *The Meaning of Educational Change*. Fourth. Oxon: Routledge. Available at: https://ebookcentral.proquest.com/lib/uocuk/detail.action?docID=4513498.

Gale, T. (2018) 'What's not to like about RCTs in education?', in A. Childs and I. Menter (eds) *Mobilising Teacher Researchers: Challenging Educational Inequality*. Routledge, pp. 207–223.

Gale, T. and Parker, S. (2018) '101 The Future of (Scottish) Education: An International Perspective', in T.G.K. Bryce et al. (eds) *Scottish Education*. Edinburgh University Press, pp. 937–949. Available at: https://doi.org/10.1515/9781474437851-102.

Gale, T.C. (1994) 'Beyond caricature: Exploring theories of educational policy production and implementation', *The Australian Educational Researcher*, 21(2), pp. 1–12. Available at: https://doi.org/10.1007/BF03219564.

Gardner, J. *et al.* (2008) 'Changing Assessment Practice: Process , Principles and Standards'.

Gardner, J. and Cowan, P. (2007) 'The fallibility of high stakes "11-plus " testing in Northern Ireland', *Assessment in Education: Principles, Policy and Practice*, 12(2), pp. 145–165. Available at: https://doi.org/10.1080/09695940500143837.

Gardner, J. and Galanouli, D. (2016) 'Teachers' Perceptions of Assessment', in D. Wyse, L. Hayward, and J. Pandya, *The SAGE Handbook of Curriculum, Pedagogy and Assessment: Two Volume Set*. 1 Oliver's Yard, 55 City Road London EC1Y 1SP: SAGE Publications Ltd, pp. 710–723. Available at: https://doi.org/10.4135/9781473921405.n44.

Gerring, J. (2017) *Case Study Research*. Cambridge University Press.

Gewirtz, S. *et al.* (2021) 'What's wrong with "deliverology"? Performance measurement, accountability and quality improvement in English secondary education', *Journal of Education Policy*, 36(4), pp. 504–529. Available at: https://doi.org/10.1080/02680939.2019.1706103.

Gibb, N. (2015a) 'Assessment after levels'. Available at: https://www.gov.uk/government/speeches/assessment-after-levels (Accessed: 4 November 2024).

Gibb, N. (2015b) 'Reforming education through international evidence [Online]'. Available at: https://www.gov.uk/government/speeches/nick-gibb-reforming-education-through-international-evidence.

Gibb, N. and Pearl, R. (2025) *REFORMING LESSONS: Why English Schools Have Improved Since 2010 and How This Was Achieved*. S.l.: ROUTLEDGE. Available at: https://doi.org/10.4324/9781003533474.

Gillies, D. (2018) 'Policy and politics in Scottish education', in T.G.K. Bryce et al. (eds) *Scottish education*. Fifth edition. Edinburgh: Edinburgh University Press, pp. 86–96. Available at: https://www.degruyter.com/document/doi/10.1515/9781474437851-009/html (Accessed: 6 May 2024).

Gipps, C. (2011) 'National assessment: a comparison of English and American trends', in P. Broadfoot, R. Murphy, and H. Torrance (eds) *Changing Educational Assessment : International Perspectives and Trends*. Oxford, UNITED KINGDOM: Taylor & Francis Group. Available at: http://ebookcentral.proquest.com/lib/gla/detail.action?docID=957317.

Gove, M. (2010) 'Pisa slip should put a rocket under our world-class ambitions and drive us to win the education space race', *The Times Educational Supplement*, pp. 25–25.

Gove, M. (2012) 'Michael Gove speech on academies', 4 January. Available at: https://www.gov.uk/government/speeches/michael-gove-speech-on-academies (Accessed: 21 July 2025).

Granoulhac, F. (2017) 'Excellence at What Cost? Austerity and the Reform of the School System in England (2010-2016)', *The Political Quarterly*, 88(3), pp. 434–442. Available at: https://doi.org/10.1111/1467-923X.12375.

Gray, J. (2010) 'Probing the limits of systemic reform: the English case', *Second international handbook of educational change*. Dordrecht: Springer, pp. 293–307.

Grek, S. (2009) 'Governing by numbers: The PISA "effect" in Europe', *Journal of Education Policy*, 24(1), pp. 23–37. Available at: https://doi.org/10.1080/02680930802412669.

Gu, Q., Sammons, P. and Chen, J. (2018) 'How Principals of Successful Schools Enact Education Policy: Perceptions and Accounts From Senior and Middle Leaders', *Leadership and Policy in Schools*, 17(3), pp. 373–390. Available at: https://doi.org/10.1080/15700763.2018.1496344.

Guardian (2020) 'Ofqual suspends A-level appeal criteria hours after announcing them', *The Guardian* [Preprint].

Gunter, H. and Chapman, C. (2009) 'A decade of New Labour reform of education', in C. Chapman and H. Gunter (eds) *Radical Reforms: perspectives on an era of educational change*. Routledge, pp. 1–13.

Hall, D. and Raffo, C. (2009) 'New Labour and breaking the education and poverty link A conceptual account of its educational policies', in H.M. Gunther and C. Chapman (eds) *Radical Reforms: Perspectives on an era of educational change*. First, pp. 155–168.

Hamilton, L. and Corbett-Whittier, C. (2013) *Using Case Study in Education Research*. London: Sage.

Hansard, H.D. (2007) '"Making Good Progress" Document'.

Harlen, W. (2005a) 'Teachers' summative practices and assessment for learning – tensions and synergies', *Curriculum Journal*, 16(2), pp. 207–223. Available at: https://doi.org/10.1080/09585170500136093.

Harlen, W. (2005b) 'Trusting teachers' judgement: Research evidence of the reliability and validity of teachers' assessment used for summative purposes', *Research Papers in Education*, 20(3), pp. 245–270. Available at: https://doi.org/10.1080/02671520500193744.

Harlen, W. (2006) 'The role of assessment in developing motivation for learning', in J. Gardner (ed.) *Assessment and learning*. London: Sage, pp. 61–80.

Harlen, W. (2007) *Assessment of Learning*, *Assessment of Learning*. Sage Publications Ltd.

Harlen, W. (2010) 'Professional learning to support teacher assessment', in J. Gardner et al. (eds) *Developing Teacher Assessment*. Open University Press, pp. 100–129.

Harlen, W. (2013) *Assessment & inquiry-based science education: issues in policy and practice*. Trieste, Italy: Global Network of Science Academies.

Harlen, W. and James, M. (1997) 'Assessment and Learning: differences and relationships between formative and summative assessment', *Assessment in*

Education: Principles, Policy & Practice, 4(3), pp. 365–379. Available at: https://doi.org/10.1080/0969594970040304.

Hatcher, R. and Troyna, B. (1994) 'The policy cycle a ball by ball account', *Journal of Education Policy*, 9(2), pp. 155–170. Available at: https://doi.org/10.1080/0268093940090204.

Hayward, E.L. (2007) 'Curriculum, pedagogies and assessment in Scotland: the quest for social justice. "Ah kent yir faither"', *Assessment in Education: Principles, Policy and Practice*, 14(2), pp. 251–268. Available at: https://doi.org/10.1080/09695940701480178.

Hayward, E.L. and Hutchinson, C. (2013) '"Exactly what do you mean by consistency?" Exploring concepts of consistency and standards in Curriculum for Excellence in Scotland', *Assessment in Education: Principles, Policy and Practice*, 20(1), pp. 53–68. Available at: https://doi.org/10.1080/0969594X.2012.742423.

Hayward, L. *et al.* (2008) *Just Making Them Think : A tension between teaching and assessment in the high stakes stages. (Commissioned from the Universities of Glasgow and Strathclyde by the Scottish Qualifications Authority)*, pp. 49–49. Available at: https://www.sqa.org.uk/sqa/files_ccc/Report 13_Make_Them_Think.pdf.

Hayward, L., Menter, I., Baumfield, V., Daugherty, R., Akhtar, N., Doyle, L., Elliot, D., Hulme, M., MacBride, G., *et al.* (2012) *Assessment at transition – Research report to Scottish Government*. Glasgow: University of Glasgow.

Hayward, L. (2012) 'Has anyone really learned anything? Assessment in the primary school', in D. Wyse et al. (eds) *Creating the Curriculum*. 0 edn. Routledge, pp. 71–85. Available at: https://doi.org/10.4324/9780203115916-10.

Hayward, L., Menter, I., Baumfield, V., Daugherty, R., Akhtar, N., Doyle, L., Elliot, D., Hulme, M., Hutchinson, C., *et al.* (2012) 'University of Glasgow Report Assessment at Transition'.

Hayward, L. (2015) 'Assessment is learning: the preposition vanishes', *Assessment in Education: Principles, Policy & Practice*, 22(1), pp. 27–43. Available at: https://doi.org/10.1080/0969594X.2014.984656.

Hayward, L. (2018) 'Notes from a Small Country: Scotland', in C. Wyatt-Smith and L. Adie (eds) *Innovation and Accountability in Teacher Education*. Singapore: Springer (Teacher Education, Learning Innovation and Accountability). Available at: https://doi.org/10.1007/978-981-13-2026-2.

Hayward, L. *et al.* (2023) 'National qualifications in Scotland: A lightning rod for public concern about equity during the pandemic', *European Journal of Education*, 58(1), pp. 83–97. Available at: https://doi.org/10.1111/ejed.12543.

Hayward, L. and Hedge, N. (2005) 'Travelling towards change in assessment : Policy , practice and research in education Travelling towards change in assessment : policy , practice and research in education', *Assessment in Education: Principles, Policy & Practice*, 12(1), pp. 55–75. Available at: https://doi.org/10.1080/0969594042000333913.

Hayward, L., Macbride, G. and Spencer, E. (2016) 'The Intersection of International Student Assessment and Educational Policy Development', in L. Volante (ed.) *The intersection of international student assessment and educational policy development*. Routledge, pp. 57–74.

Hayward, L. and O'Leary, M. (2022) 'High stakes assessment in the era of COVID-19: interruption, transformation, regression or business as usual?', *Assessment in Education: Principles, Policy & Practice*, 29(5), pp. 505–517. Available at: https://doi.org/10.1080/0969594X.2022.2139339.

Hayward, L., Priestley, M. and Young, M. (2004) 'Ruffling the calm of the ocean floor: merging practice, policy and research in assessment in Scotland', *Oxford Review of Education*, 30(3), pp. 397–415. Available at: https://doi.org/10.1080/0305498042000260502.

Hayward, L. and Spencer, E. (2010) 'The complexities of change: formative assessment in Scotland', *The Curriculum Journal*, 21(2), pp. 161–177. Available at: https://doi.org/10.1080/09585176.2010.480827.

Heath, A. *et al.* (2013) 'Education under New Labour, 1997 - 2010', *Oxford Review of Economic Policy*, 29(1), pp. 227–247.

Henry, H. (2015) '"Judge me on closing the attainment gap"', *The Times Educational Supplement* [Preprint].

Hopfenbeck, T.N. (2018) 'Classroom assessment, pedagogy and learning – twenty years after Black and Wiliam 1998', *Assessment in Education: Principles, Policy & Practice*, 25(6), pp. 545–550. Available at: https://doi.org/10.1080/0969594X.2018.1553695.

House of Commons Children, Schools and Families Committee (2010) *School Accountability: First Report of Session 2009–10, Volume I*. HC 88-I. London: House of Commons. Available at: https://www.education-uk.org/documents/pdfs/2010-csfc-school-accountability.pdf.

Humes, W. (1986) *The leadership class in Scottish education*. Edinburgh: John Donald Publishers (Book, Whole). Available at: https://go.exlibris.link/cLtzCd97.

Humes, W. *et al.* (2018) '103 The Future of Scottish Education', in T.G.K. Bryce et al. (eds) *Scottish Education*. Edinburgh University Press, pp. 969–979. Available at: https://doi.org/10.1515/9781474437851-104.

Humes, W. (2023) 'Scottish Education: A Crisis of Confidence and Trust', *FORUM*, 65(1), pp. 19–29. Available at: https://doi.org/10.3898/forum.2023.65.1.05.

Hutchinson, C. and Hayward, L. (2005) 'The journey so far: assessment for learning in Scotland', *The Curriculum Journal*, 16(2), pp. 225–248. Available at: https://doi.org/10.1080/09585170500136184.

Hutchinson, C. and Young, M. (2011) 'Assessment for learning in the accountability era: Empirical evidence from Scotland', *Studies in Educational Evaluation*, 37(1), pp. 62–70. Available at: https://doi.org/10.1016/j.stueduc.2011.03.007.

Isaacs, T. *et al.* (2013) *Key concepts in Educational Assessment*. London: Sage Publications Ltd, p. 58.

James, M. (2008) 'Assessment and Learning', in S. Swaffield (ed.) *Unlocking Assessment Understanding for reflection and application*. Oxon: Routledge, pp. 20–35.

Jerrim, J. (2021) 'PISA 2018 in England, Northern Ireland, Scotland and Wales: Is the data really representative of all four corners of the UK?', *Review of Education*, 9(3), p. e3270. Available at: https://doi.org/10.1002/rev3.3270.

Jerrim, J. and Sims, S. (2021) 'School accountability and teacher stress: international evidence from the OECD TALIS study', *Educational Assessment, Evaluation and Accountability* [Preprint]. Available at: https://doi.org/10.1007/s11092-021-09360-0.

Johnson, M. and Crisp, V. (2009) 'An Exploration of the Effect of Pre-Release Examination Materials on Classroom Practice in the UK', *Research in Education*, 82(1), pp. 47–59. Available at: https://doi.org/10.7227/RIE.82.4.

Johnson, S. (2016) 'National assessment and intelligent accountability', in D. Wyse, L. Hayward, and J. Pandya (eds) *SAGE Handbook on Curriculum, Pedagogy and Assessment*. Sage, pp. 853–870.

Kelly, A. (2021) 'A tale of two algorithms: The appeal and repeal of calculated grades systems in England and Ireland in 2020', *British Educational Research Journal*, 47(3), pp. 725–741. Available at: https://doi.org/10.1002/berj.3705.

Kennedy, Aileen (2018) '89 Scottish Teachers: Then and Now', in T.G.K. Bryce et al. (eds) *Scottish Education*. Edinburgh University Press, pp. 825–835. Available at: https://doi.org/10.1515/9781474437851-090.

Kim, L.E. and Asbury, K. (2020) '"Like a rug had been pulled from under you": The impact of COVID-19 on teachers in England during the first six weeks of the UK lockdown', *British Journal of Educational Psychology*, 90(4), pp. 1062–1083. Available at: https://doi.org/10.1111/bjep.12381.

Kirton, A. *et al.* (2007) 'Revolution, evolution or a Trojan horse? Piloting assessment for learning in some Scottish primary schools', *British Educational Research Journal*, 33(4), pp. 605–627. Available at: https://doi.org/10.1080/01411920701434136.

Kluger, A.N. and DeNisi, A. (1996) 'The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory.', *Psychological Bulletin*, 119(2), pp. 254–284. Available at: https://doi.org/10.1037/0033-2909.119.2.254.

Knodel, P., Martens, K. and Niemann, D. (2013) 'PISA as an ideational roadmap for policy change: Exploring Germany and England in a comparative perspective', *Globalisation, Societies and Education*, 11(3), pp. 421–441. Available at: https://doi.org/10.1080/14767724.2012.761811.

Koretz, D.M., McCaffrey, D.F. and Hamilton, L.S. (2001) 'Toward a Framework for Validating Gains Under High-Stakes Conditions: (648282011-001)'. Available at: https://doi.org/10.1037/e648282011-001.

Kreijkes, P. and Johnson, M. (2023) 'Who controls what and how? A comparison of regulation and autonomy in the UK nations' education systems', *Research Matters: A Cambridge University Press & Assessment publication*, (35), pp. 60–79.

Lather, P. (1986) 'Research as Praxis', *Harvard Educational Review*, 56(3), pp. 257–278. Available at: https://doi.org/10.17763/haer.56.3.bj2h231877069482.

Le Grand, J. (2006) *Motivation, agency, and public policy: of knights and knaves, pawns and queens*. Oxford: Univ. Press.

Leahy, S. and Wiliam, D. (2012) 'From Teachers to Schools: Scaling up Professional Development for Formative Assessment', in J. Gardner (ed.) *Assessment and Learning*. London, pp. 49–71.

Leat, D., Livingston, K. and Priestley, M. (2013) 'Curriculum deregulation in England and Scotland: Different directions of travel', in W. Kuiper and J. Berkvens (eds) *Balancing Curriculum Regulation and Freedom across Europe*. Enschede, Netherlands: Gildeprint, pp. 1–12.

Lincoln, Y.S. and Guba, E.G. (1985) *Naturalistic inquiry*. Nachdr. Newbury Park, Calif.: Sage.

Lingard, B., Martino, W. and Rezai-Rashti, G. (2013) 'Testing regimes, accountabilities and education policy: Commensurate global and national developments', *Journal of Education Policy*, 28(5), pp. 539–556. Available at: https://doi.org/10.1080/02680939.2013.820042.

Lingard, B. and Sellar, S. (2014) 'Representing for Country: Scotland, PISA and New Spatialities of Educational Governance', *Scottish Educational Review*.

Liverpool City Council (2019) *The Index of Multiple Deprivation*. Available at: https://liverpool.gov.uk/media/1359213/imd-2019-liverpool-analysis-main-report.pdf.

Liverpool City Council (2021) 'Demographics headline indicators', *Liverpool City Council*. Available at: https://liverpool.gov.uk/council/key-statistics-and-data/headline-indicators/demographics/.

Looney, A. *et al.* (2018) 'Reconceptualising the role of teachers as assessors teacher assessment identity.pdf', *Assessment in Education: Principles, Policy & Practice*, 25(5). Available at: https://doi.org/10.1080/0969594X.2016.1268090.

Madaus, G.F. (1988) 'The distortion of teaching and testing: High-stakes testing and instruction', *Peabody Journal of Education*, 65(3), pp. 29–46. Available at: https://doi.org/10.1080/01619568809538611.

Marshall, B. and Drummond, M.J. (2006) 'How teachers engage with assessment for learning: Lessons from the classroom', *Research Papers in Education*, 21(2), pp. 133–149. Available at: https://doi.org/10.1080/02671520600615638.

Mason, C. (2022) ‘School spending on staff development drops by more than a third | Tes’, *The Times Educational Supplement*, 21 October. Available at: https://www.tes.com/magazine/news/general/school-staff-development-spending-falls-over-third (Accessed: 10 September 2024).

McEnaney, J. (2021) *Class Rules: The Truth about Scottish Schools*. 1st ed. London: Luath Large Print.

McGinley, B. (2018) ‘17 Local Governance in Scottish Education’, in T.G.K. Bryce et al. (eds) *Scottish Education*. Edinburgh University Press, pp. 184–193. Available at: https://doi.org/10.1515/9781474437851-018.

McIlroy, C. (2018) ‘71 The National Improvement Framework and the Return of National Testing’, in T.G.K. Bryce et al. (eds) *Scottish Education*. Edinburgh University Press, pp. 624–638. Available at: https://doi.org/10.1515/9781474437851-072.

McIntosh, J. (2015) *Final report of the Commission on Assessment without Levels*. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/483058/Commission_on_Assessment_Without_Levels_-_report.pdf.

McMahon, M. (2018) ‘92 Teacher Leadership’, *Scottish Education*. Fifth. Edinburgh: Edinburgh University Press, pp. 858–866.

Mcphee, A.D. and Patrick, F. (2009) ‘“The pupils will suffer if we don’t work”: Teacher professionalism and reactions to policy change in Scotland’, *Scottish Educational Review*, 41. Available at: https://doi.org/10.1163/27730840-04101007.

Meadows, M. *et al.* (2025) ‘England’s qualifications system: the politics of resilience’, *Educational Assessment, Evaluation and Accountability*, 37(1), pp. 127–162. Available at: https://doi.org/10.1007/s11092-025-09451-2.

Menter, I., Mutton, T. and Burn, K. (2019) ‘Learning to Teach in England: Reviewing Policy and Research Trends’, *Knowledge, Policy and Practice in Teacher Education: A Cross-National Study*. Oxford: Bloomsbury Academic, pp. 5–80. Available at: https://doi.org/10.5040/9781350068711.0017.

Messick, S. (1996) ‘Validity and washback in language testing’, *Language Testing*, 13(3), pp. 241–256.

Millard, W., Small, I. and Menzies, L. (2017) ‘Testing the Water: How assessment can underpin, not undermine, great teaching’. Available at: www.lkmco.org,.

Mills, M. (2015) ‘The tyranny of no alternative: Co-operating in a competitive marketplace’, *International Journal of Inclusive Education*, 19(11), pp. 1172–1189. Available at: https://doi.org/10.1080/13603116.2015.1044204.

Milner, A.L. and Ydesen, C. (2025) ‘International Comparisons in Education: A Spatiotemporal Analysis of Methodological–Political Trajectories, Contributions, and Debates’, *The BERA-Sage Handbook of Research-Informed Education Practice and Policy*. London: SAGE Publications, p. 385.

Ministry of Housing, Communities and Local Government (2025) *English indices of deprivation 2025*. GOV.UK). Available at: https://www.gov.uk/government/statistics/english-indices-of-deprivation-2025 (Accessed: 20 February 2026).

Mockler, N. (2011) 'Beyond "what works": Understanding teacher identity as a practical and political tool', *Teachers and Teaching: Theory and Practice*, 17(5), pp. 517–528. Available at: https://doi.org/10.1080/13540602.2011.602059.

Mooney, G. and Poole, L. (2004) '"A land of milk and honey"? Social policy in Scotland after Devolution', *Critical Social Policy*, 24(4), pp. 458–483. Available at: https://doi.org/10.1177/0261018304046672.

Morris, Z.S. (2009) 'The Truth about Interviewing Elites', *Political Studies Association*, 29(3), pp. 209–217.

Muijs, D. *et al.* (2004) 'Improving Schools in Socioeconomically Disadvantaged Areas ? A Review of Research Evidence', *School Effectiveness and School Improvement*, 15(2), pp. 149–175. Available at: https://doi.org/10.1076/sesi.15.2.149.30433.

Muijs, D. and Chapman, C. (2009) 'Accountability for improvement: rhetoric or reality?', *Radical Reforms: Perspectives on an Era of Educational Change*, pp. 28–41.

National Records of Scotland (2021) 'National records of Scotland', *Dundee City Council area profile*. Available at: https://www.nrscotland.gov.uk/files/statistics/council-area-data-sheets/dundee-city-council-profile.html.

Newton, P.E. (2007) 'Clarifying the purposes of educational assessment', *Assessment in Education: Principles, Policy and Practice*, 14(2), pp. 149–170. Available at: https://doi.org/10.1080/09695940701478321.

Newton, P.E. (2010) 'The multiple purposes of assessment', *International Encyclopedia of Education*. Elsevier, pp. 392–396. Available at: https://doi.org/10.1016/B978-0-08-044894-7.00355-9.

Newton, P.E. (2012) 'Clarifying the Consensus Definition of Validity', *Measurement: Interdisciplinary Research and Perspectives*, 10(1–2), pp. 1–29. Available at: https://doi.org/10.1080/15366367.2012.669666.

Newton, P.E. (2017) 'There Is More to Educational Measurement Than Measuring: The Importance of Embracing Purpose Pluralism', *Educational Measurement: Issues and Practice*, 36(2), pp. 5–15. Available at: https://doi.org/10.1111/emip.12146.

Oates, T. (2011) 'Could do better: Using international comparisons to refine the National Curriculum in England', *Curriculum Journal*, 22(2), pp. 121–150. Available at: https://doi.org/10.1080/09585176.2011.578908.

OECD (1999) *Measuring Student Knowledge and Skills: A New Framework for Assessment*. OECD. Available at: https://doi.org/10.1787/9789264173125-en.

OECD (2015) *Improving Schools in Scotland: An OECD Perspective*. 9788578110796. Available at: https://doi.org/10.1017/CBO9781107415324.004.

OECD (2026) *The Theory and Practice of Upper Secondary Certification*. Paris: OECD Publishing. Available at: https://doi.org/10.1787/b3fea5ba-en (Accessed: 3 February 2026).

Office for National Statistics (2024) *Population estimates for England and Wales: mid-2023*, *Office for National Statistics*. Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/populationestimatesforenglandandwales/mid2023 (Accessed: 20 February 2026).

Ofsted (2011) *The impact of the 'Assessing pupils' progress' initiative*.

O'Neill, O. (2013) 'Intelligent accountability in education', *Oxford Review of Education*, 39(1), pp. 4–16. Available at: https://doi.org/10.1080/03054985.2013.764761.

Oo, C.Z. *et al.* (2024) 'Implementing school-based assessment reforms to enhance student learning: a systematic review', *Educational Assessment, Evaluation and Accountability*, 36(1), pp. 7–30. Available at: https://doi.org/10.1007/s11092-023-09420-7.

Ozga, J. (1990) 'Policy research and policy theory: A comment on fitz and halpin', *Journal of Education Policy*, 5(4), pp. 359–362. Available at: https://doi.org/10.1080/0268093900050405.

Ozga, J. (2000) 'Education: New Labour, New Teachers', *New Managerialism, New Welfare?* London: SAGE Publications Ltd.

Ozga, J. *et al.* (2023) 'Knowledge, expertise and policy in the examinations crisis in England', *Oxford Review of Education*, 49(6), pp. 713–731. Available at: https://doi.org/10.1080/03054985.2022.2158071.

Ozga, J. and Arnott, M. (2019) 'Governing beyond PISA: knowledge, networks and narratives', *Roteiro*, 44(3), pp. 1–22. Available at: https://doi.org/10.18593/r.v44i3.21004.

Ozga, J. and Lawn, M. (2014) 'Inspectorates and Politics: the trajectories of school inspection in England and Scotland', *Revue française de pédagogie*, (186), pp. 11–21. Available at: https://doi.org/10.4000/rfp.4388.

Pajares, M.F. (1992) 'Teachers' Beliefs and Educational Research: Cleaning Up a Messy Construct', *Review of Educational Research*, 62(3), pp. 307–332. Available at: https://doi.org/10.3102/00346543062003307.

Parker, S. and Knight, E. (2025) 'Problematizing the Use of Interview Data in Research for Educational Policy and Practice: Beyond Incorrigibility and Ideology', in D. Wyse et al. (eds) *The BERA SAGE Handbook of Research Informed Education Practice and Policy*. SAGE, pp. 361–380.

Parkhurst, J.O. (2017a) *The politics of evidence: from evidence-based policy to the good governance of evidence*. Abingdon, Oxon New York, NY: Routledge (Routledge studies in governance and public policy).

Parkhurst, J.O. (2017b) *The politics of evidence: from evidence-based policy to the good governance of evidence*. London ; New York: Routledge, Taylor & Francis Group.

Passy, R. and Ovenden-Hope, T. (2020) 'Exploring school leadership in coastal schools: "getting a fair deal" for students in disadvantaged communities', *Journal of Education Policy*, 35(2), pp. 222–236. Available at: https://doi.org/10.1080/02680939.2019.1573382.

Pastore, S. and Mincu, M. (2024) 'A feasible balance? The Italian teachers' standpoint on assessment literacy, assessment practice, and teacher professional development', *European Journal of Teacher Education*, 47(2), pp. 226–245. Available at: https://doi.org/10.1080/02619768.2024.2338853.

Peebles, C. (2020) 'Exam results day: SQA awards "worse than postcode lottery" for pupils in deprived areas', *The Courier* [Preprint]. Available at: https://www.thecourier.co.uk/fp/education/1488463/exam-results-day-sqa-awards-worse-than-postcode-lottery-for-pupils-in-deprived-areas/.

Perryman, J. *et al.* (2011) 'Life in the Pressure Cooker – School League Tables and English and Mathematics Teachers' Responses to Accountability in a Results-Driven Era', *British Journal of Educational Studies*, 59(2), pp. 179–195. Available at: https://doi.org/10.1080/00071005.2011.578568.

Poet, H. *et al.* (2018) *Assessment without levels: qualitative research Research report*. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/918928/NFER_AWL_report.pdf.

Polesel, J., Rice, S. and Dulfer, N. (2014) 'The impact of high-stakes testing on curriculum and pedagogy: A teacher perspective from Australia', *Journal of Education Policy*, 29(5), pp. 640–657. Available at: https://doi.org/10.1080/02680939.2013.865082.

Popham, W.J. (1999) *Modern educational measurement: Practical Guidelines for Educational Leaders*. Pearson.

Popham, W.J. (2001) 'Teaching to the test?', *Educational Leadership*, 58(6), pp. 16–20.

Popham, W.J. (2009) 'Assessment Literacy for Teachers : Faddish or Fundamental?', *Theory Into Practice*, 48(1), pp. 4–11. Available at: https://doi.org/10.1080/00405840802577536.

Porter, N. and Simons, J. (2015) *A rising tide: the competitive benefits of free schools*. London: Policy Exhange.

Pratt, N. and Alderton, J. (2019) 'Producing assessment truths: a Foucauldian analysis of teachers' reorganisation of levels in English primary schools', *British*

Journal of Sociology of Education, 40(5), pp. 581–597. Available at: https://doi.org/10.1080/01425692.2018.1561245.

Priestley, M. (2010) 'Curriculum for Excellence: transformational change or business as usual?', *Scottish Educational Review*, 1(42), pp. 23–36.

Priestley, M. *et al.* (2015) 'The Teacher and the Curriculum: Exploring Teacher Agency', in D. Wyse, J. Pandya, and L. Hayward (eds) *The SAGE Handbook of Curriculum, Pedagogy and Assessment*. London: SAGE Publications Ltd, pp. 187–201.

Priestley, M. *et al.* (2020) 'RAPID REVIEW OF NATIONAL QUALIFICATIONS EXPERIENCE 2020'. Available at: https://www.gov.scot/publications/rapid-review-national-qualifications-experience-2020/.

Priestley, M. *et al.* (2021) 'Curriculum making: A conceptual framing', in M. Priestley (ed.) *Curriculum making in Europe: Policy and practice within and across diverse contexts*. Emerald Publishing limited, pp. 1–28.

Priestley, M. and Minty, S. (2013) 'Curriculum for Excellence: "A brilliant idea, but…"', *Scottish Educational Review*, 1(45), pp. 39–52.

Pring, R. (2015) *Philosophy of educational research*.

Prior, L. *et al.* (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). Available at: https://doi.org/10.1002/rev3.3299.

Qualifications and Curriculum Authority (2010) *Assessing pupils' progress: learners at the heart of assessment*. Coventry: Qualifications and Curriculum Development Agency. Available at: https://dera.ioe.ac.uk/id/eprint/2620.

Reckhow, S., Tompkins-Stange, M. and Galey-Horn, S. (2021) 'How the Political Economy of Knowledge Production Shapes Education Policy: The Case of Teacher Evaluation in Federal Policy Discourse', *Educational Evaluation and Policy Analysis*, 43(3), pp. 472–494. Available at: https://doi.org/10.3102/01623737211003906.

Richardson, M. (2022) *Rebuilding Public Confidence in Educational Assessment*. London: UCL Press.

Richardson, V. (1996) 'The role of attitudes and beliefs in learning to teach', *Handbook of Research on Teacher Education*, pp. 102–119.

Robson, C. and McCartan, K. (2015) *Real world research : a resource for users of social research methods in applied settings*. 4th edn. Wiley.

Sadler, D.R. (1989) 'Formative assessment and the design of instructional systems', *Instructional Science*, 18(2), pp. 119–144. Available at: https://doi.org/10.1007/BF00117714.

Sahlberg, P. and Hargreaves, A. (2015) 'The tower of PISA is badly leaning. An argument for why it should be saved.', *Washington Post*, 24 March. Available at: https://www.washingtonpost.com/news/answer-sheet/wp/2015/03/24/the-

tower-of-pisa-is-badly-leaning-an-argument-for-why-it-should-be-saved/ (Accessed: 10 October 2024).

Sammons, P. (2008) 'Zero tolerance of failure and New Labour approaches to school improvement in England', *Oxford Review of Education*, 34(6), pp. 651–664. Available at: https://doi.org/10.1080/03054980802518847.

Santori, D. (2020) 'Test-Based Accountability in England', in D. Santori, *Oxford Research Encyclopedia of Education*. Oxford University Press. Available at: https://doi.org/10.1093/acrefore/9780190264093.013.1454.

Schleicher, A. (2009) 'International Comparisons of Student Learning Outcomes', in A. Hargreaves et al. (eds) *Second International Handbook of Educational Change*. Dordrecht: Springer Netherlands, pp. 485–504. Available at: https://doi.org/10.1007/978-90-481-2660-6_29.

Schmidt, R., Jacobs, M.M. and Meyer, H. (2017) 'Sociopolitical testing discourses in elementary teachers' talk about reading assessment', *English Teaching: Practice & Critique*, 16(3), pp. 391–406. Available at: https://doi.org/10.1108/ETPC-05-2017-0066.

Scotland, J. (2012) 'Exploring the Philosophical Underpinnings of Research: Relating Ontology and Epistemology to the Methodology and Methods of the Scientific, Interpretive, and Critical Research Paradigms', *English Language Teaching*, 5(9), p. p9. Available at: https://doi.org/10.5539/elt.v5n9p9.

Scottish Executive (2004) *Ambitious, Excellent Schools: Our Agenda for Action*. Scottish Executive, p. 21.

Scottish Executive (2006) *A curriculum for excellence : The Curriculum Review Group*. Edinburgh: Scottish Executive.

Scottish Government (2010) *Curriculum for excellence : building the curriculum 5 : a framework for assessment*. Scottish Government, p. 64.

Scottish Government (2015) *Creating a smarter Scotland a draft national improvement framework for Scottish education*. Edinburgh: Scottish Government: Edinburgh: Scottish Government.

Scottish Government (2016a) *Delivering Excellence and Equity in Scottish Education: A Delivery Plan for Scotland*. Edinburgh: Scottish Government.

Scottish Government (2016b) *National improvement framework for Scottish education: 2016 evidence report*. Edinburgh: The Scottish Government.

Scottish Government (2017) *Education Governance: Next Steps: Empowering Our Teachers, Parents and Communities to Deliver Excellence and Equity for Our Children.* Edinburgh.

Scottish Government (2020) *Scottish Index of Multiple Deprivation 2020*. Scottish Government. Available at: https://www.gov.scot/collections/scottish-index-of-multiple-deprivation-2020/ (Accessed: 20 February 2026).

Scottish Government (2023) *It's Our Future - Independent Review of Qualifications and Assessment: report*. Edinburgh: Scottish Government.

Available at: https://www.gov.scot/publications/future-report-independent-review-qualifications-assessment/ (Accessed: 2 February 2026).

Scottish Government (2025) *Education (Scotland) Act 2025*, *asp*. Available at: https://www.legislation.gov.uk/asp/2025/11 (Accessed: 2 February 2026).

Shapira, M. and Priestley, M. (2018) *Narrowing the Curriculum? Contemporary trends in provision and attainment in the Scottish Curriculum*, pp. 75–107.

Shaw, M. (2004) 'Why Exam Reform is Desperately Needed', *The Times Educational Supplement* [Preprint].

Shepherd, J. (2010) 'Conservatives aim to lure private schools into state system', *The Guardian* [Preprint].

Shulman, L.S. (1986) 'Those who understand: Knowledge growth in teaching', *Educational Researcher*, 15(2), pp. 4–14. Available at: https://doi.org/10.30827/profesorado.v23i3.11230.

Sibieta, L. and Jerrim, J. (2021) 'A comparison of school institutions and policies across the UK'. Education Policy Institute. Available at: https://epi.org.uk/wp-content/uploads/2021/04/EPI-UK-Institutions-Comparisons-2021.pdf.

Sikes, P. (2006) 'On dodgy ground? Problematics and ethics in educational research', *International Journal of Research & Method in Education*, 29(1), pp. 105–117. Available at: https://doi.org/10.1080/01406720500537502.

Silverman, D. (2022) *Doing qualitative research*. Sixth edition. Los Angeles London New Delhi Singapore Washington DC Melbourne: Sage.

Simpson, M. and Hayward, L. (1998) 'Policy, research and classroom-based development: Changing the assessment culture in Scottish schools', *European Journal of Education*, 33(4), pp. 445–445.

Smith, K. (2010) 'Research, policy and funding - academic treadmills and the squeeze on intellectual spaces', *British Journal of Sociology*, 61(1), pp. 176–195. Available at: https://doi.org/10.1111/j.1468-4446.2009.01307.x.

Smith, K. and Haux, T. (2017) 'Evidence-based policy-making (EBPM)', *Handbook of Social Policy Evaluation*. Edward Elgar Publishing Ltd., pp. 141–159. Available at: https://doi.org/10.4337/9781785363245.

Smith, K.E. (2014) 'The politics of ideas: The complex interplay of health inequalities research and policy', *Science and Public Policy*, 41(5), pp. 561–574. Available at: https://doi.org/10.1093/scipol/sct085.

Smith, K.E. and Katikireddi, S.V. (2013) 'A glossary of theories for understanding: Policymaking', *Journal of Epidemiology and Community Health*, 67(2), pp. 198–202. Available at: https://doi.org/10.1136/jech-2012-200990.

Spielman, A. (2017) *HMCI's Commentary: Recent Primary and Secondary CurriculumResearch*. Available at: https://www.gov.uk/government/speeches/hmcis-commentary-october-2017 (Accessed: 1 January 2023).

Spencer, E. and Hayward, L. (2016) ‘More Than Good Intentions: Policy and Assessment for Learning in Scotland’, in D. Laveault and L. Allal (eds) *Assessment for Learning: Meeting the Challenge of Implementation*. Cham: Springer International Publishing (The Enabling Power of Assessment), pp. 111–127. Available at: https://doi.org/10.1007/978-3-319-39211-0_7.

Spillane, J.P. (2004) *Standards deviation: how schools misunderstand education policy*. Cambridge (Mass.): Harvard university press.

Staufenberg, J. (2018) ‘Scrapping “levels” led to confusion between schools, warns new research’, *Schools Week*. Available at: https://schoolsweek.co.uk/scrapping-levels-led-to-confusion-between-schools-warns-new-research/.

Stewart, W. (2009) ‘Expert group steers clear of Sats reform’, *The Times Educational Supplement* [Preprint].

Stiggins, R. (1991) ‘Assessment Literacy’, *The Phi Delta Kappan*, 72(7), pp. 534–539.

Stobart, G. (2008) *Testing times The uses and abuses of assessment*. 1st edn. Routledge.

Stobart, G. (2021) ‘Upper-secondary education student assessment in Scotland: A comparative perspective’. Education working papers 253. OECD. Available at: https://doi.org/10.1787/d8785ddf-en.

Stobart, G. and Eggen, T. (2012) ‘High-stakes testing – value, fairness and consequences’, *Assessment in Education: Principles, Policy & Practice*, 19(1), pp. 1–6. Available at: https://doi.org/10.1080/0969594X.2012.639191.

Stronach, I. and Morris, B. (1994) ‘Polemical notes on educational evaluation in the age of ‘policy hysteria‘’, *Evaluation & Research in Education*, 8(1–2), pp. 5–19. Available at: https://doi.org/10.1080/09500799409533351.

Sturgeon, N. (2025) *Frankly*. London: Macmillan.

Sturrock, S. (2024) ‘‘“Gaming” in the English primary school: “do whatever you need to do to make your data look good”’, *Journal of Education Policy*, pp. 1–23. Available at: https://doi.org/10.1080/02680939.2024.2360993.

Taylor, M. (2023) ‘Power, Participants and Pandemics – steering a reciprocal path through research methods’, *Hillary Place Papers* [Preprint], (8). Available at: https://doi.org/10.48785/100/215.

Teacher Development Trust (2014) *Developing Great Teaching: Lessons from the international reviews into effective professional development*. London. Available at: https://chartered.college/wp-content/uploads/2026/03/DGT-Full-report.pdf (Accessed: 26 May 2026).

Teelken, C. (2000) ‘Market Forces in Education A Comparative Perspective in England and Scotland’, *Scottish Educational Review*, 32(1), pp. 21–32.

The Expert Group on Assessment (2009) *Report of the Expert Group on Assessment*.

The Scottish Parliament (2019) *20 years of scrutinising school education policy – a learning experience.*, *SPICe Spotlight*. Available at: https://spice-spotlight.scot/2019/12/05/20-years-of-scrutinising-school-education-policy-a-learning-experience/ (Accessed: 12 September 2025).

Thompson, G., Lingard, B. and Ball, S.J. (2020) '"Indentured autonomy": headteachers and academisation policy in Northern England', *Journal of Educational Administration and History*, 0(0), pp. 1–18. Available at: https://doi.org/10.1080/00220620.2020.1850433.

Thomson, P. (2002) *Schooling the Rustbelt Kids: Making the Difference in Changing Times*. Crows nest: Allen and Unwin.

Tomlinson, M. (2004) '14-19 Curriculum and Qualifications Reform Final Report of the Working Group on 14-19 Reform'.

Tomlinson, S. (2005) 'Race, ethnicity and education under New Labour', *Oxford Review of Education*, 31(1), pp. 153–171. Available at: https://doi.org/10.1080/0305498042000337246.

Trowler, P. (2011) *Researching your own institution*. British Educational Research Association on-line resource. Available at: https://www.bera.ac.uk/publication/researching-your-own-institution-higher-education (Accessed: 20 October 2025).

Troyna, B. (1994) *Critical Social Research and Education Policy*, pp. 70–84.

Walford, G. (2005) 'Introduction: education and the Labour Government', *Oxford Review of Education*, 31(1), pp. 3–9. Available at: https://doi.org/10.1080/0305498042000337156.

Wallace, C.S. and Priestley, M. (2011) 'Teacher beliefs and the mediation of curriculum innovation in Scotland: A socio-cultural perspective on professional development and change', *Journal of Curriculum Studies*, 43(3), pp. 357–381. Available at: https://doi.org/10.1080/00220272.2011.563447.

Walshaw, M. (2007) 'Getting to grips with Foucault', *Working with Foucault in Education*. Sense Publishers, pp. 1–24.

Warrell, H. (2016) 'Sturgeon: "No excuses" for Scots school performance', *BBC News*. Available at: https://www.bbc.co.uk/news/uk-scotland-scotland-politics-38250199 (Accessed: 7 March 2026).

Webb, R. and Vulliamy, G. (2007) 'Changing classroom practice at Key Stage 2: the impact of New Labour's national strategies', *Oxford Review of Education*, 33(5), pp. 561–580. Available at: https://doi.org/10.1080/03054980701476204.

Wells, P. (2007) 'New Labour and evidence based policy making: 1997-2007', *People, Place and Policy Online*, pp. 22–29. Available at: https://doi.org/10.3351/ppp.0001.0001.0004.

West, A. (2010) 'High stakes testing, accountability, incentives and consequences in English schools', *Policy Press*, 38(1), pp. 23–39. Available at: https://doi.org/https://doi.org/10.1332/030557309X445591.

West, A. and Pennell, H. (2000) 'Publishing School Examination Results in England: Incentives and consequences', *Educational Studies*, 26(4), pp. 423–436. Available at: https://doi.org/10.1080/03055690020003629.

Weston, D. (2018) 'A postcode lottery for teachers', *Teacher Development Trust*, 18 January. Available at: https://tdtrust.org/2018/01/18/post-code-lottery-teachers/ (Accessed: 10 September 2024).

Wiggins, A. and Tymms, P. (2002) 'Dysfunctional Effects of League Tables: A Comparison Between English and Scottish Primary Schools', *Public Money and Management*, 22(1), pp. 43–48. Available at: https://doi.org/10.1111/1467-9302.00295.

Wiliam, D. *et al.* (2004) 'Teachers developing assessment for learning: impact on student achievement', *Assessment in Education: Principles, Policy & Practice*, 11(1), pp. 49–65. Available at: https://doi.org/10.1080/0969594042000208994.

Wiliam, D. (2011) *Embedded formative assessment*. Bloomington: Solution Tree Press.

Wiliam, D. and Thompson, M. (2007) 'Integrating assessment with instruction: what will it take to make it work?', in C.A. Dwyer (ed.) *The future of Assessment: Shaping Teaching and Learning*. New York: Routledge, pp. 53–84.

Wrigley, T. (2006) *Another School is Possible*. London: Bookmarks publications.

Wrigley, T., Lingard, B. and Thomson, P. (2012) 'Pedagogies of transformation: keeping hope alive in troubled times', *Critical Studies in Education*, 53(1), pp. 95–108. Available at: https://doi.org/10.1080/17508487.2011.637570.

Xu, Y. and Brown, G.T.L. (2016) 'Teacher assessment literacy in practice: A reconceptualization', *Teaching and Teacher Education*, 58, pp. 149–162. Available at: https://doi.org/10.1016/j.tate.2016.05.010.

Yanow, D. (2011) 'Accessing Local Knowledge: Identifying Interpretive Communities and Policy Artifacts', *Conducting Interpretive Policy Analysis*. SAGE Publications, Inc., pp. 27–41. Available at: https://doi.org/10.4135/9781412983747.

Yin, R.K. (2013) 'Validity and generalization in future case study evaluations', *Evaluation*, 19(3), pp. 321–332. Available at: https://doi.org/10.1177/1356389013497081.

Yin, R.K. (2018) *Case study research and applications: design and methods*. Sixth edition. London: SAGE.