

**The implications of the Government's "assessment without levels"  
reform for teachers in primary schools.**

By

**Victoria Anne Birmingham**

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of  
Philosophy in Education at Birmingham City University

August 2021



## ACKNOWLEDGMENTS

---

If you would have asked me years ago if I thought I could study and complete a PhD I would have laughed and said no. However, there are people in my life who would not have given it a second thought and told me to go for it and that they believed I could. Right from the nervous excited phone call to mum when I found the PhD advertised, I was met with that belief. I would like to take this opportunity to acknowledge the unwavering faith and belief my parents, Susan and Patrick Birmingham, have always had in me. Thank you to my sisters, Heather and Beth and their ability to still smile and nod at my reply every time they ask how my PhD is going. I would like to thank my closest friends, Dr Liz Aston for her reassurance and proof reading and Jennifer Taylor for her constant supportive presence in my life since our first day of secondary school. To my husband Christian, thank you for supporting me, putting my studies first and moving countries so we could be together.

I could not have gotten this far without the feedback and support of my supervisors Martin Fautley and Matt O’Leary. Martin, from day 1 you gave me the agency to shape my research and learn as I go without having a fear of failure. Matt, your insight into my research has been so valuable and helped me to shape it into what it has become. You have both encouraged me to think way beyond my ‘teacher’ box and opened many new perspectives to view Education and life through. A big thank you to Dr Carolyn Blackburn for being my critical reader and providing the feedback I needed to refine my final draft.

The HELS PGR community, when I have needed it, you have been a support, an ear, a critical friend and a coffee break. I am not the most at ease in social situations but I have always felt like I belonged among you. And finally to the DRC, and the research community in the School of Education and Social Work, you have all made it possible for me complete this thesis and made me feel welcome as a PhD student and now as a research assistant. It has been a wonderful and unique experience and I really have enjoyed every minute. I had a feeling 5 years ago that I knew I would regret it if I did not take the chance. You have all proved me right.

## ABSTRACT

---

Levels formed the basis of primary school assessment since the introduction of the National Curriculum in 1988. After nearly 30 years in use in 2011, under 2010 Coalition Government reform, levels were removed from non-statutory assessment. As such, Assessment without Levels (AwLs) is an emerging field of research with little known about the impact it has had on teaching and assessment practices in primary schools to date. It is on that basis that this research emerged.

Mixed methods were used to investigate the impact of AwLs on primary teaching and assessment practices. An online survey canvassed the views of teachers across Key Stage 1 and Key stage 2 and interviews were carried out with year 2 and year 6 teachers and senior leaders in a small sample of schools. The teacher interviews aimed to complement the survey data by providing in-depth insights from teachers who taught in years with Standardised Assessment Tasks (SATs). Senior Leader Team (SLT) interviews provided a context for the school and the experiences of the classroom teachers adapting to AwLs. The study drew on three theoretical lenses through which to analyse the data. The first made use of assessment theory, particularly assessment concepts of validity and reliability. The second lens was that of neoliberalism, which was deployed as a tool for understanding the impact of this ideology on education policy. And the third lens was teacher agency, especially the ecological theory of agency from Priestley et al. (2013).

One of the underpinning aims for AwLs reform was to confront the longstanding issue of teaching to the test, my findings reveal that not only has it failed to achieve this but schools in this study relied more heavily on the limited assessment guidance available and externally purchased resources aligned towards SATs framing and content. Because of the increased focus on assessed content, findings suggest the validity of SATs must be questioned as representing the learning they purport to assess. The continuation of test-based accountability to hold primary schools to account while scaling back assessment guidance has led to schools becoming overly reliant on the guidance they do possess. This undermines the reform's goal to increase autonomy to schools for their own assessment and has opened up the market for assessment resources aligned to SATs rather than curriculum learning.

# TABLE OF CONTENTS

---

<b>ACKNOWLEDGMENTS</b> .....	<b>III</b>
<b>ABSTRACT</b> .....	<b>IV</b>
<b>LIST OF TABLES</b> .....	<b>XIII</b>
<b>LIST OF FIGURES</b> .....	<b>XIV</b>
<b>GLOSSARY OF TERMS</b> .....	<b>XVI</b>
<b>LIST OF ABBREVIATIONS</b> .....	<b>XVIII</b>
<b>1 INTRODUCTION TO THE RESEARCH</b> .....	<b>1</b>
1.1 ORIGINS OF INTEREST IN MY RESEARCH TOPIC.....	1
1.2 RESEARCH AIMS.....	2
1.3 CONTEXT AND RATIONALE.....	3
1.3.1 <i>A brief history of Assessment with Levels</i> .....	3
1.3.2 <i>Rationale for AwLs research</i> .....	6
1.4 STRUCTURE OF THESIS.....	7
<b>2 LITERATURE REVIEW</b> .....	<b>9</b>
2.1 INTRODUCTION.....	9
2.2 UNDERSTANDING ASSESSMENT IN AN EDUCATIONAL CONTEXT.....	10
2.2.1 <i>Assessment definition</i> .....	10
2.2.2 <i>A contextualised example of educational assessment</i> .....	11
2.2.3 <i>Objective considerations of assessment uses and purposes</i> .....	13
2.2.4 <i>Introduction of terms validity and reliability of assessment</i> .....	13
2.2.5 <i>In-situ assessment types</i> .....	14
2.3 THE PLACE OF ASSESSMENT IN THEORIES OF LEARNING?.....	17
2.3.1 <i>Educational learning theories- an overview</i> .....	17
2.3.2 <i>The relationship between learning theory and assessment</i> .....	19
2.3.2.1 Persistent influence of behavioural learning theory on standardised testing.....	20
2.3.2.2 Tensions between learning theory developments and educational assessment.....	21
2.4 VIEWS OF ASSESSMENT TYPES.....	23
2.4.1 <i>Formative assessment</i> .....	23
2.4.2 <i>Critique of formative assessment</i> .....	25
2.4.3 <i>Summative assessment</i> .....	25
2.4.3.1 Relationship between summative and formative assessment.....	26
2.4.3.2 Assessment referencing.....	27

2.5	VALIDITY AND RELIABILITY .....	28
2.5.1	<i>Assessment validity</i> .....	29
2.5.2	<i>Types of validity</i> .....	30
2.5.2.1	Ethical considerations of validity .....	31
2.5.3	<i>Reliability</i> .....	33
2.5.4	<i>Error and reliability</i> .....	33
2.5.5	<i>Interaction between validity and reliability</i> .....	34
2.5.6	<i>Judging assessments beyond validity and reliability</i> .....	34
2.6	CONSIDERATIONS OF VALIDITY AND RELIABILITY OF PRIMARY SCHOOL ASSESSMENT .....	35
2.6.1	<i>Validity and reliability of tests as summative assessments</i> .....	35
2.6.1.1	High-stakes summative assessment- construct validity impact.....	36
2.6.1.1.1	Impact on FA.....	39
2.6.1.2	High-stakes assessment and accountability.....	40
2.6.2	<i>Validity and reliability of teacher SA</i> .....	41
2.6.2.1	Trust in teacher SA.....	41
2.6.2.2	Teacher SA bias .....	42
2.6.2.3	Benefits of teacher SA.....	44
2.6.2.4	Use of comparative judgment.....	45
2.6.2.4.1	Focus on consistency and inter-rater reliability at the cost of validity .....	46
2.6.2.4.2	CJ to supplement rather than replace traditional SA.....	48
2.7	RESEARCH INTO ASSESSMENT WITHOUT LEVELS IN PRIMARY SCHOOLS .....	49
2.7.1	<i>AwLs and high-stakes testing</i> .....	50
2.8	CONCLUSION .....	52
<b>3</b>	<b>EDUCATION POLICY .....</b>	<b>54</b>
3.1	INTRODUCTION .....	54
3.2	POLICY AS AN IMPACT ON EDUCATION .....	54
3.3	OVERVIEW OF EDUCATION POLICY BETWEEN 1988-2010 .....	55
3.3.1	<i>Success/failure of New Labour reforms</i> .....	56
3.4	CHANGE OF GOVERNMENT, CHANGE OF POLICY...?.....	57
3.4.1	<i>Justification for Policy change</i> .....	58
3.4.2	<i>Curriculum and assessment commissioned reviews</i> .....	59
3.4.2.1	The Bew review.....	59
3.4.3	<i>The National Curriculum review (2011) expert panel</i> .....	60
3.4.4	<i>Review recommendations taken on or rejected</i> .....	62
3.4.4.1	Same evidence, different interpretations .....	62
3.5	CONCLUSION .....	63
<b>4</b>	<b>NEOLIBERALISM AND EDUCATION POLICY.....</b>	<b>64</b>
4.1	A BRIEF HISTORY OF NEOLIBERAL EDUCATION REFORM .....	65

4.2	MARKETISATION OF EDUCATION, EDUCATION AS A COMMODITY .....	66
4.2.1	<i>Choice</i> .....	67
4.2.2	<i>Education viewed as an economic product</i> .....	67
4.2.3	<i>Implications of marketised education</i> .....	68
4.3	COMPETITION.....	68
4.4	ACCOUNTABILITY AND AUTONOMY WITHIN THE NEOLIBERAL MESSAGE .....	69
4.5	MANAGERIALISM .....	71
4.5.1	<i>Performativity</i> .....	72
4.5.2	<i>Effect of surveillance</i> .....	72
4.5.3	<i>Conflict and performative behaviours</i> .....	73
4.6	EFFECTS OF NEOLIBERAL REFORMS .....	74
4.7	CONCLUSION .....	75
<b>5</b>	<b>TEACHER AGENCY AND IDENTITY.....</b>	<b>76</b>
5.1	THEORISATION OF AGENCY .....	76
5.2	TEACHER IDENTITY.....	78
5.3	IDENTITY AND AGENCY.....	79
5.3.1	<i>Inclusion of teacher identity theory into the ecological model</i> .....	79
5.4	TYPES OF AGENCY.....	81
5.5	INFLUENCES ON AGENCY AND IDENTITY .....	83
5.5.1	<i>Discourses</i> .....	83
5.5.1.1	Accountability .....	84
5.5.1.1.1	Performativity.....	86
5.5.1.2	Professionalism .....	87
5.5.2	<i>School values and ideologies</i> .....	89
5.5.2.1	School mediation and translation .....	90
5.5.3	<i>Age and teaching experience</i> .....	90
5.5.4	<i>Is agency desired?</i> .....	92
5.6	CONCLUSION .....	95
<b>6</b>	<b>METHODOLOGY AND METHODS.....</b>	<b>96</b>
6.1	INTRODUCTION .....	96
6.2	RESEARCH QUESTIONS.....	97
6.3	MULTIPLE RESEARCH PARADIGMS .....	97
6.3.1	<i>Interpretivist ontology</i> .....	98
6.3.2	<i>Social constructivism</i> .....	99
6.3.3	<i>Radical constructivism</i> .....	99
6.3.4	<i>Hermeneutics</i> .....	100
6.4	RESEARCH DESIGN .....	101

6.4.1	<i>Pragmatic research design</i> .....	101
6.4.2	<i>Mixed methods design</i> .....	102
6.4.3	<i>Overview of data collected</i> .....	103
6.4.4	<i>Teacher interviews</i> .....	104
6.4.4.1	Interviewees and rationale .....	104
6.4.4.2	Interview format .....	105
6.4.4.3	Interviews: Ethical considerations .....	106
6.4.4.3.1	Equality, diversity in participant recruitment .....	106
6.4.4.3.2	Informed consent .....	107
6.4.4.4	Minimise harm .....	107
6.4.5	<i>Online survey</i> .....	109
6.4.5.1	Data collection .....	111
6.4.5.2	Survey change of purpose.....	112
6.4.5.3	Survey: Ethical considerations .....	112
6.5	RESEARCH SETTINGS AND PARTICIPANTS.....	113
6.5.1	<i>Survey participants</i> .....	113
6.5.2	<i>Interview participant demographics Setting 1</i> .....	115
6.5.2.1	Teacher D.....	116
6.5.2.2	Teacher E .....	116
6.5.2.3	Teacher F.....	116
6.5.2.4	Head teacher and Deputy Head teacher Setting 1.....	116
6.5.3	<i>Interview participant demographics Setting 2</i> .....	116
6.5.3.1	Teacher G.....	116
6.5.3.2	Teacher H.....	117
6.5.3.3	Assistant Head teacher and assessment lead Setting 2 .....	117
6.6	ANALYSIS PROCEDURE.....	118
6.6.1	<i>Qualitative Analysis procedures and theories</i> .....	118
6.6.1.1	Pilot study- reflections and adaptations .....	118
6.6.1.2	Main Study: Qualitative analysis procedures and theories.....	119
6.6.1.2.1	Thematic Analysis .....	119
6.6.2	<i>Quantitative analysis procedures</i> .....	122
6.6.2.1	What was the difference between responses of participants with experience of NCLs and those without?	125
6.6.2.2	What was the difference between responses of participants who taught in KS1 compared to KS2?	125
6.6.2.3	What was the difference between responses of participants teaching in a year with SATs assessments and those who did not? .....	126
6.6.2.4	Identifying interesting differences in the binary independent variable responses.....	126
6.6.2.4.1	Analysis process.....	127
6.7	VALIDITY, RELIABILITY, CREDIBILITY, AND TRUSTWORTHINESS.....	130
6.7.1	<i>Awareness of one's own bias</i> .....	131

6.7.2	<i>Triangulation</i> .....	133
6.8	CONCLUSION.....	134
<b>7</b>	<b>SURVEY DATA FINDINGS</b> .....	<b>135</b>
7.1.1	<i>Abbreviations and coding of participant demographics</i> .....	135
7.1.2	<i>Question 7- All survey responses to questions relating to teachers’ summative assessment of Maths and Reading</i> .....	135
7.1.2.1	Question 7- text responses.....	137
7.1.3	<i>Question 8- All survey responses to questions relating to External Summative Testing of reading and maths</i> .....	138
7.1.3.1	Question 8- text responses.....	146
7.1.4	<i>Question 9. All survey responses to questions relating to summative assessment of non-core subjects, for example: history, geography, RE, PE, art</i> .....	147
7.1.4.1	Question 9- text responses.....	153
7.1.5	<i>Question 11- any other comments</i> .....	154
7.1.6	<i>Further analysis via demographic information</i> .....	154
7.1.6.1	Further analysis of responses to Q7 according to binary IV.....	156
7.1.6.2	Further analysis of responses to Q8 according to binary IV.....	160
7.1.6.3	Further analysis of responses to Q9 according to binary IV.....	164
7.1.7	<i>Survey data discussion and conclusion</i> .....	166
7.1.7.1	Question 7- Summary of survey responses to questions relating to teacher’s summative assessment of Maths and Reading.....	166
7.1.7.2	Question 8 Summary of survey responses to questions relating to External Summative Testing of reading and maths.....	166
7.1.7.3	Question 9 Summary of survey responses to questions relating to summative assessment of non-core subjects, for example: history, geography, RE, PE, art.....	167
7.1.7.4	What was the difference between responses of participants with experience of NCLs and those without?	167
7.1.7.5	What was the difference between responses of participants who taught in KS1 compared to KS2?	168
7.1.7.6	What was the difference between responses of participants teaching in a year with SATs assessments and those who did not?	168
<b>8</b>	<b>INTERVIEW DATA FINDINGS- MAIN STUDY SETTINGS 1 AND 2</b> .....	<b>170</b>
8.1	SETTING 1- TEACHER INTERVIEW DATA THEMATIC ANALYSIS.....	170
8.1.1	<i>Transition difficulties and adjustments</i> .....	170
8.1.1.1	New policy, same practices.....	174
8.1.1.2	Still uncertainty of change.....	176
8.1.2	<i>Different assessment uses and interpretations</i> .....	176
8.1.3	<i>Questions of validity</i> .....	177
8.1.3.1	Different assessment methods for different subject areas.....	179

8.1.3.2	Secure fit assessment of writing using interim assessment frameworks.....	181
8.1.4	<i>School expectation that teacher SA and test match</i> .....	181
8.1.5	<i>Trust in teacher assessment, the need for proof</i> .....	182
8.1.6	<i>Pupil mental health</i> .....	185
8.1.7	<i>High-stakes influences of SATs</i> .....	185
8.1.7.1	Wash-back .....	186
8.1.8	<i>Summary</i> .....	190
8.2	SETTING 1- SENIOR LEADERSHIP TEAM INTERVIEW THEMATIC ANALYSIS .....	191
8.2.1	<i>Transition and adaptation</i> .....	191
8.2.2	<i>Difficulties with tracking</i> .....	193
8.2.3	<i>School's responsibility to meet raised curriculum standards</i> .....	194
8.2.4	<i>Test and independence valued</i> .....	195
8.2.5	<i>Labelling or not labelling children</i> .....	197
8.2.6	<i>Improvements without levels- student dialogue and parental engagement</i> .....	198
8.2.7	<i>Summary</i> .....	199
8.2.8	<i>Interview Themes Triangulation Setting 1</i> .....	200
8.2.8.1	Overlapping themes.....	200
8.2.8.2	Conflicting themes .....	200
8.2.8.3	SLT impact on teacher themes.....	201
8.3	SETTING 2- TEACHER INTERVIEWS DATA THEMATIC ANALYSIS .....	202
8.3.1	<i>Adjusting to assessment without levels</i> .....	203
8.3.1.1	Late and/or insufficient guidance so seeking support elsewhere .....	205
8.3.2	<i>Questions of Validity</i> .....	208
8.3.3	<i>Evidence- needing to prove assessment judgments</i> .....	211
8.3.4	<i>Tracking</i> .....	213
8.3.5	<i>Accountability</i> .....	214
8.3.5.1	High-stakes influences of SATs- Teaching children to educate them or to pass a test? .....	215
8.3.6	<i>Summary</i> .....	215
8.4	SETTING 2- SLT INTERVIEW DATA THEMATIC ANALYSIS .....	216
8.4.1	<i>Transition and adaptation</i> .....	216
8.4.2	<i>Tests objectively measure progress</i> .....	219
8.4.3	<i>Labelling or not labelling children</i> .....	219
8.4.4	<i>Parental engagement</i> .....	220
8.4.5	<i>Summary of SLT S2 interview themes</i> .....	221
8.5	S2 TEACHER AND SLT INTERVIEW THEMES TRIANGULATION. ....	221
8.5.1	<i>Overlapping themes</i> .....	221
8.5.2	<i>Conflicting Themes</i> .....	221
8.5.3	<i>SLT impact on teacher Themes</i> .....	222
8.6	CONCLUSION .....	222

<b>9</b>	<b>TEACHER AGENCY ANALYSIS AND DISCUSSION.....</b>	<b>224</b>
9.1	INTRODUCTION.....	224
9.2	CATEGORIES OF AGENCY IDENTIFIED.....	224
9.2.1	<i>Conflicted compliance</i> .....	225
9.2.1.1	Teacher F.....	226
9.2.1.2	Teacher E.....	229
9.2.1.3	Teacher D.....	232
9.2.1.4	Teacher G.....	236
9.2.1.4.1	Espoused theory of action inconsistencies with theories in use.....	236
9.2.2	<i>Non-conflicted compliance</i> .....	240
9.2.2.1	Teacher D.....	240
9.2.2.2	Passive compliance.....	241
9.2.3	<i>Agency</i> .....	244
9.2.3.1	Agentic resistance.....	244
9.2.3.2	Strategic compliance.....	247
9.2.3.2.1	Teacher D.....	248
9.2.3.2.2	Teacher H.....	248
9.3	NEOLIBERAL DISCOURSES AND TEACHER AGENCY.....	253
9.3.1	<i>Conflicted compliance</i> .....	253
9.3.2	<i>Non-conflicted and passive compliance</i> .....	255
9.3.3	<i>Agency</i> .....	257
9.3.4	<i>Conclusion</i> .....	259
<b>10</b>	<b>FURTHER ANALYSIS AND DISCUSSION.....</b>	<b>261</b>
10.1	SATS PRESSURES AND THE IMPACT OF HIGH-STAKES TESTING WITHOUT LEVELS.....	261
10.2	TIMING AND EXTENT OF GOVERNMENT GUIDANCE ON ASSESSMENT FOR ALL YEARS.....	263
10.3	WRITING MODERATION AND RELIANCE ON ITAF.....	265
10.4	VALIDITY OF SATS.....	267
10.5	TEACHER ASSESSMENT: TRUST AND RELIABILITY.....	269
10.6	NEOLIBERAL IMPLICATIONS OF AWL POLICY REFORM.....	271
10.6.1	<i>Marketisation of education, education as a commodity</i> .....	271
10.6.1.1	Competition.....	272
10.6.2	<i>Accountability and autonomy within the neoliberal message</i> .....	273
10.6.3	<i>Managerialism</i> .....	273
10.6.3.1	Performativity.....	273
<b>11</b>	<b>CONCLUSION.....</b>	<b>275</b>
11.1	INTRODUCTION.....	275
11.2	WHAT HAS BEEN THE IMPACT OF THE ASSESSMENT REFORM 'ASSESSMENT WITHOUT LEVELS' ON TEACHING AND ASSESSMENT PRACTICE?.....	275

11.3	WHAT ARE THE POLICY INTO PRACTICE IMPLICATIONS OF THE ASSESSMENT REFORM ‘ASSESSMENT WITHOUT LEVELS’?	278
11.4	CONTRIBUTION TO KNOWLEDGE .....	279
11.4.1	<i>Adding to and moving forward debates around assessment practices and policy in primary education.</i>	280
11.4.2	<i>Agency theory and the Neoliberal impact</i> .....	281
11.5	RESEARCH DESIGN REFLECTIONS .....	283
11.6	RECOMMENDATIONS.....	284
11.6.1	<i>Further research</i> .....	284
11.6.2	<i>Schools, teachers, and policy makers</i> .....	284
<b>REFERENCES.....</b>		<b>287</b>
<b>APPENDICES.....</b>		<b>306</b>
APPENDIX 1 RELEASE OF ASSESSMENT GUIDANCE .....		306
APPENDIX 2 APPROVAL OF ETHICAL APPLICATION RELEASE OF ASSESSMENT GUIDANCE .....		308
APPENDIX 3 ONLINE SURVEY.....		309
APPENDIX 4 PILOT STUDY INTERVIEW SCHEDULE .....		317
APPENDIX 5 INFORMATION SHEETS AND CONSENT FORMS .....		318
APPENDIX 6 PILOT STUDY DATA ANALYSIS THEORY AND PROCEDURE.....		324
APPENDIX 7 FULL STUDY REVISED INTERVIEW SCHEDULE .....		338
APPENDIX 8 INTERVIEW SETTING SCHOOL DEMOGRAPHICS AND SATS PERFORMANCE.....		340
APPENDIX 9 INTERVIEW TRANSCRIPT.....		345
APPENDIX 10 FULL STUDY CODING PROCESS NOTES .....		348
APPENDIX 11 MY PHD JOURNEY .....		350

## LIST OF TABLES

---

Table 1-1 Standardised assessment tasks (SATs) for assessment with and without levels in Primary schools.....	5
Table 1-2 Assessment criteria for end of Key Stage 1 and Key Stage 2 NCL and AwLs. ....	6
Table 2-1 Example of Child A's and B's attempts at two assessment tasks for counting 1-5. 12	
Table 6-1 Diagram to show an overview of the methods and where the participants were recruited from.....	103
Table 6-2: Duration of each interview .....	105
Table 6-3 Survey participant demographics. ....	113
Table 6-4 Reflections and adaptations following the Pilot Study.....	118
Table 6-5 Conversion of agreement responses to frequency of values .....	128
Table 7-1 Cross tabulation of survey responses to Q8.1 and Q8.3. ....	141
Table 7-2 Crosstabulation of survey responses to Q8.5 and Q8.6. ....	143
Table 7-3 Crosstabulation of survey responses to Q8.5 and Q8.1. ....	145
Table 7-4 The crosstabulation of survey answers from Q9.1 and Q9.3 .....	150
Table 7-5 Crosstabulation of survey responses to Q8.5 and Q9.6. ....	152
Table 7-6 Crosstabulation of survey responses to Q9.6 and Q8.3. ....	152
Table 7-7 Results of Mann-Whitney U test on participant responses from binary IV for Q7 .....	156
Table 7-8 Results of Mann-Whitney U test on participant responses from binary IV for Q8 .....	160
Table 7-9 Results of Mann-Whitney U test on participant responses from binary IV for Q9 .....	164
Table 0-1 The initial organisation of pilot study interview themes.....	326
Table 0-2 Pupildemographic data from school research settings. ....	340
Table 0-3 Pilot study school pupil demographic information. ....	341
Table 0-4 Setting 1 school pupil demographic information .....	342
Table 0-5 Setting 2 school pupil demographic information. ....	343

## LIST OF FIGURES

---

Figure 2-1 'Three dimensions of learning and competence' Illeris (2017: 12).....	19
Figure 5-1 Ecological theory of Agency from Priestley et al. (2013: 152) .....	80
<i>Figure 5-2 Annotated version of Priestley et al. (2013) model of ecological agency</i> .....	81
Figure 6-1 Likert question structure for survey.....	110
Figure 6-2 The number of years teaching experience of survey participants. ....	114
Figure 6-3 The progress of internet link access to survey .....	115
Figure 6-4 Jisc Online surveys presentation of survey response for each statement in Q7-9 .....	123
Figure 6-5 Example of survey data presentation for Q7-9. ....	124
Figure 6-7 Flow chart showing concurrent triangulation strategy for data analysis, adapted from Creswell (2009) .....	134
Figure 9-1 Comments from Teacher F used to indicate how agency achievement was enabled and constrained using Priestley et al. (2013: 152) ecological theorisation of agency. .....	228
Figure 9-2 Comments from Teacher E used to indicate how agency achievement was enabled and constrained using Priestley et al. (2013: 152) ecological theorisation of agency. .....	231
Figure 9-3 Comments from Teacher D used to indicate how agency achievement was enabled and constrained using Priestley et al. (2013: 152) ecological theorisation of agency. .....	234
Figure 9-4 Comments from Teacher H used to indicate how agency achievement was enabled and constrained using Priestley et al. (2013: 152) ecological theorisation of agency. .....	246
Figure 9-5 Comments from Teacher D used to indicate how agency achievement was enabled and constrained using Priestley et al. (2013: 152) ecological theorisation of agency. .....	251
Figure 9-6 Summary of Neoliberal impact of conflicted compliance evident in teacher interview data.....	254

Figure 9-7 Summary of Neoliberal impact of non-conflicted and passive compliance evident in teacher interview data.....256

Figure 9-8 Summary of Neoliberal impact of agency evident in teacher interview data .....258

Figure 9-9 Adapted theory of ecological agency from neoliberal discourse impact.....259

## GLOSSARY OF TERMS

---

**Assessment**-the forming of judgments based on the collection of evidence relevant to a learning goal.

**Assessment without Levels (AwLs)**-the assessment of educational attainment in England from the ages of 5-11 in Primary schools, according to curriculum defined age related expectations at the end of KS1 and KS2.

**Criterion-referencing**-assessments are judged on the bases of a set criteria which defines achievement.

**External Moderation**- regulation of statutory teacher summative assessments by Department for Education trained moderators to ensure consistency between different schools regionally and nationally.

**Formative assessment (FA) also Assessment for learning (AFL)**-any assessment whose outcome is used to inform future learning as a means of deepening or furthering understanding.

**Key Stages**- the separation of educational attainment phases in England.

**National Curriculum Levels (NCLs)**-the assessment of educational attainment in England from the ages of 5-11 in Primary schools, according a scheme of levels from 1-6 at the end of KS1 and KS2.

**Norm-referencing**-assessments are ranked according to how individual scores compare to others who were assessed.

**Moderation**-a collective exercise in assessment where teachers use their professional judgment through shared enquiry to reach a consensus of assessment criteria understanding.

**Reliability**-the consistency of results in an assessment if replicated.

**Summative assessment (SA) also Assessment of learning**-the systematic and planned collection of evidence of learning to judge attainment within a specified time frame.

**Teacher Summative Assessment**- the systemic and planned collection of evidence by a teacher used to inform their professional judgments of learning attainment within a specified time frame.

**Teacher Assessment**-assessment which relies significantly on a teacher's professional judgment to draw inferences relevant to a learning goal.

**Test**-a selection of activities or tasks designed to produce evidence against a specified learning area or goal.

**Validity**-the accuracy in which the assessment task represents the learning it is inferred to.

## LIST OF ABBREVIATIONS

---

- ACJ- Adaptive comparative judgement
- AERA- American Educational Research Association
- AfL- Assessment for learning
- APA- American Psychological Association
- APP- Assessing pupil progress
- APTG- Assessment Policy Task Group
- ARG- Assessment Reform Group
- AWL- Assessment without levels
- BERA- British Educational Research Association
- CAWL- Commissions on assessment without levels
- CJ- Comparative judgement
- DCSF The Department for Children schools and Families
- DfE- Department for Education
- DHT- Deputy head teacher
- ERA- Education reform act
- FA- formative assessment
- HT- Head teacher
- ITAFs- Interim assessment frameworks
- IV- Independent variable
- KS1- Key stage 1, school years 1-2, ages 5-7
- KS2- Key stage 2, school year 3-6, ages 7-11
- LEA- Local Education Authority

LA- Local Authority

N/A- Non applicable

NA/ND- Neither agree nor Disagree

NCLs- National Curriculum Levels

NCR- National Curriculum Review

NFER- National Foundation for Educational Research

NNS- National numeracy strategy

NLS- National literacy strategy

NS- National strategies

OECD- Organisation for Economic Co-operation and Development

OfSTED- Office for Standards in Education, Children's Services and Skills

PIRA- Progress in Reading Assessment

PoS- Programmes of study

PUMA- Progress in Understanding Mathematics Assessment

PISA- Programme for International Student Assessment

QCA- The Qualifications and Curriculum Authority

S1- Setting 1

S2- Setting 2

SA- Summative assessment

SATs- Statutory assessment tasks

SEN- Special educational needs

SPAG- Spelling grammar and punctuation

SPTO- Student Pupil Tracker Online

TA- teacher assessment

TGAT - Task Group on Assessment and Testing

TIMMS- Trends in International Mathematics and Science Study

TTTT- teaching to the test

ZPD- Zone of proximal development

# 1 INTRODUCTION TO THE RESEARCH

---

## 1.1 ORIGINS OF INTEREST IN MY RESEARCH TOPIC

Educational assessment is the main focus of this thesis, as a Primary school teacher it felt like the main focus of my job. My performance as a teacher and my students' attainment was judged using assessment, with the latter contributing to the former. Before becoming a teacher assessment was the gate keeper to further and higher education, gaining qualifications allowed me to progress academically, or, for my career. As I progressed through education into my teaching career I had not questioned assessment's function in measuring my ability, if I did not perform well on a test I needed to work harder. Practising test questions as well as content was fundamental in my preparations. I viewed these assessments purely as a measure of my own ability. As a student, I was aware that my secondary school targeted every student with at least 5 A\*-C GCSEs, and cared a great deal about increasing results each year. As a sixth form student my College strongly encouraged everyone to apply to University, even if they had other plans. What only occurred to me when I became a teacher was that *my* results as a student were also a measure of success or failure of my teachers and school.

As a teacher I accepted this, the more effort I put in to educating students the better they would perform when assessed. During my initial years of teaching I used assessment with levels (NCLs). This had detailed tiered descriptors for Reading, Writing and Maths in the form of APP (assessing pupil progress), a catalogue of exemplars for teacher summative assessment (SA) and optional yearly tests by the QCA (Qualifications and Curriculum Authority). I had assumed along with the new Curriculum in 2014 there would be a replacement assessment framework for National Curriculum levels (NCL). When I learnt that this would not be the case I began to question the purpose of Primary school educational assessment. It frustrated me that I no longer knew what the success of my students, and therefore myself, looked like. For example, Year 5 and Year 6 shared some curriculum objectives for writing and I did not know how the new expected standard would differ between the years. My salary would now be directly related to my performance and based,

at least from my experience, on lesson observations and on student end of year assessment targets. As I looked to the senior management team for guidance I realised they too were in the dark. We were in a situation of managing a new curriculum *and* grappling with the new unknown of Assessment without Levels (AwLs).

I wondered at the time why the decision not to replace levels had been made, questioning the logic of removing an embedded assessment framework and replacing it with nothing, while still holding schools accountable with Key Stage 2 (KS2) Standardised Assessment Tasks (SATs). It felt as if we had all been given a new destination of SATs without levels, with limited directions and no way of knowing how far away we were along the journey. When I saw an opportunity to study AwLs as a PhD, I cautiously jumped at it. I wanted to find a solution to the confusion and struggles I had felt, and observed in others, associated with AwLs. I was offered the position and began a journey towards becoming a researcher. While my initial research proposal was very much influenced by a solution-focused “what works” agenda, this evolved into a desire to adopt a more nuanced exploration of the complexities associated with assessment and its impact on teacher’s thinking and their practice.

## 1.2 RESEARCH AIMS

Through the research project two key research questions were formulated:

- What has been the impact of the assessment reform ‘Assessment without Levels’ on teaching and assessment practice?
- What are the Policy into Practice implications of the assessment reform ‘Assessment without Levels’ for Primary school teachers?

These research questions will aim to provide:

- An analysis of how AwLs has changed assessment practices in schools from the point of view of teachers who teach in SATs years.
- Primary school teacher views based on their experience of AwLs so far.
- An analysis of the overall impact Assessment without Levels has had on Primary school teachers.

To fulfil these aims I will address the gap in research about the impact of AwLs on primary teachers' through their perspectives and experiences of this education policy reform. In doing so I hope to contribute new knowledge to the literature on assessment in Primary schools by finding a novel way to explore the unknown territory of AwLs, discover collective and individual experiences of other teachers and guide how policy changes are better managed in the future.

The conclusion in my thesis will present a set of recommendations informed by my research data and analysis as to how to best support teachers through education policy reforms as well as highlighting future research opportunities.

### 1.3 CONTEXT AND RATIONALE

This research focused on the removal of levels from state-maintained schools in England, specifically Primary schools. Following a change in government in 2010 a new curriculum was introduced officially in September 2014. NC Levels continued to be used to assess the end of Key Stage 1 (KS1 age 6-7) and Key Stage 2 (KS2 age 10-11) until May 2016, when SATs would test the new curriculum. Therefore, this research targeted teachers' experiences transitioning to AwLs and the first 1-2years with SATs under AwLs. These experiences will be contextualised in the history of National Curriculum Levels (NCLs) and literature concerning educational assessment.

#### 1.3.1 A brief history of Assessment with Levels

In England, assessment with levels was introduced in State Maintained schools as a corresponding framework to the first National Curriculum in 1988. The Task Group on Assessment and Testing (DES, 1988a), headed by Professor Paul Black, recommended a tiered assessment framework for the end of each Key Stage. What was originally intended as a set of Standardised Assessment Tasks (SATs) assessed by the teacher were replaced with externally marked tests for English, Maths and Science. The original idea was criticised for being unworkable due to logistics, a lack of training on assessment, and an unmanageable workload (Whetton, 2009). KS1 SATs were relaxed in 2005, with test results informing the teacher SA based KS1 SATs judgments (Whetton, 2009), following this change

Science KS2 SATs were removed in 2009 by The Department for Children schools and Families (DCSF).

The inclusion of 'standardised' in the SATs refers to set assessment: conditions, content and marking. Every student is assessed to the same 'standard'. Teacher SA was required by the Local Education Authority (LEA) at the time but did not form part of end of KS2 SATs judgments. Levels ranged from 1-5 in Primary schools with students at the end of KS1 and KS2 expected to achieve a Level 2 and 4 respectively. In 1996, Primary school league tables were introduced using KS2 SATs data to rank schools.

NCLs were subsequently shaped by New Labour after 1997. New Labour introduced 'Assessment for Learning' guidance for formative assessment and 'Assessing Pupil Progress' (DCSF, 2009) for teacher SA. The former focused on KS2 SATs as the measure of improved standards and the latter comprised tiered level descriptors for reading, writing, maths and science. APP Level descriptors were comprehensive, divided subjects into topics, and accompanied later by the 'Standards' files exemplifying each Level. Although initially only intended for end of Key Stage teacher SA, NCLs eventually permeated all years in Primary school, developing as a means of tracking progress. Sub-level assessments labelled 'a', 'b', and 'c' became common place though cautioned against by government guidance on APP (Getting to Grips with Assessing Pupils' Progress, DCSF, 2009).

When Labour lost the 2010 election, the new Coalition Government initiated Curriculum and Assessment reform (explored in more detail in [Chapter 3- Education Policy](#)). Guidance for NCLs was archived including APP and all other National Strategies resources. The new Curriculum draft was released in 2013 before becoming mandatory in 2014. The transition to AwLs lasted until 2015/2016, up to then KS1 and KS2 SATs assessed the previous curriculum using NCLs. Table 1-1 summarises the major changes made to SATs in 2016. Notably a grammar test was introduced in both KS1 and KS2, writing was now moderated teacher SA, and the KS2 mental maths paper was replaced with an arithmetic paper, also added to KS1. SATs testing remained part of the KS1 TA judgment, with flexibility over the timing of test administration.

Table 1-1 Standardised assessment tasks (SATs) for assessment with and without levels in Primary schools.

	NCLs			AwLs		
KS1	English			English		
	Reading Paper	Spelling Paper		Reading Paper	Grammar Paper	
	Maths			Maths		
	Level 2 Paper <i>(Optional Level 3 Paper)</i>			Arithmetic Paper	Reasoning Paper	
KS2	English			English		
	Reading Paper <i>(Optional Level 6 Paper)</i>	Writing Paper <i>(Optional Level 6 Paper)</i>	Spelling Paper <i>(Optional Level 6 Paper)</i>	Reading Paper	SPAG-Grammar Paper 1	SPAG- Spelling Paper 2
	Maths			Maths		
	Mental Maths Test <i>(Optional Level 6 Paper)</i>	Paper 1 <i>(Optional Level 6 Paper)</i>	Paper 2 <i>(Optional Level 6 Paper)</i>	Arithmetic Paper	Reasoning Paper 1	Reasoning Paper 2

Schools were given autonomy over their non-statutory in-school assessments with the exceptions of end KS1 and KS2, years 2 and 6. For these years in March 2016 Interim teacher assessment frameworks (ITAFs) for reading, writing, maths and science were introduced. These ITAFs were brief when compared to APP. Its use was outlined as end of key stage assessment specifying they were not designed for tracking, assessing individual pieces of work or informing school programmes of study (Interim teacher assessment frameworks at the end of key stage 1, STA, 2015).

Published the following autumn, exemplars provided examples of pupil work to the standards of the criteria shown in

	NCL		AwLs	
Grade criteria	KS1	Level 1	KS1	Working towards expected
		Level 2		Working at expected
				Working with greater depth in expected
	KS2	Level 3	KS2	Working towards expected
		Level 4		Working at expected

		Level 5		Working with greater depth in expected
--	--	---------	--	--

. For all other years, the curriculum was available for teachers to see end of year expectations within KS1 and KS2 but gave no criteria for the assessment of: working towards expected, expected, or greater depth within expected.

Table 1-2 Assessment criteria for end of Key Stage 1 and Key Stage 2 NCL and AwLs.

		NCL		AwLs	
Grade criteria	KS1	Level 1	KS1	Working towards expected	
		Level 2		Working at expected	
				Working with greater depth in expected	
	KS2	Level 3	KS2	Working towards expected	
		Level 4		Working at expected	
		Level 5		Working with greater depth in expected	

### 1.3.2 Rationale for AwLs research

Educational reform provides opportunity to explore teaching practices and capture the impact of reform through teachers' experiences. For example, studies after the introduction of 'Curriculum for Excellence' in Scottish Secondary schools explored the impact of policy change through teachers' experiences (Biesta et al., 2015; 2017; Priestley 2011b, Priestley et al., 2013; 2015; 2016). Curriculum for Excellence was a major change for Secondary teachers in Scotland and draws parallels with the 2010 Coalition Reforms in England. Biesta and Priestley demonstrated the importance of teacher experiences of reform as well as the opportunity it brings to explore teachers' practices.

Primary assessment has entered unknown territory, with increased autonomy for non-statutory assessment and a need to adjust to SATs with new curriculum expectations. This research is important in producing a view of AwLs in schools and its impact on teachers. Primary schools are now left with a significantly stripped back assessment framework in addition to acclimatising to a new curriculum. As such the implications to school autonomy

for non-statutory assessment need to be addressed in light of previous studies indicating that even with NCLs, with guidance and exemplars, teacher SA required more intervention and support not less (Black et al., 2011; Cox 2008; Harlen, 2005; Stobart 2009).

#### 1.4 STRUCTURE OF THESIS

The thesis structure is as follows:

**Chapter 1 - Introduction to the research.** This chapter contextualises and justifies the topic of the research. Research aims are presented as well as reasons for my interest in the topic.

**Chapter 2 - Literature review.** This chapter reviews assessment literature relating to Primary school assessment and the various forms it takes. Considerations are made of the place learning and learning theory have in assessment and, moreover, how accurately assessments are assumed to represent the learning intended. Research concerned with the assessment in-situ will be reviewed drawing out the implications of high-stakes assessment in Primary schools as part of school accountability. Finally, research addressing AwLs specially will be reviewed.

**Chapter 3 - Education Policy.** This chapter reviews education policy regarding assessment from 1988 to 2015. Specific focus is given to the justifications for removal of NCLs and how recommendations of Government commissioned reviews in to the National Curriculum and its assessment where utilised.

**Chapter 4 – Neoliberalism and Education Policy.** This chapter views Education Policy through a Neoliberal lens. Key discourses within this are identified and explored in terms of how they impact on both teaching and assessment practises.

**Chapter 5 – Teacher Agency and Identity.** In this chapter theory relating to teacher Agency and Identity are presented. Discourses identified in Chapter 4 are then viewed through this lens of teacher Agency and Identity to consider the impact this has on Teaching and Assessment practices.

**Chapter 6 – Methodology and Methods.** Within this chapter I set out my positionality as a researcher and the methodology I have adopted for this study. The methods I have adopted

are discussed, justifying their relevance to the research. Analysis procedures for my data are presented alongside ethical considerations embedded into the research design.

**Chapter 7 – Survey Findings and Analysis.** This chapter presents the findings from the survey data. Findings were analysed by comparing responses between demographic groups of teachers such as experience assessing with NCLs or not. Relevant themes identified in the literature reviewed in Chapter 2 are used to contextualise and support these findings.

**Chapter 8 – Interview Findings and Analysis.** This chapter presents the findings from teacher and SLT Interview data. Key themes are drawn out for each Interview setting and triangulated between settings. Findings will be contextualised and supported with relevant literature from Chapters 2, 3 and 4.

**Chapter 9 – Teacher Agency Analysis and Discussion.** Teacher interview data is analysed in this chapter using the Ecological Theory of Agency presented in Chapter 5. The impact of AwLs on teacher agency is explored in terms of how agentic teachers were in terms of managing the impact of AwLs on student learning.

**Chapter 10 – Further Analysis and Discussion.** This chapter draws together the findings of the research through triangulation of the Survey and Interview data and the literature presented in the thesis thus far.

**Chapter 11 – Conclusion and Recommendations.** In this final chapter, I address how the research findings answer the research questions. Contributions to knowledge are presented. The thesis concludes by reflecting on the research design and my positionality as researcher. From these conclusions I present recommendations for research and practice.

## 2 LITERATURE REVIEW

---

### 2.1 INTRODUCTION

This chapter aims to review the current literature pertaining to Primary Assessment. How assessment is understood as a concept with common forms of assessment, formative and summative, is explored through a contextualised example of assessment in-situ. The interplay between assessment and learning will be considered drawing attention to the persistence of more traditional learning theories, which through assessments such as testing encourage pedagogy more aligned to behaviourist than contemporary constructivist theories (Baird et al., 2017; James, 2006).

This research focuses on primary school AwLs in England but will draw on literature outside of this context, where it is felt to add depth and perspective, for example when discussing the impact of standardised testing used for school accountability. This practice is not unique to England (Biesta, 2008; Buchannan, 2015), therefore literature and studies addressing its impact on teaching and learning help build a more comprehensive view of assessment in primary schools.

Literature regarding assessment theory will be addressed in two sections. The first will explore in-situ assessment types generally understood as formative and summative. This is built on in Chapter 2.4 engaging critically with this field where the complexities of assessment terminology are presented. The literature does not agree on how or if assessment can have a dual formative and summative use or purpose. Black and Wiliam (2018) suggests the lines between these distinctions are not as important as the knowledge the assessments provides the teacher, Harlen & James (1997) find a distinction useful because it prevents summative assessment over shadowing formative assessment. The second assessment theory section will review the literature on the validity and reliability. In providing an overview of the historical development of these concepts is it clear that methodological perspectives impact on how academics measure the quality and effectiveness of an assessment. Of the foundations laid by Cronbach (1971) and Messick

(1986) an assessment is viewed as valid if it is able to measure what it purports to. Key to this area of literature is the subjectivity of interpretation (Guba and Lincoln, 1986). This causes tensions where statistical models are used to measure assessment validity and reliability. This section of the literature review is then put into context by relating this theory to studies on primary assessment and its impact on teaching and learning in [Chapter 2.6](#).

A major facet of Primary Assessment is teacher based assessment which is difficult to judge as reliable and often accused of being invalid because of how subjective it is by nature (Johnson, 2013). The impact of this is seen in the literature through a lack of trust in teacher assessment and a propensity towards numerical and statistical measures of learning such as standardised testing (Ball et al., 2012; Green, 2009; 2011). Though literature strongly suggests that these measures of learning may be no more valid or reliable when they cannot sample the curriculum in its breadth and are subject to pressures which result in teaching to the tested content (Koretz, 2017; Hoyle and Wallace, 2007; Torrance, 2011; Whetton, 2009).

This chapter concludes with the research published since the removal of levels. It tells an early, incomplete picture of AwLs in Primary schools, which does not record the gains proposed by Policy makers of removing NCLs, seeing instead a higher reliance on any assessment guidance available and a continuation of TTTT practices (Ward and Quennerstedt, 2019; Pratt and Alderton, 2017; 2019).

## 2.2 UNDERSTANDING ASSESSMENT IN AN EDUCATIONAL CONTEXT

### 2.2.1 Assessment definition

Assessment is a process of judgment making by assigning a value to something. In terms of educational assessment, certain knowledge or skills which are of value are judged. These educational assessments are experienced as definitive judgments of attainment when viewed summatively.

When considering the meaning of 'assessment', historically its root word 'assess' meant estimating the value of someone or something for taxation (Oxford English Dictionary, 2020). The definition expanded to include judgments of a person beyond financial value.

Assess, *v.*

To evaluate (a person or thing); to estimate (the quality, value, or extent of), to gauge or judge.

1934—1979

Oxford English Dictionary (OED, 2020)

Education value judgments were included within the definition of assessment in 1956.

Assessment, *n.*

5.b. Education. The process or means of evaluating academic work; an examination or test.

Cf. 1956-1985

OED (2020)

Assessment in education has become a major part of schooling in England, central to teaching and learning. However, the definition above makes no reference to learning, but instead academic work. This is important when considering a common ‘folk’ theory, after Bruner, of assessment where an element of learning is observed in a task and judged as learnt or not. Such ‘folk’ theories arise because;

*...our interactions with others are deeply affected by our everyday intuitive theories about how our minds and the minds of others work. These theories, rarely made explicit, are omnipresent in practical and educational decisions (Olson and Bruner, 1996: 10)*

The next section will explore this ‘folk’ concept of assessment while contextualising assessment in Primary Education.

## 2.2.2 A contextualised example of educational assessment

As a Primary school teacher, I adopted a common ‘folk’ theory of assessment, as a judgment of whether learning had occurred or not. For example, after counting instruction, a learning aim/objective counting from 1-5 could be assessed by the pupils verbally reciting 1 to 5. Success of this task (or not) could be interpreted as showing the student has met the counting objective. However, as Gipps (2011: 140) argues ‘Assessment is not an exact

science, and we must stop presenting it as such', it is not as simple as judging verbal performance of '1, 2, 3, 4, 5'.

This is not to say that intended learning has not been achieved but that this is only one interpretation. Table 2-1 shows an example of two students' attempts at the task of targeting counting 1-5.

*Table 2-1 Example of Child A's and B's attempts at two assessment tasks for counting 1-5.*

	Assessment 1- counting verbally 1-5	Assessment 2- identifying in tubs amounts of 5 with toy blocks
Child A	Yes	No
Child B	Yes	Yes

In this example, both students can verbally count 1-5, however, with the second task, only Child B could identify, through counting, tubs containing five blocks. Child A miscounted, counting the same blocks repeatedly or missing some altogether. What this means is that what is understood as counting needs to be established. National Curriculum non-statutory guidance exemplifies counting '...as reciting numbers and...enumerating objects' (DfE, 2013: 6). In this context, the first assessment represents learning of counting as reciting numbers in order but cannot represent counting as 'enumerating objects'. However, this could be demonstrated by the second example as this requires a number to be assigned to objects or people (enumeration). In examining and defining the visible learning evident in this task we are addressing the 'construct' of the assessment, the part of the task which can be interpreted to represent learning.

It is important to consider where assessment 'constructs' originate and what is demanded of them. From Baird et al. (2017)'s perspective, assessment constructs are not merely a derived indicator of meeting curriculum aims, but can be mediated by policy, dominant contemporary learning theories, contemporary experts, and performance on assessments themselves.

Returning to the example in Table 2-1, Task 1 does not contain the construct required to judge counting as 'enumerating', but it does represent 'counting as reciting numbers'. This is

not a criticism of the assessment task, only that it is important to consider what is being assessed and its purpose. If the purpose of assessment 1 was to indicate preparedness for verbally reciting numbers beyond 5 to 10, then task 1 is sufficient. However, if the purpose was to indicate readiness to recite and enumerate numbers to 10 then both tasks are needed to cover both assessment constructs.

### 2.2.3 Objective considerations of assessment uses and purposes

This example above is a simple one, but even in its simplicity it demonstrates issues with conceptualising assessment. Assessment is often more complex than an objective representation of learning. Historically, assessment has been viewed as an outside measurement of a 'thing', existing independent of the context in which it occurs (Scharaschkin, 2017). This objective view of assessment lingers in educational assessment when used as a measure of, amongst other things, students' progress and ability, teacher progress and ability, and quality of schools. As such, the term 'assessment' has become a discourse, unquestioned by some in its use to hold schools and teachers to account, informing a 'folk' view of assessment referenced above.

Educational assessments are interpreted as judgments of ability or lack thereof, with educational qualifications awarded based on assessment performance. Year 6 (Yr6 ) SATs results are used to predict secondary education ability using Progress 8 to target progress (DfE 2016). This real-world position assumes the performance of a given assessment task represents the construct intended. As Baird et al. (2017) state, 'Educational assessment constructs are not 'out there' waiting to be discovered.' (Baird et al., 2017: 322). It is necessary to examine the performance element of assessment and assumptions based on it.

### 2.2.4 Introduction of terms validity and reliability of assessment

In an ideal world, assessment tasks would be chosen with constructs which represent specific learning aim/objectives/goals easily providing a firm foundation for interpretations of learning. However, educational assessments are based on student actions which may or may not communicate their learning on a task. Referring back to Table 2-1, Child A was unsuccessful at Task 2, they could not count the bricks accurately, but when repeated Child

A could perform the task accurately. Possibly, Child A was unfamiliar with components of the task which could have impacted on their performance. This considers the extent to which the assessment actually assesses what it aims to – this being the *validity* of the assessment; and the repeatability of assessments to gain consistent judgments - the *reliability*. As James (2017: 410) states, ‘the confidence placed in them [assessments] depends on the strength of the arrangements for fairly judging performances and outcomes.’. Validity considers not only appropriateness of assessment constructs, but also the interpretations of the assessment (Cronbach, 1971). Considerations of validity and reliability of educational assessment are discussed in more detail in [Chapter 2.5](#) and [2.6](#) but this chapter has set the foundations for a more theoretical exploration of these aspects of assessment.

### 2.2.5 In-situ assessment types

An additional consideration to the example in Table 2-1 is the *purpose* of the assessment. So far these tasks have been discussed in relation to student readiness to progress in counting, with the judgment perhaps informing the teacher of which students were ready and which needed more intervention and where. These tasks could also be used to judge how much the student has learnt without the assessment being used to inform future teaching.

Teachers, in-situ, refer most commonly to these assessments as formative assessment (FA) and summative assessment (SA) respectively.

Assessment terminology within the field of Primary assessment literature does not fit neatly into these categories, nonetheless, to understand the intricacies of assessment terminologies it is useful to start with these most commonly used assessment terms for teachers in-situ. The use of FA and SA terminology was influenced by the creation of a national assessment framework for the newly created National Curriculum in 1988. The Task Group on Assessment and Testing (TGAT in DES, 1988a) created recommendations in line with the new curriculum, defining assessment via four purposes;

- *formative, so that the positive achievements of a pupil may be recognised and discussed and the appropriate next steps may be planned;*

- *diagnostic, through which learning difficulties may be scrutinised and classified so that appropriate remedial help and guidance can be provided;*
- *summative, for the recording of the overall achievement of a pupil in a systematic way;*
- *evaluative, by means of which some aspects of the work of a school, an LEA or other discrete part of the educational service can be assessed and/or reported upon. (DES, 1988A, para 23)*

Aspects of diagnostic assessment have been integrated into FA definitions. Recent definitions of FA still encapsulate the diagnostic category:

*It [FA] enables teachers to identify when pupils are struggling, when they have consolidated learning and when they are ready to progress. (DfE 2015: 19)*

Similarly, evaluative assessment has become part of the general SA definition. In school SA for the monitoring of attainment is reflective of TGAT's (DES, 1988a) evaluative assessment.

*Nationally standardised summative assessment allows the Government to hold providers of education (Schools, local authorities, academy chains etc.) to account and to measure the impact of educational policy making. (DfE 2015: 21)*

Both the Assessment Policy Task Group (APTG) and then later the Assessment Reform Group (ARG), aimed to bridge assessment research with teacher accessibility. Investments were made to into FA and SA leading to a reconceptualisation of FA and SA into assessment *for* and *of* learning respectively (Daugherty, 2007). The APTG funded a review of research on formative assessment practices, producing both an academic paper and teacher accessible pamphlet '*Inside the black box*' (Black and Wiliam, 1998b). This combination of pamphlet style publications and research increased the familiarity of assessment terminology in schools with national guidance for schools using assessment *of* and *for* learning interchangeably with summative and formative assessment respectively (DfES, 2006).

The use of the definitions of assessment from TGAT, APTG, ARG and NS has resulted in the language of formative, summative, and assessment *of* and *for* learning commonly used in schools by teachers. This can be viewed as a general split between assessment to inform future learning, or assessment which judges what has been learnt. However, the ways in which educational assessment in practice is not kept in distinct categories will be explored in Chapter 2.4.

## 2.3 THE PLACE OF ASSESSMENT IN THEORIES OF LEARNING?

The use of FA to support learning has developed since Black and Wiliam (1998b) 'Inside the Black Box' and become integrated into Primary teaching pedagogy. FA's fundamental aim is to improve learning; therefore, FA practices have developed alongside developments in learning theory. Summative assessments in contrast have remained static. Therefore, this section will address the place and role of learning theory in summative and evaluative assessments. This is to be understood within the context of the previous section, that assessment does not exist as a dichotomy between formative and summative, rather that the purpose of the assessment acts as the key driver.

Before exploring how assessment relates to learning theory, influential learning theories will be reviewed. As the field of learning theory is extensive, I will draw mainly on literature which addresses the interaction between learning theory and assessment.

### 2.3.1 Educational learning theories- an overview

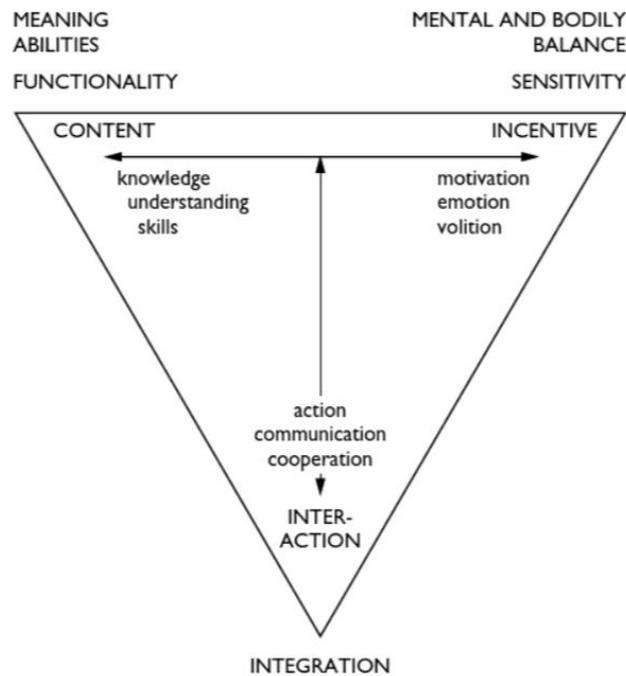
The field of learning theory is complex, comprising differing theories of learning which vary in their capacity to complement or contest each other. Learning theories are not static through time, with dominant theories shifting. Historically, behaviourist learning theories dominated education until the 1970s and 1980s (James and Lewis, 2012). Learning was viewed as linear, resulting from behavioural conditioning, achieved by the accrual of skills broken down into practicable elements with students trained to respond to instruction (James, 2006; Gipps, 2011).

Increasing in influence in the 1980s and 1990s, cognitive theories focus on the science of the brain relating learning to developments in computer programming at the time (Bruner 2009). Differing from behavioural learning theories the '...role of the teacher is to help 'novices' to acquire 'expert' understanding of conceptual structures and processing strategies' (James, 2006: 55). Piaget (1952 [1936]) was influential in developing cognitive learning theories, framing new learning within the context of previous learning where assimilation and accommodation is conceptualised within interactions with the immediate environment. Additionally, learning was viewed by Piaget (1952[1936]) in ages and stages, where a child was only capable of achieving certain learning goals at a given age. In a similar

fashion to behaviourism, prior knowledge is used to predict future learning, however in cognitive theories new learning is interpreted through previous knowledge (Pritchard, 2017). Within cognitive theories, learning can be differentiated into a hierarchy of skills, for example, Bloom's taxonomy (1956) orders learning in six stages with each stage progressing in difficulty from lower order skills of knowledge and comprehension to higher order skills of synthesis and evaluation.

Social-constructivist, also referred to as socio-cultural theories of learning are a recent development in learning theories from pragmatic and interpretivist theory (Daniels, 2016). Learning happens as a result of the interaction between the individual and the *social* environment, constructing new knowledge through these experiences (see [Chapter 6](#)). Social constructivist theories derive from the work of Vygotsky 1978, 1986 [1934]), and are evident in Bruner's (2009) emphasis on the importance of culture in learning, Engeström's (1987) expansive learning theory, and ecological model of learning from Bronfenbrenner (1977; 1979) viewing learning as holistic across different experiences. Vygotsky's *zone of proximal development* enables the theorisation of learning mediated through the social interaction of the teacher and the student (Vygotsky 1978; 1986 [1934]). This theory contributes towards the notion of 'scaffolding' to support learners conceptual understanding, forming new knowledge through their interaction with others and objects which they could not do unassisted. Though these theories of learning differ they all demonstrate a shift in learning theory towards a socially constructive learning and away from an instrumentalist theory of learning. Through these developments, teaching has become more aligned with student perceptions and experiences and less about bestowing the teacher's knowledge through instruction. Illeris (2017) posits differences between learning theories arose because individually they focus on part of the process. For example, behavioural and cognitive learning theories focus predominantly on internal processes, whereas social theories emphasise external interactions. In the Figure 2-1 below, Illeris (2017) extrapolates these dimensions of learning.

Figure 2-1 'Three dimensions of learning and competence' Illeris (2017: 12)



(Illeris, 2017: 12)

For learning to happen, students need to recall previous learning while interacting with the teacher; listening, looking and asking questions if needed. Additionally, learners need to be sufficiently 'incentivised' to participate and recall previous knowledge. Assuming all this, effective interaction from the teacher through accessible, clear and appropriate content is needed. Learning, therefore, is contingent on all three dimensions and may be inhibited by: student motivation and concentration (incentive), previous knowledge (content), and effective teacher communication (environment) (Illeris 2017: 13).

Reviewing historically dominant learning theories provides a view that how learning is understood to occur has shifted from behavioural to constructivist in nature. This shift will be important when analysing how assessment theory has evolved along-side learning theories and the impact AwLs practices have on learning itself.

### 2.3.2 The relationship between learning theory and assessment

The overview of learning theories above is provided to give context to discussions on assessment theory, specifically conceptualisations of assessment validity and how pressures on testing, such as the KS2 SATs, can impact on the very learning they purport to measure.

### 2.3.2.1 Persistent influence of behavioural learning theory on standardised testing

Influence from behavioural learning theories persists through some aspects of standardised testing. There has been an increasing influence from cognitive theories of learning, aiming to capture higher order thinking skills (Snow and Lohman, 1993). However, a psychometric approach found in large scale assessments such as the SATs still dominates (Baird et al., 2017; Schoenfeld, 2017; Scharaschkin, 2017). KS1 and KS2 SATs tests require students to memorise and recall information, at times out of context, characteristic of behavioural learning theories (Gipps, 2011). Tests favouring recall do not encourage a deeper resolved learning typical of constructivist views of learning now favoured in education and alternatively encourage instrumentalist learning (Jarvis et al., 2003). Further influence of behavioural learning theories is evident when the goal is ‘...differentiating between individuals who possess certain attributes, or in determining the degree to which they do so.’ (James, 2006: 48). This interpretation of assessment data views learning as a fixed, predictable capacity, for example, the use of KS2 data to predict achievement at KS4, rather than aligned with social constructivist theories where learning is regarded as fluid, and educational constructs considered goals (Baird et al., 2017: 329-330).

Where learning is a joint not solo venture, dependent on and relative to the context of the individual, James and Lewis (2012) argue knowledge cannot be separated out of the learner’s own context therefore assessment should reflect this.

*According to sociocultural theory, the transfer and translation of cultural knowledge – learning and teaching – involves both externalisation and internalisation through shared activity (interaction) and individual learning activity (action)... (James, 2017: 407-408)*

Sfard (1998) similarly refers to two type of theory; a participatory metaphor for learning and an acquisition metaphor for learning, stating theory is not neat and universally applicable, so in practice a combination of theories is beneficial. Learning theories do not exist within neat boxes and theorising something as complex as learning results in blurring the boundaries between theories (James, 2006: 59)

*...behaviourist approaches seem to work perfectly well when the focus is on the development of some basic skills or habitual behaviours...On the other*

*hand, cognitivist approaches seem to be best when deep understanding of conceptual structures within subject domains is the desired outcome.*

### 2.3.2.2 Tensions between learning theory developments and educational assessment

The concept of standardised large-scale educational assessment as a measure of educational standards carries with it an assumption that attainment itself is quantifiable. Viewed from this perspective, assessment is purported to be an objective measurement of reality, assuming knowledge is amenable to being captured independent of the assessment process (Scharaschkin, 2017). Within this view, the link between learning and assessment requires scrutiny when considering learning as an ‘innate mental characteristic’, inseparable from the individual (James 2006: 56). In disagreement with this, Scharaschkin (2017) suggests that any measurement of someone does not exist independently of the assessment itself, and therefore understandings of educational assessment should differentiate between assessment of attainment and assessment of the individual.

Baird et al. (2017) question how ‘useful and predictive’ educational assessments are at assessing the learning intended, under varying demands and uses of large-scale standardised assessments. They suggest that aligning assessments with developments in learning theory is a necessary first step in producing useful assessments. Alternatively, Wiliam (2017) suggests that the issue in capturing learning in assessments is not test theory, or a lack of alignment between developments in learning theory and assessments, rather, the expectations of what tests *can* achieve does not match what test *do* achieve.

*The problem is not that the theory is inadequate. The problem is that the theory gives answers that people don't like. (Wiliam, 2017: 296)*

Wiliam (2017) is referring to limitations of the information large-scale standardised assessment can provide. Similarly, commentary from Goldstein (2017) agrees test theory is not at fault, but, different to Wiliam (2017), Goldstein (2017) criticised Baird et al. (2017) for failing to demonstrate why learning theory *should* inform assessment construction.

According to Goldstein (2017: 388):

*...assessment is essentially a technology that can have no theoretical basis of itself but seeks to be motivated by the area to which it is applied, in this case education.*

Goldstein is dismissive of claims by Baird et al. (2017) that standardised tests impact on the teaching of what is being assessed (assessment back-wash) because they do not align with learning, arguing that:

*...quantitative assessments do not have to be constructed on that basis – they may be designed simply as monitoring devices or instruments to evaluate an educational reform. (Goldstein, 2017: 389-90)*

Goldstein (2017) in favouring quantitative measures, ignores the wealth of qualitative assessment literature supporting Baird et al. (2017)'s position; assessments may not be designed to influence what is taught restrictively, but this does not prevent it happening. As Baird et al. (2017: 319) argue;

*Despite the apparent lack of a solid relationship between learning theory and assessment practice, strong relationships between assessment and teaching and learning practices are claimed in the literature.*

Moreover, normative assessments are '...intended to affect the attribute being assessed.' (ibid: 320) with teaching and learning intentionally being affected by the assessment, known as wash-back. For example, year 1 phonics and year 4 times table test were introduced by Government to target specific areas of learning. These policy initiatives carry a discourse ignored by Goldstein (2017), namely that tests are purposefully used to target learning through teaching.

To summarise, understanding of learning has developed beyond previous theories as being both linear and atomised into smaller buildable units, and thus should be assessed as such. As Gipps (2011: 4) argues;

*We need to put on to the assessment agenda issues of learning style and depth...educational assessment for the next century must be based on our best current understanding of theories of learning.*

The extent an assessment represents attainment is a problem when it neither captures nor encourages intended learning. Learning needs to be central to assessment to infer the meaning it aspires to. Approaches attempting to do this are in their infancy and dominated by standardised assessment practices. Alignment between developments in learning theory and assessment is still aspirational and central to validity concerns discussed in [Chapter 2.5](#).

## 2.4 VIEWS OF ASSESSMENT TYPES

As mentioned in [Chapter 2.2](#), educational assessment is typically categorised as summative or formative depending on its purpose or use (Baird et al., 2017: 337). Therefore, it cannot be assumed, for example, that a test is a summative assessment or teacher assessment is formative, instead the intended inference of the assessment is central to how assessment is defined rather than what specific task itself is (Black and Wiliam, 2018)

These types of assessment do not exist as a dichotomy, nor is there a universally agreed upon definition, resulting in tension within educational assessment literature when attempting to define different types of assessment, as well as if it is even useful to do so (Black and Wiliam, 2018). Idiosyncrasies of formative and summative assessment will be considered separately before exploring their interaction in the literature.

### 2.4.1 Formative assessment

Throughout the literature FA is seen, in terms of children's learning, as the most important form of assessment (Black 1998; 2015; Black and Wiliam 1998a; 1998b; Harlen and James, 1997). Historically, assessment referred to the evaluation of learning and did not encapsulate tasks within the teaching sequence designed to drive learning, these were considered part of good teaching practice (Wiliam, 2011). Previously developed by Scriven (1967) referencing formative evaluation in relation to teaching, Bloom (1969) and shortly after Bloom et al. (1971) used the expression *formative evaluation* meaning 'to provide feedback and correctives at each stage in the teaching-learning process' (Bloom, 1969: 48).

Baird et al. (2017) argue that although the term can be traced back to Scriven (1967) and Bloom (1969), its use in English education today is more reflective of a constructivist approach to learning, compared to behaviourist theories of Scriven (1967) and Bloom (1969). What *feedback* means is central to this argument, with Wiliam (2011: 4) agreeing that Bloom's definition is unhelpful because it separates, '...information from its instructional consequences' arguing that '...the use of assessment information to improve learning cannot be separated from the instructional system within which it is provided.' Therefore, formative feedback is not merely the information of what could improve but requires the communication of what the learner needs to do to improve (Shute, 2008: 154). For example, Black & Wiliam's (2009: 9) definition;

*Practice in a classroom is formative to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers, to make decisions about the next steps in instruction that are likely to be better, or better founded, than the decisions they would have taken in the absence of the evidence that was elicited.*

Clear in this definition is the importance of assessment for learning to happen. It is important to note that the term Assessment *for* Learning (AfL) is often conflated with formative assessment, with the use of the term AfL recently preferred to mitigate complexities associated with a FA definition (Wiliam, 2011). However, as Bennet (2009) argues, this only shifts the previous definition issues to the new term. Another motivation for the shift in language is the emphasis AfL places on the learner as Broadfoot et al. (1999: 7) argue assessment;

*...may be formative in helping the teacher to identify areas where more explanation or practice is needed. But for the pupils, the marks or remarks on their work may tell them about their success or failure but not about how to make progress towards further learning.*

Similarly, Baird et al. (2017) prefers the distinction between AfL and formative assessment, because it emphasises student autonomy in their learning.

#### 2.4.2 Critique of formative assessment

Black and Wiliam's (1998b) pamphlet 'Inside the Black Box' helped to promote FA's worth in teaching. This paper became very influential in the field of education. Its pamphlet format made it accessible to practising teachers increasing the profile of the term FA within schools. However, it did not go without criticism (Dwyer, 1998; Bennet, 2011). Bennet (2011) is critical of Black and Wiliam (1998b) emphasising the need to quantify improvements in learning as a result of FA intervention and argued their meta-analysis, Black and Wiliam (1998a) was too diverse and large to produce meaningful results which Black and Wiliam (1998b) was founded on. Black and Wiliam (1998a) did not call their review a meta-analysis, however they do describe effect sizes which may have contributed to Bennet's (2011) conclusion. To Harlen and James (1997: 378), FA is described as more of a 'human act of judgment' than a 'technical matter of measurement' suggesting they disagree with Bennet's argument for definitive quantitative measures of FA's validity. Concurrently, Baird (2010) believed some of Bennet's criticism of FA related to a positivist perspective in North America where FA is strongly associated with teacher led tests and quiz driven feedback (Baird et al., 2017). Bennet (2011), additionally, neglects to mention further research from Black and Wiliam (Black and Wiliam, 2003; Wiliam et al., 2004) reflecting on their research, highlighting how it 'represents our opinions and prejudices as much as anything else' (Black and Wiliam, 2003: 633). They also state that they used non-traditional 'activities' to publicise their work; this is in reference to a variety of publication types for example, teacher friendly pamphlets. It may be that this non-traditional approach is seen as less academic and more suitable for those working in schools.

However, more recent reflection from Black (2015) agrees with some criticism from Bennet (2011). Black (2015: 163), with hindsight, comments that he did not understand how complex 'the slow pace of teacher change' was in education, and that the 1998 review was 'too optimistic where it said that there was enough evidence to justify the research findings to practical action'.

#### 2.4.3 Summative assessment

Guidance for summative assessment (SA) regarding AwLs indicates two roles for summative assessment: 1) summative assessment as an indicator of attainment and 2) external national

standardised assessments for accountability and school monitoring purposes. As such, SA relies on standardisation for results to be comparable. This can be the format of TA or tests (Bew, 2011; Harlen, 2004b), although there are differing views within both literature and policy as to which is better in terms of producing valid and reliable, or even useful inferences.

The form SA takes in Primary schools can be traced back to recommendations from TGAT'S Report (DES, 1988A) for standardised assessment tasks (SATs) at the end of each key stage. Originally these tasks were to be teacher assessment and formed around a portfolio of evidence. Whetton (2009: 141) criticised these recommendations calling them aspirational rather than achievable; the tasks would have amounted 'to over 100 statements of attainment, which had to be assessed for each child' and describes TGAT'S Report (DES, 1988a: 141) as sowing 'seeds of trouble' for assessment in Primary schools. As such SATs were replaced with external tests. In defence of his recommendations, Black (1998: 64) felt that they needed to have had 'more explanation and implied a need for far more, and far slower, development than the report indicated'. However, he further explains that if this had been said, the report would not have been accepted. This indicates a conflict in providing recommendations that Black (1988) felt the government would accept, even if too simple, compared with accurate recommendations that the government would reject.

#### 2.4.3.1 Relationship between summative and formative assessment

As assessment tends to be categorised by its use or purposes rather than activity (e.g. as a test), the relationship between SA and FA is complex. There is continued debate within the education literature as to whether assessments can fulfil simultaneously summative and formative uses.

While Wiliam and Black (1996: 54) concluded that using the same task does not maximise the benefits for formative or summative assessment practices, stating they were at the 'extreme ends of a continuum', Black and Wiliam (2018: 20) felt '...there should be no conflict between formative and summative assessment – indeed, the distinction would not be useful – because all assessment would be about producing valid inferences about students.' The importance from this view is placed on the inferences drawn from

assessments and how valid they are instead of focusing on a theoretical separation of the terms. Recent literature supports an overlap between tasks designed for formative or summative purposes, one assessment task can provide formative and summative inferences (Baird et al., 2017; Bennet, 2011; Black and Wiliam, 2018; Wiliam, 2017). For example, periodic summative assessment can be used to contribute towards a formative judgment according to Baird et al. (2017). Government definitions of summative assessment from 2015 also advocate the use of summative assessments formatively.

*In-School summative assessment enables teachers to evaluate both pupil learning at the end of an instructional unit or period (based on pupil-level outcomes) and the impact of their own teaching (based on class-level outcomes). Both these purposes help teachers to plan for subsequent teaching and learning. (DfE, 2015: 20)*

Similar blurring of assessment types to suit differing interpretations from TGAT'S Report (DES, 1988a) was felt by Harlen and James (1997) to cause confusion over how and if formative and summative assessments should be distinguished. The report suggested the use of portfolios for both formative and summative purposes. This, Harlen and James (1997) felt, encouraged teachers to sum up their FA to provide an overall summative judgment risking assessment 'washback/backwash' and teaching to the test. This could not only undermine the function of standardising summative assessments by aggregating past assessments, but also compromise the use of FA to develop 'learning with understanding' (Harlen and James, 1997: 367). Black and Wiliam (1998b) addressed the concerns of Harlen & James (1997) that formative assessment, after TGAT'S Report (DES, 1988a), did not have the emphasis and use it should have had.

#### *2.4.3.1.1.1 Assessment referencing*

How the assessment judgment is reached is regarded as a distinction between FA and SA. FA was argued by Harlen and James (1997) to use criterion-referencing (performance is compared to set criteria) and pupil-referencing (also referred to as ipsative, performance compared to an individual's prior performance on the same task), while summative

assessment uses criterion-referencing and norm-referencing is similar to cohort-referencing where,

*...the norm-referenced standard simply represents the level of attainment of a particular student in relation to the level of attainment of all other students who sat the examination in question. (Newton, 2011: 20)*

Whereas cohort referencing is described by Wiliam:

*The notion of a 'standard' in an assessment system is defined as the attachment of specific meanings to specific test or examination scores. When these meanings are in terms of the performance of a group of individuals, the standard can be described as norm-referenced (when the individual is not a member of the reference group) or cohort-referenced (when the individual is a member of the reference group). (Wiliam, 1996: 293)*

Essentially, SATs with levels were criterion-referenced whereas SATs without levels are cohort-referenced and produced by a scaled score related to how well the cohort performed against the criteria of the assessment itself. This is viewed as unfair by Gipps (2011) as the assessment grade is influenced not only by the student themselves but also the cohort sitting the same exam.

*...results from criterion-referenced assessment can also be used for norm-referenced-type purposes, and indeed norms are often used to set and interpret criteria of performance. But nevertheless, the point is well made, that in order to move away from a norm-referenced approach the only other reference we have come up with is that of criteria or standards, whether the result is described as criterion-referenced assessment, graded assessment, or standards-referenced assessment. (Gipps, 2011: 7)*

## 2.5 VALIDITY AND RELIABILITY

Validity and reliability are central considerations of Primary assessment. Before exploring this in [Chapter 2.6](#), theoretical underpinnings will be presented here first.

### 2.5.1 Assessment validity

Chapter 2.2 demonstrated theoretical assumptions that an assessment task works by making visible an aspect of learning - a construct. The extent to which an assessment task/activity '...measures a trait or theoretical construct' is central to how well assessments can be interpreted as 'measuring what they purport to measure'; this is the validity of an assessment (Stenner et al., 1983). Differing views exist of a universal understanding of validity, and how validity theory can be used to determine the effectiveness of an assessment (*inter alia* Borsboom et al., 2009; Newton and Baird 2016).

Historically, assessment validity theory is built around the use of tests as assessments and underpinned by the behaviourist foundations of psychometric testing (Cronbach, 1971; Messick 1990; Stenner et al., 1983). In this view, tests are developed using constructs which sample the domain being assessed (Baird et al., 2017), as Cronbach and Meehl (1955: 283) state, 'A construct is some postulated attribute of people, assumed to be reflected in test performance' thus aiming to '...connect data to inferences' (William, 2017: 397). An assessment's validity is therefore argued to be reliant on the inferences made that the construct represents the relevant domain being assessed (Baird et al., 2017). In the same way as learning is viewed as socially constructed, so are assessment constructs, making them difficult to define, thus impacting on the ease of drawing inferences directly from assessment data, as William (2017: 397) argues;

*...if construct definition were easy, then the domain would probably be so easy to define, or the relationship between data and inference so straightforward, then we would not need the constructs in the first place.*

How valid an assessment is, therefore, is more dependent on the interpretation of the assessment outcome rather than the assessment 'tool' itself; 'one validates, not a test, *but an interpretation of outcomes from a measurement procedure*' (Cronbach 1971: 447, emphasis in original, in Newton and Shaw, 2014: 112). Moreover, this accounts for an overlap in an assessment's use as formative and summative, in addition to an assessment having multiple purposes. For example, Yr6 SATs are used to measure progress, evaluate

teaching and, rate the quality of a school, ‘...the claim to validity is conditional and specific to [each] particular interpretation and use of results.’ (Newton and Shaw, 2014: 112).

### 2.5.2 Types of validity

The dominant position of assessment validity theory today is underpinned by the work of Messick (1990), which built on that of Cronbach and Meehl (1955), and Cronbach (1971); this defines all validity as construct validity- how constructs are interpreted as measuring what was intended (Newton and Shaw, 2014).

Through the field’s continued evolution, validity has been defined by its various facets. In their comprehensive account of developments in assessment validity theory, Newton and Shaw (2014) map out developments therein, from logical and empirical theories of determining validity, to more recent deconstructions of validity. They attribute the development and adoption of these validity categories to the American Psychological Association (APA) Committee on Test Standards, and consecutive additions of Technical Recommendations for Psychological Tests and Diagnostic Techniques (*Standards*). In the process, the focus of validity shifts from a feature of the assessment task to interpretations made from the task (Newton and Shaw, 2014). Four types of validity are generally represented in the literature, though this thesis will focus specifically on the construct validity of Primary Assessment in terms of how well the learning based in the curriculum is represented.

- *Content* validity- concerning the coverage of appropriate and necessary content, i.e. the test covering the skills necessary for good performance, or all the aspects of the subject taught
- *Predictive* validity- relating to test accurately predicting some future performance well
- *Concurrent* validity- concerning test correlation with, or gives substantially the same results as, another test of the same skill
- *Construct* validity- relating to whether the test is an adequate measure of the construct, that is, the underlying (explanatory) skill being assessed.

(Gipps, 2011)

Conceptualising validity in its fragmented attributes has been viewed as problematic in its ambiguity, however, it was useful in demonstrating how one assessment could be interpreted for different purposes (Newton and Shaw, 2014). Subsequent editions of the *APA Standards*, were influenced by Messick (1989a) who conceptualised validity as;

*...a unitary concept, in the sense that score meaning as embodied in construct validity underlies all score-based inferences... (Messick 1989a: 19)*

In Messick's view, all validity is related to construct validity, with two major concerns regarding 'construct-irrelevant variance' and 'construct under-representation' and these forms of validity are particularly relevant to Primary Assessment. For example, there may be construct-irrelevant variance in the KS2 reading tests results where answers depend on students being able to express their thoughts in writing, where *writing* rather than *reading* skills are relied on which are different though related constructs. For construct under-representation, the maths SATs contains a significant proportion of questions relating to numbers resulting in an under-representation of areas like shape and space.

#### 2.5.2.1 Ethical considerations of validity

Addressing the impact of construct-irrelevance and under-representation on students expands validity concepts to include ethical consequences of assessment (Gipps, 2011). Kane (2016) felt that something along the lines of ethical validity considerations should be included when addressing assessment validity. Kane (2006; 2010; 2016) considered validity of both assessment design and interpretation. In doing this he was able to maintain the importance of ethical considerations such as social consequences of testing; while at the same time, the validity of the test, in terms of design, is maintained but inseparable from the *interpretative* arguments (Kane, 2006; 2016). Similarly, Baird et al. (2017: 329-30) argue that assessments are intrinsically linked to learning and have a key purpose '...to generate the very attributes that they assess by making transparent what students should know and be able to do...' this argues that the consequences of educational assessment are entangled within the assessment and thus are concerns of validity. Therefore, including ethical considerations within validity theory as part of consequential validity is divisive, as it relies on validity being a property of interpretations of assessments, which is not universally

acknowledged (Wolming and Wikström 2010). For example, Borsboom et al. (2009) criticise the view that assessment inferences have validity, to them validity is '*...how the test works, and this is certainly not a property of the test score interpretations...but of the measurement itself.*' (Borsboom et al., 2009: 149, emphasis in original).

The most recent (2014) publication of the American Educational Research Association (AERA) *Standards* acknowledges both facets of validity concerning assessment interpretation and design, stating;

*Validity refers to the degree to which evidence and theory support the interpretation of test scores for proposed uses of tests...Construct underrepresentation refers to the degree to which a test fails to compare important aspects of the construct...Construct-irrelevance refers to the degree to which test scores are affected by processes that are extraneous to the test's intended purpose. (AERA, 2014: 11-12)*

Sireci (2016) praised AERA's (2014) incorporation of both purpose and interpretation in the *Standards*, though cautions against the emphasis placed on interpretation as the site of validity to the exclusion of a test itself, deeming tests as 'useless' if their construction is not an equal consideration. Tensions seem to exist in the extent to which validity lies in the assessment and/or interpretations. Whereas Sireci (2016) aims for a balance between the two, these facets of validity theory are seen as incompatible by Cizek (2016), who defends a similar view to Borsboom et al. (2009) that validity lies within the test. In the context of Primary Assessment and particularly the KS2 SATs it is difficult to align this argument with the implications schools face when students are judged as below the expected standard. KS2 assessments do not fulfil a singular purpose of student SA, they have additional uses which have real consequences for schools and pupils explored in [Chapter 2.6](#) and [4](#).

Part of the variation in theoretical positions regarding assessment validity theory could originate from tensions '*...between achieving educational objectives, on the one hand, and (psychometric) measurement, on the other*' (Scharaschkin 2017: 454), and how educational assessment might require different validation focus and procedures. The differing purposes of educational assessments arising from their roots in psychometric measurement require

different emphases. Other arguments for a unified definition of validity come from Newton and Shaw (2016), who feel agreement is needed on how validity is used as a word, embracing ambiguity, and moving away from a technical definition, and instead relying on an understanding of its practical application in varying circumstances, this would mean, radically, abandoning the need for the word altogether.

### 2.5.3 Reliability

In addition to validity, reliability is a main consideration of educational assessments and their suitability. Gipps (2011: 2) defines reliability as;

*...the extent to which an assessment would produce the same, or similar, score if it was given by two different assessors, or given a second time to the same pupil using the same assessor.*

Therefore, a reliable assessment is one that achieves consistency if repeated (Baird and Black, 2013; Harlen, 2000; Wiliam, 2001).

### 2.5.4 Error and reliability

Reliability, traditionally, is rated on the difference between a 'true score', and any 'error'. A 'true score' would be the average a student achieved on a number of tests, for example, which were similar and sampled the same curriculum content (Black and Wiliam, 2006).

Factors which may impact on a student achieving their 'true score' include:

- Students performing differently depending on the particular wording or sampling of questions (Wiliam, 2001)
- External factors impacting on how the student performs on the day of the assessment
- Variance between markers or between assessments with the same marker (Black and Wiliam, 2006)

This is referred to as error, 'the extent to which the result on particular testing occasion departed from the true score' (Wiliam, 2001: 18). Error is theorised to take two forms, 'measurement error' and 'sampling error' (Koretz, 2008; Nisbet and Shaw, 2019; Wiliam, 2001). Measurement error includes bias where '...the test makes systematic errors in measuring a particular characteristic' (Nisbet and Shaw, 2019: 624). Whereas sampling error

concerns how the domain being assessed is sampled (Koretz, 2008; Wiliam, 2001). Additionally, measurement error includes inconsistencies between assessors, either between different assessments or between different assessors, these being *inter-rater* reliability, as well as differences from the same assessor when judging the same task '*intra-rater*' reliability (Gipps, 2011: 57).

### 2.5.5 Interaction between validity and reliability

It cannot be assumed that a valid assessment is reliable or that a reliable assessment is valid, though for a test to have high validity it must be reliable. Wiliam (2001) explains this interaction using a metaphor of lighting an unlit stage, where the stage represents learning and the light represents assessment making learning visible;

*For a given amount of lighting power (cf testing time), one can use a spotlight to illuminate a small part of the stage very brightly, so that one gets a very clear picture of what is happening in the illuminated area (high reliability)...Alternatively, one can use a floodlight to illuminate the whole stage, so that we can get some idea what is going on across the whole stage (high validity), but no clear detail anywhere (low reliability). The validity/reliability relationship is thus one of focus. (Wiliam, 2001: 21)*

This describes a 'tension' Wiliam feels between a reliable test or a valid one and has real implications for Primary assessment which will be explored in [Chapter 2.6](#).

### 2.5.6 Judging assessments beyond validity and reliability

As learning theory has developed beyond behavioural roots, the psychometric underpinnings of the assumption that assessment can objectively represent learning is viewed as inappropriate by some. Gipps (2011: 171), finds issue with the assumption that a reliable assessment is conceptualised as one which can produce an accurate score. Gipps suggest reliability should be replaced with '*...comparability, which is based on consistency*'. Fairness is an additional key aspect of judging assessment, taking account of how educational assessments allow for equity, as each child has differing life experiences, therefore may interpret assessment constructs differently (Gipps and Stobart, 2004). It has also been suggested the terms validity and reliability could be abandoned in exchange for

others of 'authenticity', 'credibility', 'transferability' and 'dependability'. The contested nature of these terms is not restricted to the field of assessment but also has relevance to research methodology referred to in [Chapter 6.7](#) (Guba and Lincoln 1989). Considering this alternative vocabulary may bring an alignment between developments in learning theory away from traditional criteria, which inform the former (Gipps 2011). Further exposure to differing concepts of rigor in assessment may provide teachers access to wider discourses opening up the space for example, for teacher SA to be judged more in terms of its authenticity and credibility in representing student learning than its lack of generalisability. This is discussed in more detail in [2.6.1 Validity and reliability](#). The following sections of this chapter will explore the current theorisation and discourses relating to validity and reliability of primary school assessment.

## 2.6 CONSIDERATIONS OF VALIDITY AND RELIABILITY OF PRIMARY SCHOOL ASSESSMENT

Having discussed the theoretical underpinnings of validity and reliability in educational assessments, this section will consider how this relates to primary school assessment literature. Particular attention will be paid to the two forms of summative assessment in Primary schools, teacher assessment and tests.

### 2.6.1 Validity and reliability of tests as summative assessments

In [Chapter 3](#) Education Policy, the argument is made that national statutory testing has become accepted by Policy makers as the best way to measure learning, allowing schools to be measured, monitored and managed from test data which appears to make learning tangible (Ball et al., 2012; Green, 2011). An essential component of these assessments is standardisation, as Gipps (2011) explains,

*If individuals are to be compared with one another then we need to be certain that the test or assessment was carried out in the same way for all individuals, scored in the same way and the scores interpreted in the same way. Standardization is thus vital...(Gipps 2011: 5)*

What Gipps is referring to is the need to prioritise reliability on statutory assessments in order to enable comparability nationally. Even when considering the best tests available,

Wiliam (2001) estimates that only 80% of students at KS2 are awarded the correct level based on test reliability, meaning theoretically 20% do not. To increase reliability, Wiliam (2001) concluded that tests would need to test more on a small section, or include more questions about each topic, adding 30 hours to decrease error to 10%. He suggested instead that teacher SA should be more prominent in KS2 assessments. Taking this into account, since ‘...only some material and certain tasks are amenable to this type of testing.’ (Gipps 2011: 5) standardised testing has further implications for test validity due to construct under-representation (See 2.6.1.1.). When referring to Wiliam’s (2001) metaphor of the unlit stage with spotlights of a few places,

*...one has no idea what is going on elsewhere, and the people in darkness can get up to all kinds of things, knowing that they won't be seen (not teaching parts of the curriculum not tested). (Wiliam, 2001: 21)*

Therefore, test results cannot universally be interpreted to represent the wider curriculum content intended to be sampled. Within this view tests are not the problem, but interpretations taken from them in judging the performance of a school or teacher therefore devalues tests as a measure of student attainment as intended. Where a test has multiple interpretations based on it, the validity is argued to be even lower (AERA, 2014).

#### 2.6.1.1 High-stakes summative assessment- construct validity impact

An assessment is described as high-stakes if its outcome carries rewards or sanctions which impact on the capacity of the school and its teachers to function. The use of league tables to rank schools based on KS2 SATs, with potential parents choosing to send their children to a better performing school thus decreasing school funding in those schools which lose out under this, fits in these criteria (Biesta, 2008; Moss, 2017). How high-stakes assessments impact on the very learning they aim to improve is a significant concern in the literature (Baird et al., 2017; Black and Wiliam 1998b; Hoyle and Wallace, 2007).

Campbell’s Law can be used to explore this;

*The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be*

*to distort and corrupt the social processes it is intended to monitor.*

*(Campbell, 1976: 46)*

This has implications for Primary school assessments where a quantitative measure of learning, the KS2 SATs, is used to capture student attainment, but additionally to nationally monitor and judge schools and teachers (Stobart 2009). Koretz (2008), who studied standardised assessments used this way, noted six effects, these are;

- Working more effectively
- Teaching more
- Working harder
- Reallocation
- Coaching
- Cheating

The top three of these are desired outcomes from test pressures, however reallocation (strategically allocating teaching time to reflect test representation of curriculum content), coaching and cheating distort the curriculum taught (Koretz, 2017; Hoyle and Wallace, 2007), thus the assessment is no longer able to function as a measure of educational attainment, as suggested in Campbell's law. Part of the distortion to learning caused by assessment under high-stakes pressure comes from assessment becoming performative, where the emphasis is on *showing* and *proving* learning has occurred at the expense of authentic teaching practices (Ball, 2012). As teacher and school performance is tied to KS2 SATs externally, performative practices encourage content and structure of SATs assessments to impact on what and how things are taught. This is referred to as 'wash-back' or 'back-wash' in research interchangeably. Gipps (2011: 27) refers to this as 'measurement-driven instruction' stating;

*Measurement-driven instruction is defined as when a high-stakes test, because of the important contingencies associated with students' performance, influences the instructional programme that prepares students for the test...*

Back-wash or wash-back have been associated with curriculum narrowing in KS2 to the 'core' subjects of maths, reading and writing at the expense of the rest of the curriculum (Black and Wiliam, 1998b; Torrance, 2011; Whetton, 2009; Wyse and Torrance, 2009). Even within core subjects, teaching time can be reallocated to specific curriculum content '...to better match the content of a specific test' Koretz (2008: 252) or aligning what is taught to specific test weighting, thus inflating test scores (Koretz 2008). For example, if shape and space only represent a handful of questions in a KS2 SATs maths paper, lesson content can be aligned to teaching number and calculation which represent most of the marks. Both strategies can result in invalid test scores, as the results are interpreted as sampling a wider domain than that tested. Stobart (2009: 168) provides an example of this for KS2 reading SATs where the assessment relies on a 'written response rather than their reading'.

Similarly, Tennent (2020) argues the test is not fit for purpose as it focuses on comprehension of unseen texts, which require a specific knowledge of the context which not all children will have access to. This literature is concerned with the effect coaching students has in improving test performance where inferences of improved student attainment, based on these scores, are not valid because they value construct-irrelevant criteria which enabled better performance (Koretz, 2008). Whetton (2009), although agreeing with Koretz (2008: 155) does attribute some benefit to assessment washback, stating that it 'has helped to give priority to particular subject elements from time to time'.

Wyse and Torrance (2009: 216) blame Neo-liberal aligned 'political pressure' for turning 'a potentially positive educational deployment of standard assessment tasks' into a 'national testing system' (see [Chapter 4](#)). The outcome being the creation of teachers who have learnt to coach children to pass tests, resulting in apparent improvements of test reform after their initial implementation (Torrance, 2011; Wyse and Torrance 2009).

It is difficult to remove the high-stakes element of Primary school assessment when it is normalised as necessary, and often viewed as the most economical way to hold schools to account, evident from the report by Bew (2011: 22) which stated high stakes testing needs to be used, as the Government 'believe that holding each school accountable externally is essential', and 'that 'high-stakes' accountability systems are the most cost-effective method for raising achievement' (Bew, 2011: 9). Emphasising the cost effectiveness of assessment

demonstrates where the priorities are for Government regarding Primary school SATs. Cost effectiveness does not equate to valid assessment interpretations especially viewed alongside concerns of high-stakes testing recommended in Bew's report; its potential narrowing of curriculum and over-coaching of children. However, the report was commissioned to review KS2 assessment within the confines of a test-based accountability system. Similar to Goldstein (2017), the report suggests that test-based accountability does not have to result in curriculum narrowing. However, the opposite conclusion is drawn in literature concerning educational assessment in England (Moss, 2017; Torrance, 2011; Whetton, 2009; Wyse and Torrance, 2009), or other cultures where high-stakes testing is prevalent (Linn, 2000; Diamond 2007; Berliner, 2011; Plank, 2013). Goldstein (2017) and Bew's (2011) argument fails to acknowledge how national testing, for example, year 1 phonics and year 4 times tables tests, have been introduced to improve standards in those areas, using tests to target specific curriculum areas. The financial argument of the cost effectiveness of SATs seems to be an issue of weighing up overall cost against the benefit for the school in informing accountability systems. This echoes concerns made by Black (1998), who did not think enough resources were put into developing teacher SA as recommended by TGAT (DfES 1988) as it was the more costly option.

One suggestion from the literature is to provide tests worth teaching to, to mitigate the issues caused from high-stakes testing (Baird et al., 2017; James 2017). These would be assessments which more accurately represented the concepts of desired learning. However, this has implications for accepting assessment as the driver for the curriculum received by the students and would not resolve the reliability and validity challenges above. Another suggestion is for school accountability to take a different form (Green, 2011).

#### *2.6.1.1.1.1 Impact on FA*

In addition to the issues raised above is that of how high-stakes SA impacts on formative assessment. High-stakes testing is argued to have a detrimental impact on FA, undermining effective implementation, and preventing time from adequately being used to improve its use (Black and Wiliam, 2003; James, 2017; Torrance, 2011;). Whetton (2009) concludes that FA is believed to better improve attainment rather than high-stakes testing. Recently in England, test based formative assessment has been recommended by Christodoulou (2017:

34) drawing on cognitive load theory, and the American learning theorist Hirsch, which contrasts with a more constructivist emphasis on the role of the learner.

#### 2.6.1.2 High-stakes assessment and accountability

Accountability is argued to be at the centre of why standardised tests are high-stakes (See [Chapter 4](#)). Bailey (2014: 664) links the rise of accountability with the rise of ‘a quality culture’ where ‘...Teachers are expected to “deliver” these results for their students...with the result that teacher competence is increasingly judged by student outcomes’. Although, these comments are made in regards to college teachers, they are echoed widely within the Primary sector. The concern with test-based accountability lies with the discourse it operates in, which proposes that the quality of teaching and learning can be captured and measured by standardised testing. Subsequently KS2 assessment data then forms the basis for comparisons of quality to be made between schools, creating an educational market place where student results are a commodity (See [Chapter 4.2](#)). Moss (2017: 63) refers specifically to the assessment of literacy in Primary schools, stating ‘...the trajectory to number-driven reform has left schools struggling with the difficulties that numbers on the public stage produce’, resulting in a technical process of accountability rather than a professional process where ‘...professionals are held accountable for the degree in which their actions meet certain standards’ (Biesta, 2017: 320-321). For Biesta (2017), this places students’ outcomes - test results – as the measure of success rather than the education of the pupils. Performative practices which encourage better test outcomes as a result are prioritised as argued above, resulting in a disjointedness between what high-stakes assessments indicate, and what they are used to signify (Baird et al., 2017). Biesta (2017) questions whether we have come to value what we *can* measure rather than measuring what we value. It is the over-use of data, he suggests, that is dominating our education system. In agreement with Biesta (2017), Moss (2017: 62) describes the current situation in which,

*...high-stakes testing remains. If anything, it has become more potent, placing more schools at risk of external intervention...instead of building the curriculum and then deciding how it can best be assessed, the assessment*

*tools themselves simply become the curriculum...Schools must teach children whatever the assessment asks of them.*

Thus, school success relies on test results of Yr6 pupils. Moss (2017) demonstrate how the number-driven marketised culture, where assessment data dominates school monitoring and encourages competition has come to fruition and is solidly instated into English education and assessment (Bew, 2011; DfES, 1988;).

It is not that the concept of accountability is viewed universally as wrong, literature critiquing high-stakes accountability based on testing is not typically criticising accountability itself, or high-stakes accountability, rather, the system based on testing requires further scrutiny (Black and Wiliam, 2018; Green 2011; Koretz 2008, 2017). Such a focus on test-based accountability diverts attention away from the impact it is having on student learning (Black and Wiliam, 2007, 2018). Allowing teachers to play a larger part in summative assessment is one suggestion to improve the link between what is being used to hold teachers and schools to account, and the validity of those measures (Black and Wiliam 2018).

## 2.6.2 Validity and reliability of teacher SA

Much of the published research explored supports the use of teachers' summative assessment and suggests that it should be more prominent in assessment practices (Black et al., 2011; Cox, 2008; Harlen, 2005; Stobart, 2009). Similar to testing, teacher SA is judged according to its validity and reliability (Bennet, 2011; Johnson, 2013).

### 2.6.2.1 Trust in teacher SA

Teacher SA is criticised mainly for its consistency. Teacher SA between schools is argued to be inconsistent, and within schools of being untrustworthy due to the risk of unconscious or conscious bias of the assessor (Gardner, 2010). A number of papers exploring teacher SA suggest that the persistence of statutory testing has created a view that teacher assessment cannot be trusted (Allal, 2013; Collins et al., 2010; Marlow et al., 2014). Collins et al.'s (2010) research into KS2 Science assessments, in the absence of compulsory SATs, found a quarter of Yr6 teachers did not feel confident in providing accurate assessments without a test. Yr6 teachers felt their own assessments were judged as unreliable by other teachers in their

setting, as well as in secondary schools which received them so they felt they had to use optional tests to validate their assessments. Even when highlighting the narrow scope of the data collected (in only one subject), this shows that in the absence of high-stakes testing most teachers in this study still opted for external testing, instead of trusting their own judgment. It was also found that in Wales, the removal of high-stakes testing produced very little change in teacher assessment practice in science. In a study of end of year report cards, Allal (2013) also found that teachers generally used external test data to inform their judgments, teachers only sought further evidence from their own assessment if they did not agree with the test score given. Bew (2011) supports this explaining that teachers felt their own assessments were given less value. A lack of trust in a teacher's own assessment could be due to a lack of investment in teacher SA skills, as indicated by Harlen (2005). Collins et al. (2010: 284) attribute the continued use of external tests to 'both teachers' own lack of confidence in their assessments...de-skilled over years of high-stakes testing, and...perceived demands of parents and local authorities for 'objective' evidence'. Similarly, Marlow et al. (2014: 414) found, 'formal assessments can be perceived as more accurate and objective than teachers' assessments' when comparing teacher SA to tests.

#### 2.6.2.2 Teacher SA bias

Johnson (2013) refers to unintentional bias as the 'halo effect' where teachers may approach assessment differently based on their values, stating this is well known to distort assessments by teachers. Bennet (2011) supports these statements by arguing that any teacher-based assessment inferences come with bias impacting on their reliability. To Harlen (2005), this does not mean teacher SA should be avoided, arguing that bias has to be taken into consideration so it can be acknowledged and compensated for. Johnson (2013) does not recommend this, suggesting instead that only tests with a determined reliability should be used for teacher assessment. There are two issues with this suggestion; the first is that it is unfeasible and practical to do. Measuring this is too costly and lengthy to implement (Baird and Black, 2013). There is no mention of what degree of reliability or validity tests should have, a consideration when creating a test that is deemed high in reliability as this would increase the test time significantly (Stobart, 2009; Wiliam, 2001). The second and more prominent problem with Johnson's criticism of teacher SA is that it is criticised in view of high-stakes assessment. This is where research agrees that teacher SA

should not be used for high-stakes testing (Black 1998; Wiliam, 2001; Harlen 2005, Baird and Black 2013; Stobart 2009; DfES 1988). This conclusion is drawn not because it is felt teacher assessment cannot be trusted, but that teacher SA has not been given the resources or support to develop, and that under high-stakes pressures it can easily be distorted. Along these lines, Bew (2011) does not recommend that teacher SA be used to assess children's' achievement in school because it is not suitable to hold schools to account in league tables. Marlow et al. (2014: 424) conclude their research by saying,

*Perhaps it is unrealistic to assume that teachers should be able to accurately assess intellectual functioning, especially given the significant other demands placed upon them. In this case, a review of the scope, function and purpose of teacher assessment would need to be completed.*

This is a worrying conclusion, and it will be interesting to see how changes linked with AwLs will impact a larger teaching issue of workload. Acknowledging this and the previous discussion surrounding the lack of support and resources put into evolving teacher SA since TGAT'S report (DES, 1988A), teacher SA in Primary schools may not be currently suited to perform as the sole type of SA at the end of KS2. Those who advocate, above, for the increased presence of teacher SA do not suggest it should replace tests, only that it should have an increased presence. Subjecting a teacher SA system to high-stakes accountability risks the validity and reliability of the assessment through the pressure for the school to be *seen* to succeed (Harlen, 2005). As Gipps (2011) cautions;

*If we move teacher assessment and performance assessment too closely towards standardization in order to satisfy traditional reliability, we are in danger of throwing out the baby with the bath water. (Gipps, 2011: 147)*

Similarly, Stobart (2009) feels there should be distance between test SA and teacher SA. He states that SATs tests do not assess the whole curriculum, therefore, if teachers are concerned with how their assessments measure up with the SATs test, they may only teach what they think will be assessed on the SATs tests. This suggests it is the high-stakes nature of assessments which impacts teacher SA rather than the practice itself.

### 2.6.2.3 Benefits of teacher SA

Teacher SA, when not subjected to the pressure of high-stakes assessment, arguably has many benefits. Stobart (2005) and Tennent (2020) recommend that teacher SA is given a bigger role within end of KS2 assessment so that the whole curriculum is represented.

Teacher SA mitigates the time issues associated with increasing the reliability of tests (William, 2001); has the potential to impact on the learning process through interaction with the student; can assess the whole curriculum and complement tests by assessing areas not suitable for written tests (Harlen and Gardner, 2010); and also reduces the pressure on the student of being assessed because of any test-based anxiety or loss in self-esteem (Harlen and Crick, 2003).

Where moderation processes and training take place, teacher SA could be as valid and reliable as standardised tests (Black et al., 2011; Gardner, 2010). Support for teacher assessment was a prominent finding by Collins et al. (2010: 283), but added 'that summative teacher assessment was subject to rigorous internal and external moderation informed by clearly articulated criteria' for it to be effective. Johnson (2013: 101) recommended this too, though cautiously stating 'consensus moderation can only prove its worth if formally evaluated' suggesting a formal moderation procedure and not one at the discretion of individual schools. As AwLs requires moderated teacher SA of writing, it will be interesting to see how this formal process of moderation is experienced by participants.

A more radical approach is suggested by James (2008), who wishes to better align assessment with socio-cultural learning theory. Teacher SA would be an essential part of assessment, envisaged as situated within the learning context, and captured using a variety of means, such as developing a portfolio of how available resources are used to work in a productive manner (James, 2017: 409). James (2017) is mindful of what interpretations may be based on this form of assessment, cautioning that judgments may be generalisable beyond the context they were formed in. For James, the teacher is best placed to assess their students, as this allows the assessment to better align with learning, rather than assessment backwash (James 2017; James and Lewis 2012).

#### 2.6.2.4 Use of comparative judgment

Moderation is argued above to be essential in developing teacher SA as an alternative to testing. Assessment without levels requires teacher SA to produce the final grade for writing, which was previously an externally marked writing test. This has raised concerns of inter-rater reliability of these assessments, as well as how time consuming and resource heavy external moderation is, so attempts have been made to introduce an alternative approach to writing assessment in the context of Primary writing using a comparative judgment method (Bramley and Vitello, 2019; van Daal et al., 2019).

*Comparative judgement, comparative pairs, pairwise comparison or paired comparison are terms that are used to describe a measurement method, which involves making inferences based on specific task criteria about one student's performance compared with another's. (Terricone and Newhouse, 2016: 2)*

To develop the inter-rater reliability of teacher assessment judgments, methods of comparative judgments (CJ) are being trialled by schools. Advocating for CJ, Terricone and Newhouse (2016) state advancements in computer technology should enable the widespread use of CJ for high-stakes assessment. Reliability is gained by many markers (commonly referred to as judges) making judgments, with computer programmes developed to identify outlying markers whose judgments are not consistent with others, and removing them from the process (Benton and Elliott, 2016; Terricone and Newhouse, 2016). This draws on Thurstone's (1927) 'Law of Comparative Judgment' which states that a comparison of two specimens has only two possible outcomes; that one is better and one is worse. It reasons that one person making a comparison is unreliable, therefore many comparisons on the same specimens by one person, or, one comparison each from a group of people make comparisons reliable. The specimens are not always physical but can be psychological and qualitative, as long as it is clear what dictates what is better and adapts what pieces are selected for comparisons of very similar pieces of work. Additional developments include adaptive comparative judgment (ACJ) developed by Pollitt (2012). As per Thurstone's law, no two comparisons can be equal, therefore very similar pieces of writing may be more difficult to judge. This method decreases the random matching of pairs

of writing scripts through advancing rounds of comparisons, subsequently argued to make the comparison more reliable. Applying ACJ, Pollitt (2012) says judgments can continue until the level of reliability desired has been met. However, Bramley and Vitello (2019) argue that using adaptive CJ inflates reliability scores making them more attractive as a marking resource above traditional marking, where reliability can also be increased with additional markers.

#### *2.6.2.4.1.1 Focus on consistency and inter-rater reliability at the cost of validity*

The literature deviates on what is needed to make the comparisons valid with the primary focus on reliability and inter-rater reliability specifically (Heldsinger and Humphry, 2013). Pollitt (2012: 292) states ACJ applied through their study was as valid as any other mark scheme because 'it relies on the judges keeping their minds constantly on the aim to maximise validity', which in this case was 'showing more evidence of what it means to be good at writing', referring to a short 160-word-length statement referencing English as a subject, not exclusively writing. Judges have also been found to differ on what they valued as good academic writing when using CJ, indicating that judgments were based on different standards. van Daal et al. (2019) found judges in their research could interpret narrow criteria differently and miss out construct-relevant criteria in making their judgment. As Thurstone's law states, it must be clear what will be used to judge one specimen as better. Thurstone (1927: 267) does caution the full applicability of the law:

*To transfer the reasoning in the same way from a single observer to a group of observers for specimens such as handwriting or English Composition is not so certain. For practical purposes it may be assumed that when a group of observers perceives a specimen of handwriting, the distribution of excellence that they read into the specimen is normal on the psychological continuum of perceived excellence. At least this is a safe assumption if the group is not split in some curious way with prejudices for or against particular elements of the specimen.*

Thurstone suggests here that a group of observers can be 'assumed' to be all perceiving excellence as the same unless there is a 'split in some curious way'. This does not

demonstrate a sense of certainty or reliability as discussed in the recent literature implying its use for the assessment of writing.

Benton and Elliott (2016) defend the use of CJ based on reliability and consistency, expressing that even though individual expert judges find it difficult to separate two similar pieces of work, this should not be a barrier, as it is not only one judgment at play but many. In researching the reliability of CJ, Benton and Elliott (2016) applied a formula they developed to existing literature. They aimed to predict the reliability of expert judgments to improve comparative judgment as an assessment tool. In doing so, Benton and Elliott (2016: 371) argued that not only can two pieces of work be compared against each other but that a 'direct comparison with benchmark scripts' could also be used successfully, removing the need for expert judges. It is important to state that they use a broad definition of 'expert', where it is unclear what an expert marker for writing is, for example, an experienced writer or an experienced KS2 teacher. This is interesting because it draws attention to the importance, or lack, of subject knowledge, and makes it unclear what each individual's priorities might be when making a judgment. As van Daal et al. (2016: 13) found, that while the majority of judges in their study justified their reasons in relation to 'construct-relevant arguments', variation was found where some judges' reasons did not corroborate with judgments given. This brings into question the bases for reliable scores found by Benton and Elliot (2016), because it is not clear why or how these judgments are reliable, only that they are consistent. Inter-rater reliability is being used as the judgment of reliability rather than intra-rater reliability of the judges, or how judges form expertise, or how assessment knowledge impacts on validity.

What is interesting is the potential effects of filtering out differing or outlying judges. Potentially an outlying judge may have more valid judgments based on either better subject knowledge or experience with the year and stage being assessed, but their judgments would be removed because they do not fit with the general trend of markers. Humphry and McGrane (2015: 445) argued doing this means 'there is little or no scope for marker harshness (or lenience)' suggesting the issue of teacher bias in could be eliminated through their comparison of CJ alongside standardised rubric marking of writing. However, the participants in this study, were experienced markers using the national assessment rubric,

and conducted the CJ 'immediately after having marked performances using the rubric.' (Humphry and McGrane: 449). The high-agreement found between rubric marking and CJ could be because of their well-established knowledge of the rubric rather than a benefit of CJ (Humphry and McGrane 2015). In fact, Verhavert et al. (2019) found the more experienced the judge the fewer repetitions of comparisons are needed, where novice judges required more comparisons to achieve a high level of inter-rater reliability, and an expert assessor was recommended to achieve the most out of CJ. In this view the experience and knowledge of the assessor is key to reliability of CJ and not just a concern of validity. This raises a question as to whether this can be replicated by less experienced teachers in English schools, as is suggested by Cambridge assessment with 'No More Marking' (2020).

#### *2.6.2.4.1.2 CJ to supplement rather than replace traditional SA*

Wheadon et al. (2019) strongly recommends CJ for standardised assessment of writing because of low comparable cost to moderation and use of in-service teachers as judges. They feel it will mitigate the issues of rubric driven writing assessments and the washback which can occur encouraging more mechanistic writing, lacking in fluency and engagement. However, they do caution against the use of CJ for high-stakes assessment because of pressures that are associated and the possibility of score inflation. This was a real concern in the literature with CJ recommended as a supplement for the current national writing assessments in place (Heldsinger and Humphrys, 2010). This recommendation was not just based on the pressures associated with high-stakes testing but also the workload teachers may experience, especially in older Primary year groups where writing scripts can be lengthy (Humphry, 2013; Wheadon et al. 2019).

## 2.7 RESEARCH INTO ASSESSMENT WITHOUT LEVELS IN PRIMARY SCHOOLS

The literature review sections so far have considered theoretical and contextual considerations of Primary Assessment as a whole and will now explore the research reported on so far specifically with reference to Assessment without Levels (AwLs). Some of this research has already been referenced as it contributes to the wider picture of Primary assessment.

When researching the official removal date of Assessment with Levels (NCLs) a report by McIntosh (2015), 'Commission on assessment without levels' (CAWL, 2015) was released. This provided a preliminary look at how schools had already begun to construct their assessment policy without levels, with the aim to communicate 'a clear understanding of the purposes and principles of assessment' (CAWL, 2015: 10) against the backdrop of a critique of levels, with their removal providing an opportunity for schools to reshape assessment policy in schools, but with responsibility to consider the 'purposes and principles' outlined in the report.

Concern was raised over evidence that some schools had attempted to recreate the structure of levels with the new curriculum, attributing this to the 'conditioning' levels had caused. They reasoned that school assessment cultures had to change for AwLs, and the Curriculum from 2014 to improve educational standards. Part of this culture McIntosh (2015) attributed to how linked NCLs was to school accountability, i.e. KS2 SATs. They acknowledged that the original concept of levels was not to influence assessment throughout schools, its function was to evaluate learning at the end of key stages only. Despite this;

*...the pressure generated by the use of levels in the accountability system led to a curriculum driven by Attainment Targets, levels and sub-levels, rather than the programmes of study. (CAWL, 2015: 12)*

NCLs, from this analysis, focused on attainment targets instead of the curriculum encouraging teaching to the test (TTTT) practices to prioritising progress through the level descriptors which spanned all Primary years. Similarly, Alderton and Pratt (2021) found all 11 schools in their study purchased tracking software for maths which focused on NC attainment targets for each school year and teaching according to the 'gaps' identified.

Their concern was that this is producing teachers who are ‘little more than the ‘fillers of mathematical gaps’’ (Alderton and Pratt, 2021: 12). There is much agreement in the literature of this view that NCLs was subject to pressures which encouraged TTTT but it was not felt that changing and reforming Primary assessment will prevent this in the future. Rather than improving the impact of test-based accountability on teaching.

### 2.7.1 AwLs and high-stakes testing

Moss (2017) contends the high-stakes testing in Primary assessment have become more precarious, as the removal of levels has meant the removal of resources and support for in school assessment, essentially moving the goal posts and taking away the map helping schools to reach them.

What McIntosh (2015) fails to address is the argument that TTTT is influenced significantly by accountability pressures, rather than the language and criteria used to assess pupils. Ward and Quennerstedt (2019) found such impact in a recent study researching the influence of SATs on maths teaching in Primary school. Accountability pressures from SATs, the senior leadership team (SLT) and OfSTED meant learning was prioritised if it was able to be evidenced and measured, resulting in a loss of creative aspects of maths and a privileging of test based in-School assessments. SATs pressure had become self-governing under the evidencing pressures of the SLT and OfSTED. This study indicates that even with levels gone, pressures are still felt from statutory assessment mediated through SLT, favouring that which can be measured in SATs. Similarly, Pratt and Alderton (2017) studying Primary teachers maths assessment argued assessment practices may have changed, but they operate within the same discourses as they did with levels. For example; NCLs was criticised for encouraging the selective teaching of borderline students of the expected level for that year (Ball et al., 2012). In their study, Pratt and Alderton (2017) saw a continuation of this, with teacher attention paid to pupils on the border of the expected standard. Assessments were seen as less formative, and instead were used performatively to target borderline children. In further exploration of educational discourses in Primary maths assessment, Pratt and Alderton (2019) studied the way teachers have formed new assessment ‘truths’ in the absence of levels. Similar again to Ward and Quennerstedt (2019), Pratt and Alderton (2019: 590) found an increased reliance on testing for teacher SA attributed to unease felt from a loss of control over assessment.

*largely validated through testing and normalised – to define what is (ir)relevant, (in)essential and central/marginal to schools' practice...This provides teachers with a (reconstructed) discourse of control, allowing them to participate again in taking responsibility for pupils' learning and to merit their performance as teachers.*

The continued presence and pressure of test-based accountability with AwLs, and increased performance pressure on teachers now it is related to pay, has resulted in a reconstructed understanding of assessment in terms of testing and what is viewed as 'normal' for a student to achieve (Pratt and Alderton 2019).

Teacher confidence in developing their own SA was seen a concern in a more recent commissioned review of AwLs in Poet et al. (2018). This focused on qualitative data gathered on non-statutory assessment practices in schools. Among their main findings they reported:

- Teacher confidence was mixed;
- Some reporting of increase workload;
- Primary schools used externally purchased tests;
- core subjects the main focus of AwLs in all years.

The use of externally purchased tests was selected by SLT to allow for comparability with other schools nationally, as well as to inform predicted grades for end of KS2. Primary teachers especially felt tests were relied on because their SA was not trusted;

*...some teachers felt that their school did not trust them to determine whether their pupils were meeting age-related expectations without the use of standardised tests...[they] were concerned that this lack of trust had led to an increase in formal testing to provide the school with nationally standardised and comparable summative assessments of pupil performance.... (Poet et al., 2018: 37)*

Teacher confidence in their SA was questioned by Williams (2015), who used teacher perceptions of confidence assessing and developing their own test materials to judge the effectiveness of teachers to assess students, echoing findings from Poet et al. (2018). A

reliance on assessment schemes replicating SATs does not fulfil the hopes of McIntosh (2015), that school autonomy on non-statutory assessment would divert the effects of accountability pressures from statutory tests away from in-School assessment. Poet et al. (2018) further reported that collaboration with other schools and the use of online assessment resources meant that little further guidance was required by the government other than exemplars.

These recent studies into AwLs in Primary school question the effectiveness of removing levels to decrease assessment back-wash/wash-back, such as TTTT, when Primary schools are still largely held to account with KS2 SATs (Priestley et al., 2015b; Wilkins 2015).

## 2.8 CONCLUSION

In concluding this review of the literature, it is clear that formative assessment is seen as the most effective form of assessment to improve learning. However, for accountability purposes summative assessment, specifically test based, is the priority. This has resulted in a tension where the high-stakes nature of KS2 SATs, teaching to the test strategies, performative behaviours, and an increase in managerialism, conflicts with developments in learning theory which encourage a constructivist pedagogy which is at odds with the behaviouralist pedagogy testing encourages.

Where SA is concerned, the literature overwhelmingly supports the use of teacher SA rather than testing (Allal, 2013; Collins et al., 2010; Harlen, 2005; Marlow et al., 2014; Stobart 2009) but with caution in high-stakes situations where the multiple interpretations on the summative judgments compromise validity. Before the transition into AwLs, KS2 SATs contained no teacher SA judgments. It is very concerning that teacher SA has been underutilised and undervalued for KS2 SATs, but it is also promising that the assessment of writing without levels will now include teacher SA despite the introduction of a grammar test. As Torrance (2011: 459) argued a new curriculum and assessment model was 'urgently needed'. Whetton (2009: 156) said 'it cannot be certain that any new system will be necessarily better and without its own unforeseen consequences'. Whetton's (2009) comment shows how vital this research into AwLs is.

In reviewing the literature, it is apparent but understandable that there is not much research exploring the impact of AwLs on teachers' teaching and assessment practices because of how recent the reform is. This is an emerging field of research and requires building upon from the studies which have already been published. My research will address the gaps identified in primary assessment literature as set out in my research aims to provide:

- An analysis of how AwLs has changed assessment practices in schools from the point of view of teachers who teach in SATs years.
- Primary school teacher views based on their experience of AwLs so far.
- An analysis of the overall impact Assessment without Levels has had on Primary school teachers.

In fulfilling these aims the research will explore both the impact of AwLs on teaching and assessment practice and the policy into practice implications of education policy, challenging the assumptions made that statutory testing captures the learning intended. In doing so this research hopes to identify whether it will be the change Torrance (2011) was hoping for is or a more pessimistic view of Whetton (2009), that each system will come with its own problems.

The next two chapters will consider the political element of AwLs. The first will review the transition in education policy from NCL to AwLs while addressing the reasons given for the change. The second will concentrate on the neoliberal ideology which underpins education policy and impacts on assessment in schools.

## 3 EDUCATION POLICY

---

### 3.1 INTRODUCTION

This chapter will briefly explore government policy reform from 1988 to 2010, picking up on key themes and arguments, before discussing the recent education Policy reform from the Coalition Government of 2010. Attention will be given to changes in Education Policy which provide a context for the AwLs reform, to be viewed in. New Labour Education Policy, for instance, lay the groundwork for the Coalition Education Policy reform emphasising the importance of KS2 SATs as a measure of school quality.

Both assessment and curriculum policy will be included in the discussion, as *what* is being assessed as well as *how* it is assessed impacts on teachers and schools. As such, this chapter will provide detail of statutory assessment requirements present during Thatcher's Conservative Government and New Labour's reforms. New Labour's reforms will be described with reference to the National Strategies, criticisms of which were used as part of the justification for the 2010 Coalition Education reforms.

International comparisons of educational standards are identified as a major feature in Coalition Education Policy and their justification for reform. England is viewed as falling behind countries like Singapore and South Korea, who score well on the Organisation for Economic Cooperation and Development's (OECD's) Programme for International Student Assessment (PISA) tests. A critique is then provided of how the Coalition used recommendations and guidance provided from an assessment and curriculum review.

### 3.2 POLICY AS AN IMPACT ON EDUCATION

Education Policy exemplifies strategies expected to be followed by schools. In England, policy makers are responsible for defining what is taught through the National Curriculum and monitoring student and school progress through statutory testing. However policy can also be seen as responsible for '...inciting particular kinds of thinking and action' (Schmidt, 2017: 12), an impact of policy outside of a set of rules for schools to follow, but additionally as way of shaping teaching practice through their mediation of policy. Perryman et al.

(2017), argue that in implementing policy requirements, policy becomes part of the teachers' identity, shaping how they define themselves as a 'good' teacher. They argue:

*Teachers in effect become policy...through 'interest' and 'curiosity' to improve themselves, become a better teacher, a 'good' teacher. (Perryman et al., 2017: 754)*

Importantly, policy is not viewed by all as an absolute influence on teachers' practice; how teachers experience policy is mediated in individual schools through '...filtering out and selective focusing done by Head teachers and their SLTs...' (Ball et al., 2011: 626).

Accordingly, both teachers and schools themselves are an essential part of policy as it is enacted through them and mediated by their values, beliefs and experiences (Buchanan, 2015). The result of Education policy reform is then dependent on *how* it is enacted by schools and teachers as well as the expectations it exemplifies. Therefore, when considering education policy reform that removed levels, policy requires examining on two levels, i.e., policy at the 'macro' and 'micro' level. The first being the policy itself, which is analysed in this chapter. The second is educational policy at the school level which will be analysed through a Neo-Liberal lens in [Chapter 4](#) and that of teacher agency and identity in [Chapter 5](#).

### 3.3 OVERVIEW OF EDUCATION POLICY BETWEEN 1988-2010

Education was the main driving point of New Labour's election campaign, argued as the essential component for a successful 'knowledge economy' and society. Under their motto of 'pressure and support' education was an investment in 'human capital' to compete against other countries in a global market (DFES, 1997). Support came in the form of the National Strategies comprising APP guidance and pedagogy support. Past and preceding governments have intervened in curriculum but not pedagogy. For example, the National Strategies required the setting of ability groups, as well as 1hr each day earmarked for literacy and numeracy. Likewise, APP was non-statutory KS1 and KS2 assessment expectations. These aspects of the National Strategies were criticised for restricting schools to a set pedagogy and non-statutory assessment framework and underpinned arguments to increase autonomy for school assessment. Pressure from New Labour reforms accompanied support with an emphasis on 'rigorous assessment and testing at ages seven and 11.' (DCFS,

1997: 22), viewed as a measure of and encouragement for improvements in literacy and numeracy standards. Notably this placed accountability with teachers and schools through enhanced publishing of performance (SATs) data. New Labour's consecutive education reforms communicated the message that improvements in standards needed to be maximised and predictable. Thus, enhancing the publication of SATs data was argued to benefit parents as 'consumers' and teachers through the use of this data to set 'clear targets' and 'evaluate their performance' (DfES, 2001: 6-8).

New Labour gave previous Conservative market reforms 'meat and teeth' (Exley and Ball, 2014: 22) and ratcheted up the use of marketised and managerial language in educational Policy;

*...with the consequence that what may have once seemed impossible or simply unthinkable becomes sensible, obvious and inevitable... (Bailey and Ball, 2016: 128)*

However, unlike the Conservative governments of Thatcher before, Labour invested money in schools (Exley and Ball, 2014; Furlong, 2013), introduced National Literacy and Numeracy Strategies, and provided schools with extra funding for low-income pupils (this was the 'pupil premium'). This investment is rationalised by Furlong (2013) to account for the lack of industrial action. By the time the Coalition government came to power in 2010, parents had been positioned as consumers and SATs results as the measure of teacher and school quality had become normalised (Bailey and Ball, 2016).

### 3.3.1 Success/failure of New Labour reforms

KS2 SATs test data failed to demonstrate improvements set by 'ambitious' targets of 75% of students in maths and 80% in literacy reaching level 4 by 2005. Their reliance on SATs data as a measure of educational standards in Primary schools demonstrates the establishment of teachers' and schools' performances being explicitly identified as the key variable in meeting targets. Comparatively, the plateauing of KS2 SATs data following the initial increase is similar to plateaus discussed by Koretz (2008; 2017), where educators become familiar with how subject content is tested, therefore quick gains are seen first, followed by plateauing of results. This suggests the plateau in SATs data was symptomatic of standardised testing, rather than the ineffectiveness of the National Strategies.

Upon the archiving of the National Strategies (NS), making them non-mandatory guidance for schools, a commissioned review of them used KS2 test data to evidence improvements in literacy and numeracy. This described the NS not only as contributing towards overall improvements achieving level 4, but also in closing the 'poverty gap' (DfE, 2011a). Additionally, the review drew on, Trends in International Mathematics and Science Study (TIMSS, 2007), stating children in England outperformed other English-speaking nations such as Australia and USA in mathematics. Despite these positives reported, none were included in the Coalition Government's education White Paper 'The Importance of Teaching' (DfE, 2010b) or 'The Case for Change' (DfE, 2010a).

The next section lays out the rationale for the changes made to education policy in England from the Coalition Government 2010 contributing to the removal of levels in Primary Assessment.

### 3.4 CHANGE OF GOVERNMENT, CHANGE OF POLICY...?

Although education did not feature largely on the political agenda for the 2010 election (Bailey and Ball, 2016), when in office the Coalition Government commenced a review of education policy and archived curriculum and assessment guidance from New Labour. The Coalition Government's report 'The Case for Change' (DfE, 2010a: 5) contained their argument for education reform, specifically the need for education to support the economy, stating 'Those [countries] that have stood still have now found themselves surpassed by countries which, having out-taught them, are now out-growing them economically.' Education in England was not felt to have progressed and was failing to provide a workforce necessary for a changed global economy. This characterised education as the linchpin of economic success, echoing a message increasingly found in New Labour's successive reforms. Concurrently, the use of pupil performance data was linked to parental choice. The Coalition government of 2010 followed the Case for Change (DfE, 2010a) with their White paper 'The Importance of Teaching (DfE, 2010b). The Prime Minister and Deputy Prime Minister focused strongly on the results of the Programme for International Student Assessment (PISA) tests as an indicator of the English education system's inability to compete highly with other countries.

### 3.4.1 Justification for Policy change

Two key features of education practises in the countries discussed as surpassing England were; high accountability, and autonomy measures for schools (DfE, 2010a; DfE 2010b). In adopting these approaches to education, Michael Gove, Secretary of State for education at the time, argued education ‘...allows us all to become authors of our own life stories’ and therefore will allow children to ‘...overcome accidents of birth and background to achieve much more than they may ever have imagined.’ (DfE, 2010b: 6). However, these conclusions are drawn without fully considering the different context of accountability and autonomy in England, compared to countries referenced. For example, recommendations for higher autonomy were based on Singapore’s success in PISA.

*In the most recent OECD PISA survey in 2006 we fell from 4<sup>th</sup>...to 14<sup>th</sup> in science, 7<sup>th</sup> to 17<sup>th</sup> in literacy, and 8<sup>th</sup> to 24<sup>th</sup> in the world in the 2000s in mathematics. The only way we can catch up, and have the world-class schools our children deserve, is by learning the lessons of other countries’ success.  
(DfE, 2010b: 3)*

It was also stated that improvements in Singapore started from higher government intervention which only followed with higher autonomy for schools when standards had improved. A concern of the Coalition government was that England’s education system was falling behind, therefore, it could be argued that unlike in Singapore, education standards were not good enough to warrant less state intervention, as ‘In Singapore, reform focused first on securing sufficient places and then focused on quality through a highly centrally-directed process, including prescription of lesson content and teaching.’ (DfE, 2010a: 4), this aspect of the reform seems ignored.

Differences in accountability measures between countries, as well as country size, were not considered. Applying test based external accountability based on these justifications ignores how countries such as Singapore, small in comparison to other countries, and South Korea differ in their use of high-stakes testing, which are mostly used for university entrance exams; cultural attitudes to these test result in additional tuition, at extra cost to parents and differing attitudes to learning (Koretz, 2017). Alongside Singapore, Finland was cited as an example of where higher autonomy and accountability is responsible for educational improvement.

*Finland and South Korea – the highest performing countries in PISA – have clearly defined and challenging universal standards, along with individual school autonomy. (DfE, 2010b: 3-4)*

However, the Finnish education system does not use external accountability measures at the Primary school level (Webb et al., 2009). These considerations are not explored, rather, the message of high accountability and higher autonomy was given without much debate of each country's situated context

There was additionally an absence of discussion of consequences associated with a reliance on test-based accountability, only a statement that the new government wished to reduce the amount of curriculum narrowing and test rehearsal (DfE, 2010a: 23). The intended and unintended consequences are described in more detail in previous [Chapters 2.5 and 2.6](#). In short, test-based external accountability results in positives of working harder and better, but also encourages teaching to the test (TTTT), test coaching and cheating (Koretz, 2008).

### 3.4.2 Curriculum and assessment commissioned reviews

The publications 'Case for Change' (DfE 2010a) and 'The Importance of Teaching' (DfE, 2010) outlined the intent for a review of the curriculum and assessment, however, strong discourses of accountability, autonomy, and neoliberal ideologies of social improvement, impacted these reviews on curriculum and assessment.

The two reviews commissioned were the Bew Review of Key Stage 2 testing (2011), and an expert panel to create recommendation on The Framework for the National Curriculum (2011).

#### 3.4.2.1 The Bew review

The Bew review (2011) was commissioned to review assessment and make recommendations within the remit of certain conditions. Discourses of high autonomy and high externally imposed accountability were set out from the start, with recommendations to be made within this context. The removals of NCLs, as an assessment tool, had also been proposed by the Coalition Government (DfE 2010a; 2010b)

*The Secretary of State has therefore been clear that school autonomy must be accompanied by robust accountability. (Bew, 2011: 4)*

The report acknowledges concerns from The Importance of Teaching (DfE, 2010b) of unintended consequences of TTTT and shows concern that external statutory testing (KS2 SATs) was too high-stakes employing levels (Bew, 2011). The use of KS2 SATs data to inform league tables was viewed as 'crude', with the report suggesting that widening the evidence used may reduce the SATs associated pressure;

***...a greater range of published information will reduce the likelihood that league tables will be created focused on one indicator alone. (Bew, 2011: 12 emphasis in original).***

The report recommends teacher SA to bear more weight in KS2 SATs, feeling teachers were best placed to assess their students, and that this will reduce accountability pressures on one measure of success- the KS2 SATs.

*We would like to see a greater emphasis on teacher assessment within statutory assessment, and summative teacher assessment to be given greater weight within the accountability system. (Bew 2011: 9)*

The Bew Review (2011) concluded that concern over the suitability of levels was sufficient to consider a different assessment scale as part of the National Curriculum review.

### 3.4.3 The National Curriculum review (2011) expert panel

The National Curriculum Review (NCR, 2011) panel agreed with recommendations made by Bew (2011) that assessment with levels needed to change. NCR (2011) recommended 'ready to progress' assessment, focusing on fewer areas in more depth.

*Amongst the international systems which we have examined, there are several that appear to focus on fewer things in greater depth in Primary education, and pay particular attention to all pupils having an adequate understanding of these key elements prior to moving to the next body of content – they are 'ready to progress'. (NCR, 2011: 45)*

These recommendations, although framed similarly within international comparisons of practice, explored more fully the context of each country stating that universally '...there is no clear trend within high-performing jurisdictions.' (NCR, 2011: 45-46). Just as Koretz (2008) had argued that culture and tradition are of significant importance to attitudes and

preparation for testing, the NCR panel drew attention to factors outside of the school unique to each culture and society. The report found causal elements of teaching and assessment practice which they felt added to the debate for educational reform in England, these included: catch up support, holding the group together and small group teaching to bring students up to attainment of the rest of the year group (NCR, 2011: 46-47).

From their analysis the NCR (2011) made the recommendation to move away from a 'best fit' model of summative assessment and instead employ a formative tracking of achievement (NCR, 2011). To encourage high expectations for all, learning needed to focus on the mastery of a core set of learning with an assessment at the end of a key stage to decide if the children were ready to progress to the next (NCR, 2011: 47).

It is important to note that there were tensions between some members of the panel and the secretary of state for education Michael Gove. Letters provide a dialogue of tensions caused by: differing contract lengths of experts, a short time scale for completion of the review, as well as differing views of its direction. Conflicts in viewpoints were anticipated by Mary James, who in advance of working on the panel wrote to Nick Gibb, Minister for schools at the time, to set out her position on education and curriculum. Subsequently, James and Pollard wrote a resignation letter dated 10<sup>th</sup> October 2011 to Michael Gove, with major concerns that they were being bypassed and that changes, unsupported by the evidence from the panel, were made to the draft PoS (Program of Study) without their consultation. A subsequent letter followed on 20<sup>th</sup> October 2011, withdrawing their resignation after meeting Gove. This letter set out James and Pollard's understanding of their continuing contribution to the report. The priority for James and Pollard, from this letter, was the completion of the report, however, this was with the understanding that as Gove stated;

*...ultimately decisions about the form of the new National Curriculum is for ministers to take, and I do not promise to endorse or reject any of the panels specific proposals until I have had an opportunity to hear a wider range of views on the issue.*

The resulting constructed curriculum therefore did not necessarily reflect the recommendations from Bew (2011) or the NCR (2011) panel.

### 3.4.4 Review recommendations taken on or rejected

There is a question mark over the usefulness of the 'independent' reviews commissioned on curriculum and assessment, when many of the main themes within the new policy can be located in Conservative Party Documents such as the 2007 Green paper and The Case for Change (2010a). Recommendations were ignored where they differed from the view already established by the Government or had been formed based on hasty adoption of certain aspects of international assessment policy. For example, in 2017, The Primary Assessment in England, Government Consultation Response stated they would be, '...removing the statutory requirement to carry out teacher assessment in reading and mathematics at the end of key stage 2' (DfE, 2017: 29). The Bew Review (2011), even when framing their recommendations within the context of high accountability and high autonomy, recommended an increase inclusion of teacher assessment at the end of KS2.

#### 3.4.4.1 Same evidence, different interpretations

The interpretation of evidence differs between the NCR (2011) and policy. For example, both the expert panel report and The Case for Change (DfE, 2010a) draw on international comparisons for curriculum and assessment recommendations, however, how these comparisons are interpreted seem to differ. The proposal of the ready to progress assessment and high expectations of achievement by NCR (2011), was to ensure expectations are not limited for children and that they are ready to build on what they have learnt.

*The emphasis on effort is particularly marked in the Confucian-heritage countries such as China, Hong Kong SAR, Singapore, South Korea and Taiwan... In Western countries, especially in the US and England, the assumption has often been that capacity to learn...is determined by innate endowment of fixed intelligence (ability). This assumption...has had negative influence on expectations of achievement and how learning and assessment is organised. (NCR, 2011: 8.6)*

NCR (2011) recommendation of higher educational expectations was based on comparisons from cultures who favoured effort rather than a belief of a set intelligence. The expert panel focused on adapting the mind-set of teaching and learning to ensure expectations were not

limited for all students, whereas the government raised expectations in the Primary Curriculum to encourage high expectations of learners. In practice, this raises the difficulty of SATs rather than effecting deeper cultural change in attitudes towards learning, and risks continuing the TTTT practices seen with levels.

In a speech at a conference for school leaders, James discussed the tensions between political incentives for educational achievement and professional ones. She cites a difference of education as economic imperative compared to education as an ethical moral imperative (James, 2015). These differing motives contributed to the divergence between expert recommendations and political economic imperatives.

### 3.5 CONCLUSION

This chapter has provided a brief timeline of education reforms pertaining to assessment since the introduction of the National Curriculum in 1988 up to the removal of NCLs in 2010. It has presented a critical analysis of the 2010, Conservative and Liberal Democrat Coalition, education reforms. Specifically, attention has been drawn to the large focus on assessment comprising of high autonomy for non-statutory assessment and high accountability reliance on KS2 SATs results. The AwLs reform has been noted to have been formed within the bounds of these foci. The next chapter will examine the literature on Neoliberalism and how education policy is influenced by dominant marketised and managerial discourses and the impact this has on teachers' assessment and teaching practices.

## 4 NEOLIBERALISM AND EDUCATION POLICY

---

English education Policy has not escaped a wider influence of neoliberal ideology, evident since the 1980s. As an ideology, its impact is said by Metcalf (2017) to not only encourage pro-market agendas for government policies, but also as something engrained in the day-to-day regulation of peoples' choices.

*In short, "neoliberalism" is not simply a name for pro-market policies...It is a name for a premise that, quietly, has come to regulate all we practice and believe: that competition is the only legitimate organising principle for human activity. (Metcalf, 2017)*

Neoliberal ideology manifests itself in two forms. The first being the wider free market effects of policy, with the increased use of KS2 SATs data to rank schools. This is seen as a direct result of competition created by neoliberal reforms over the last 30-40 years (Ball 1998; Davies and Bansel 2007; Exley and Ball 2014; Furlong 2013). The second being a 'quiet' self-governing influence, for example, an emphasis on individual teacher accountability for student attainment with personal responsibility for improving student performance (Ball 2012; Metcalf 2017; Ong 2007). Ong (2007: 4) refers to these two aspects of neoliberalism as neoliberal with a big N;

*...a unified state apparatus totally dedicated to the interests of unregulated markets*

And neoliberal with a small n being;

*...a mode of 'governing through freedom' that requires people to be free and self managing (Ong, 2007: 4)*

In the context of AwLs Policy in primary schools the big N can be applied to competition encouraged between schools for pupil places, creating a market where KS2 SATs grades are a commodity. Neoliberal with a small n can be applied to how accountable teachers are for their practice as measured through their students grades. During the course of this chapter this duality of neoliberalism will be used to contextualise the impact of neoliberal ideologies on assessment practices in primary schools.

#### 4.1 A BRIEF HISTORY OF NEOLIBERAL EDUCATION REFORM

Neoliberalism as an ideology is not new. In the 1930s Friedrich Hayek viewed the world through the lens of the economic market, where the free market would protect from a totalitarian state. The responsibility of the state was to ensure the market remained free providing a '...fixed, neutral, universal legal framework within which market forces operate spontaneously' (Metcalf, 2017). Hayek's neoliberal system required all aspects of democracy to have economic justifications, thus political action is linked to the system leading to choices made based on economic benefit, rather than discussion and deliberation (Metcalf, 2017). Metcalf argues this created a system based on people as consumers and sellers, with an emphasis on the product being produced, for example, a work force educated sufficiently to fulfil the needs of the job market. Neoliberalism is not to be confused with a *laissez faire* ideology of minimal interference, rather the state:

*...becomes the protector of capital and its role is reduced to the enactment of monetary, fiscal, social, and educational policies to nourish and protect the market.'* (Baltodano, 2012: 493)

Much literature strongly focuses on the Reagan/Thatcher era of the 70s and 80s as the point where policies in education (and other public sector areas) began to withdraw control from the state, and instead created market driven competitive systems (Ball, 1998; Davies and Bansel, 2007; Exley and Ball, 2014; Furlong, 2013). Aside from the introduction of a national curriculum, control would not come from the government to raise and maintain standards, instead it would be the allocation of money creating a market within the education system (Exley and Ball 2014).

Davies and Bansel (2007: 248) describe this shift as,

*...the transformation of the administrative state, one previously responsible for human well-being, as well as for the economy, into a state that gives power to global corporations and installs apparatuses and knowledges through which people are reconfigured as productive economic entrepreneurs of their own lives.*

The implication of the Davies and Bansel (2007) quote above is that it has become internalised for teachers and schools to focus on the larger assessment driven market of

Primary education, where individual teachers are 'reconfigured' to maximise the test outcomes of their students for the good of themselves and their school. In Governments adopting this position it is argued a system is created where '[i]nequalities are inevitable, necessary but unintentional' (Exley and Ball, 2014: 13) because of the emphasis on what can be quantified and measured. These means it is inevitable that some schools will attract full classes of students just as others will not resulting in funding inequities between different schools.

## 4.2 MARKETISATION OF EDUCATION, EDUCATION AS A COMMODITY

A key lever of the neoliberal ideology is that a competitive market can and should regulate itself. Where a market is not present one should be created, evidenced in a higher focus on accountability, privatisation (including a push for academies and free schools) and competition to drive improvement in standards. As discussed earlier (See [Chapter 3](#)), the economic benefits of education obtained through a skilled workforce has been a re-occurring theme in education policy since the 1980s (Exley and Ball 2014; Furlong 2013; Whitty and Wisby, 2016). However, in recent educational reforms, in light of an increased emphasis of international competition (DfE, 2010a), this view of education as a product is even more prominent.

Following the Labour government, the Conservative - Liberal Democrat Coalition reforms continued to marketise education by reallocating funding powers to schools, enabling and requiring them to make more choices in a wider education resource market (Kumar and Hill, 2012: 18). Specifically regarding assessment, the removal or reduction in teacher assessment guidance and frameworks created a resource market for assessment products. Pratt (2016) explains this in terms of neoliberal approaches in assessment practice which are the dominant discourse in schools defining what is and is not thinkable. For example, research by Pratt and Alderton (2019) found assessment resources which supported SATs preparation were preferred by the schools they studied (Pratt and Alderton, 2019). This reflects views by Stevenson (2010: 342) who describes an education environment where '...the language of markets, targets and tests is not only increasingly regulating education, but is driving out the possibility of other languages and closing the educational field to other

possibilities'. Exley and Ball (2014) similarly describe a normalisation of market agendas in education contributed to by Labour's 'ratcheting' of neoliberal reforms.

#### 4.2.1 Choice

The education market is situated as promoting choice and diversity within education to drive up standards. Providing parents with the choice of schools was argued by Labour to ensure value from education '...both in terms of capital and revenue spending' (DfES, 2005: 113). School league tables, informed by KS2 SATs, form the bases for parental choice on schools, assuming that higher SATs scores equate to better school quality. This in itself is problematic as education viewed through this market lens assigns a financial value from government allocation of funding in exchange for the 'product' of said education. However, the concept of what Kumar and Hill (2012: 2-3) referred to as '...so-called "parental choice" of a diverse range of schools...' can be contested. Choice is argued to be an illusion unless parents have the wealth to truly choose where their children are educated, inequalities are thus perpetuated as those parents who can, will, pay for 'better' education, or move house to gain access to 'better' schools, therefore 'playing the system' in a way less wealthy parents cannot (Ball et al., 1996; Kumar and Hill, 2012). When considering engagement in the process of making school choices, Kumar and Hill (2012) found socio-economic background did impact parental choice for Primary school places. They focused on the importance of parents using choice as a means to select a school for their children; for example, if parents took advantage of the total number of school options available to be selected. Parents in more deprived areas selected fewer schools due to limited choices. The authors do highlight that richer parents may not need to be as strategic in their choices because they live closer to 'better' schools which in itself is subjective and open to discussion, and does not question the larger issue raised by Kumar and Hill (2012) that realistically this choice is dependent on wealth not engagement with the school application system.

#### 4.2.2 Education viewed as an economic product

It is a recurring theme within recent education policy changes to refer to economic benefits of education using OECD guidance and its PISA rankings. Using the OECD to justify educational recommendations is arguably problematic as it automatically frames successful education within a discourse of education as an economic imperative. The OECD markets

PISA as a key indicator of a country's future economic success, framing human knowledge as capital (Hopfenbeck et al., 2018). Schmidt (2017: 28) argues 'True to its economic roots the OECD places the complex world of education within the economically driven and rather straightforward framework of productivity'. In doing so, it assumes that it is possible to capture numerically relationships between teaching, curriculum, and learning (Moss, 2009). James (2015) suggested 'The driver for this seems to be the belief that higher measured scores on national and international indicator systems will raise the country's economic competitiveness in a global market.' Education is given market value in terms of successful production of a better educated population, resulting in a more employable work force (Exley and Ball, 2014; Bailey, 2004; Baltodano, 2012; Ball, 1998); education itself is now not for the greater good of the citizen or as education being a 'moral' right (James 2015), but instead for the good of the economic market (Davies and Bansel, 2007).

#### 4.2.3 Implications of marketised education

In summary, there are degrees of marketisation of education. Schools are positioned as the consumer of resources in a market of assessment software, tests, and frameworks (Ball et al., 2012: 524). This coexists with the marketing of school through parental choice, with OfSTED judgments and SATs informed league tables to inform this choice. Both of these markets are linked with school funding allocated by pupil numbers, encouraging schools to compete for maximum class sizes. Therefore, as pupil KS2 SATs data informs the league tables and parent school choices resources which recommend themselves on the bases of SATs preparation or similarity are preferential.

On a larger scale, education is viewed in policy as the answer to a competitive global economy where low-skilled work is increasingly supplied abroad, and England needs to provide education fit for higher skilled jobs available in the future, thus emphasising the market driven neoliberal message for education.

### 4.3 COMPETITION

As described previously, subsequent governments have continued focusing on Education policy in the terms of marketization (Ball, 2003). Parental choice is emphasised and controls where funding is provided for their child, which Whitty and Wisby (2016: 317) argue is to 'increase the power of the 'consumer' and reduce that of the 'producer'' where the parents

- the consumers - have greater power over schools - the producers – encouraging schools to compete for pupil places. This is an example of how Ong (2017) uses the term neoliberalism with a small n and big N; parents are ‘free’ to choose the best school for their children (n) based on information like league tables ranking and comparing schools against each other (N). As a result, schools themselves are governed by choices (n) they need to make to out-compete other schools and secure pupil numbers and funding (Davies and Bansel, 2007). To manage and encourage ‘excellence’ in teachers, to improve the school’s chances of competing, Ball (2003) argues new technologies (used to mean mechanisms and systems that drive and regulate performativity) and structures are put in place;

*...Teachers are re-worked as producers/providers, educational entrepreneurs and managers and are subject to regular appraisal and review and performance comparisons. (Ball, 2003: 218).*

Practices developed during the Labour government (Furlong, 2013), contributed to Coalition reforms making a regulatory system of targets linking teachers to ‘...their status, their pay and potentially their jobs’ (Pratt, 2018: 2). Pratt makes a link to the neoliberal view of the self, and the importance of personal performance in achievement (Ong 2007’s small n). This further internalises competition in the school market by delegating responsibility to teachers for the schools’ success measured by KS2 SATs;

*...by making the stakes so high, teachers have little choice but to invest themselves in this game and thus become complicit in the marketisation, competing for the highest grades they can extract from their pupils (Pratt, 2018: 3)*

#### **4.4 ACCOUNTABILITY AND AUTONOMY WITHIN THE NEOLIBERAL MESSAGE**

The flavour of the Coalition government’s justification for higher autonomy and accountability derived from international comparisons and evidence from OECD reports and PISA scores. The adopted viewpoint being that high accountability and autonomy are successful aspects of education practice in countries out-performing England. This autonomy is viewed by Bailey and Ball (2016: 130) as part of the marketisation of education, enabling more choice for parents in terms of what schools could provide.

The first part of this chapter drew attention to the dominance of high autonomy and accountability, as the way to improve education standards in the Coalition's education reform. A persistent influence of neoliberal ideologies of government may drive this message of educational success, as Fautley (2017: 87) states 'Neoliberal governments tend to prefer "small" government, and so removing centralization from some aspects of policy is viewed as a good thing.' Bailey and Ball (2016: 133) describe this aspect of coalition policy as tensioned '...between a weak but strong state', weak with the devolution of power towards schools and an increase in Academies, and strong in terms of school monitoring and regulation.

Schmidt (2017) is sceptical that autonomy can drive up standards in the current accountability climate. He argues high accountability with minimum control of *how* success is measured (OfSTED, KS2 SATs) undermines autonomy of teachers;

*...Teachers have become "responsible providers" who have no impact in decision making...they are accountable but have no avenues to construct accountability. Their autonomy is merely clerical... (Schmidt, 2017: 29)*

Similarly, James (2015) argued that more autonomy will only increase standards if systematic accountability pressures are reduced.

Despite increasing government rhetoric of devolving more control to schools, the reality is counter-argued by Green (2011).

*...the reality is that contracts, targets, performance indicators, and monitoring and evaluation systems act as new forms of control. They continue to govern practitioners ('governing-at-a-distance') in order that they might 'enterprise themselves' and enact policy reform. (Green, 2011: 45)*

As Green (2011), Schmidt (2017), and James (2015) argue, the controls in place in terms of accountability measures dominate decision making of schools. In order to investigate this, it is worth exploring briefly what is understood by the term 'Accountability'. A duality of the term exists, one in relation to how successfully something is audited, and the second being an ethical meaning of 'being answerable to' (Green, 2011: 12-13). In educational management these meanings can be conflated with successful auditing used as a means of

being answerable to something, tying up accountability in school monitoring such as OfSTED and KS2 SATs.

For teachers, this has implications in terms of performance related pay, with performance likely based on pupil attainment data. There are also implication for practices encouraged by the OfSTED inspection framework, ensuring those practices are visible, and, altering teaching practice in preparation of inspection to ensure all areas are explicitly met (Perryman et al., 2017). Ward and Quennerstedt (2019) present an example of this distinct impact of self-governance while exploring the impact of SATs based accountability on maths teaching. They found many of the lesson activities were reflective of testing, with priority given to lesson content which could evidence test-measurable skills (Quennerstedt, 2019: 274). This performative stance meant that creative aspects of maths were side-tracked in order to enable learning which could be more clearly evidenced. These strategies which 'play the system' are labelled as performative, ensuring what *needs* to be seen is explicitly seen, to avoid repercussions in terms of a drop in league table place, of OfSTED inspection triggers, or failures and/or forced academisation.

*Until Policy makers wean themselves off their faith in the idea that only the measurable is manageable, we are a long way off a new accountability paradigm becoming a reality (Green, 2011: 6)*

Green (2011) calls into question whether schools are more autonomous with less centralised control of the teaching practice, especially in light of current test-based accountability pressuring the education system.

#### 4.5 MANAGERIALISM

Green (2011) describes management under a neoliberal influence as a new type of governance implemented within schools as a result of the adoption of 'New Public Management'. The 'Language of 'outcomes', 'indicators', 'performance criteria'' (Green, 2011: 8) have become dominant, requiring explicit evidence through audit trails to judge performance. The measurable becomes fundamental;

*"If you can't measure it, you can't manage it"...which meant that organizations needed to be 'transparent' in their operation. (Green, 2011: 41)*

Where standards are to be raised schools must meet measurable targets in what Ball et al. (2012: 514) describe as a top down audit culture in education, instilling a 'performance culture' through use of 'league tables, national averages, comparative and progress indicators, OfSTED (Office for Standards in Education) assessments and benchmarks. The notion of measurability equating manageability prioritises aspects of education monitoring which can be measured through tools such as KS2 SATs results. This directly encourages school practices which make explicit, through what can be measured, that standards have been raised (Green, 2011; Ball et al., 2012). It is concerning that schools assessing without levels have no guidance for non-SATs years to measure standards, which may increase the use of teaching practices which can more clearly evidence performance criteria, this being referred to as 'performativity'.

#### 4.5.1 Performativity

As referred to above, institutional practices of managerialism ensure the setting and tracking of targets to monitor and encourage success, argued to encourage teaching practices and strategies which positively influence measurable outcomes of standards. Performative behaviours, in this context, are teaching strategies which will impact the most on SATs scores at the end of KS2, achieved at the expense of things which cannot be measured in that form (Ball 2012). Teacher priorities are complicated by the impact of performance related pay, and school pressures to show visible progress (Ball, 2003; Pratt, 2016; 2018). As Ball (2012: 31) reasons,

*...Indeed, performativity works best when we come to want for ourselves what is wanted from us...*

One of the reasons suggested for an increase in performative behaviours in schools is a rise in surveillance, teachers are subjected to 'the appraisal meeting, the annual review, report writing, the regular publication of results and promotion applications, inspections and peer reviews that are mechanics of performativity.' (Ball, 2003: 220).

#### 4.5.2 Effect of surveillance

Perryman et al. (2017) focus on an additional contributor to performativity, that of being observed and judged by OfSTED. They speculate that we are in a 'post-panoptic' era (Boyne, 2000; Courtney, 2016). 'Panoptic' governance refers to Foucault's use of a metaphorical

prison with inmates surrounding an internally lit guard tower, where prisoners could not identify when they were being watched (Foucault, 1980). In theory this leads to a self-governance of behaviours expected at all times due to the uncertainty of observation. Perryman et al. (2018: 149) found;

*...inspection was not just about surveillance, but the threat of surveillance, and engendered a regime in which schools self-govern their performance...*

Therefore, a 'post-panoptic' performative era is where aspects of performative behaviours previously only adopted during an inspection, are now normalised and internalised, regulating behaviour even outside of observation by inspection. The use of pupil assessment tracking as a way to monitor to teacher performance was indicated in research by Alderton and Pratt (2021) and has implications for performative assessment behaviours referred to in the next section.

#### 4.5.3 Conflict and performative behaviours

For the teacher, complying with performative systems can be conflicting, especially if what is required cannot be seen as worth doing (Ball, 2003; Metcalf, 2017). This can have adverse effects on teachers where 'The struggles are often internalized and set the care of the self against duty to others' (Ball, 2003: 216). Pratt (2018: 11) suggest that some teachers do not actively engage with the performativity system in an authentic way, but play 'the system to construct an appropriate picture of success' To Ball (2003: 255) this is 'both resistance and capitulation', because it still requires teachers to engage with the technologies of performativity and argues further that a new kind of teacher has been created to 'maximize performance, who can set aside irrelevant principles, or outmoded social commitments, for whom excellence and improvement are the driving force of their practice' (Ball, 2003: 223); with universities being required to produce teachers who will embrace this new culture (Furlong, 2013).

Why performativity has this conflicting effect on teachers may be due to neoliberal influence with a small n (Ong, 2007), where improvements come from 'the self', success is from effort and self-drive to improve (Ball, 2015; Perryman et al., 2017; 2018). The individual teacher's pursuit of success is seen as sufficient to ensure success of the whole school through a managerial delegation of small responsibilities of the whole (Green, 2012).

Each is responsible for their own improvement, with rewards and sanctions in place to encourage success, but responsibility lies with the individual to govern themselves. Perryman et al. (2017: 746) use Foucault's categorisations of power to shed light on the self-governing aspects of neoliberalism's impact on self-improvement and regulation,

*This shift is linked to governmentality, a portmanteau word combining government and mentality or rationality...Governmentality...refers to the self, so is also how and why the self shapes its own conduct in particular ways.*

Governmentality of the self influences actions, as the self is accountable and ruled by themselves, teachers become '...a truly reflective practitioner under the subtle persuasion of governmentality, dominated yet free' (Perryman et al., 2017: 755). Pratt (2016: 896) argues that:

*schooling thus demands of teachers engagement in a doxa of performativity in which they must do accountability work; that is, they must act in particular ways that are valued within the school system and make these visible to others.*

Using a lens of Bourdieu, one's *doxa* dictates an acceptable choice, therefore even though teachers theoretically 'choose' to teach a narrowed curriculum and engage with the neoliberal system in play, it can be argued to not do so is inconceivable with the dominance of the performative message (Pratt, 2016).

#### 4.6 EFFECTS OF NEOLIBERAL REFORMS

Davies and Bansel (2007: 252) argue the neoliberal constructed citizen is instructed that their personal economic drive for success is required for national economic success;

*...the newly responsabilized individuals fulfil their obligation to the nation/state by pursuing economic wellbeing for themselves and their family, for their employer, company, business or corporation.*

In this sense, the commodification of education in primary schools placing value on KS2 SATs scores become *the* measure for success, for the school, for the pupil and for the teacher.

Marketing this product (KS2 SATs results) on league tables has had a knock-on effect of prioritising teaching strategies which will positively impact on these.

The literature describes a conflict between teachers wanting to teach and being dominated by this Neoliberal discourse at large (Bailey, 2004; Ball, 2003; Pratt, 2016). A teacher may be judged as working professionally by evidencing set criteria but may miss the mark of what would be seen as professional outside of the managerial model by 'successful' teaching, repackaged into the performance driven priorities teachers now work with (Stevenson, 2010). Baltondo (2012: 489) describes teachers as being 'stripped of their most precious role: the duty to educate a generation of fully informed democratic citizens'. Pratt (2016: 901) found teachers felt constricted in their practice, expressing they felt they had

*...no choice but to put their faith in the numbers themselves as representative of something of 'real' use value for the pupil, whereupon it becomes necessary to believe that the numbers being generated represent a change in something tangible for pupils themselves.*

Although agreeing that these pressures are significant on teachers, some of the literature exploring teacher agency describes a more hopeful picture, where teachers are able to prioritise student needs strategically. This is explored in the next chapter in terms of teacher agency and identity.

#### 4.7 CONCLUSION

This chapter has outlined the many ways Neoliberal discourses influence education policy in England and how such policy is a vehicle for these marketised discourses. The impact of this on schools and teachers has been explored and will be used as a lens to view the data findings through. In doing so, the second of my research questions will be addressed concerning the policy into practices implications of AwLs. Questions have been raised over the successfulness of high autonomy for non-statutory school assessment with the continued presence of test-based accountability at the end of KS2. The main discourses identified in this chapter will now be used to explore the effects on teacher agency. In doing so it will link together the neoliberal impact on policy and teaching practice providing a theoretical lens to view teaching and assessment practices without levels.

## 5 TEACHER AGENCY AND IDENTITY

---

The education reform removing levels from primary assessment aimed to give schools more autonomy over non-statutory assessment. [Chapter 4](#) Education Policy explored how a neoliberal ideology has influenced the decision to adopt high-autonomy with high-accountability in the 2010 Education reform; that chapter also identified conflict for teachers when expected to conform to performative measures of standards as a means to hold teachers and schools to account. Teacher agency and identity will be adopted as a theoretical lens to view teacher interview data through to explore the implications of the removal of levels on teacher identity and agency in a time of high-autonomy and accountability.

### 5.1 THEORISATION OF AGENCY

Biesta and Tedder's (2006) ecological theory of agency is particularly applicable to the topic of assessment reform as this theory directly applies to education, as well as having a specific focus on curriculum changes and teacher agency, which are pertinent to this research (Biesta and Tedder, 2006; 2007; Biesta et al., 2015; 2017). In Biesta and Tedder (2006) they build on theory from Emirbayer and Mische (1998: 790), who defined agency as:

*...the temporally constructed engagement by actors of different structural environments—the temporal-relational contexts of action—which, through the interplay of habit, imagination, and judgment, both reproduces and transforms those structures in interactive response to the problems posed by changing historical situations.*

Emirbayer and Mische (1998) theorise three elements of agency, the '...iteration, projectivity, and practical-evaluation' (ibid: 790). The iteration element forms habitual aspects of agency through an actor's selection of appropriate behaviours informed by past interaction and experience. Projectivity is an imaginative element to agency where the actor 'imagines' outcomes of possible actions responding in relation to their '*...hopes, fears, and desires for the future*' (ibid: 791, emphasis in original). The practical-evaluation element encompasses aspects of agency where adaptation occurs, negotiating between the possible

directions of action projected in response to the emergent demands of the current situation. Essentially, agency is framed in how an agent reflects on the past while hypothesising the future and adapting to present (Priestley et al., 2016). This views agency as subjective because each individual has unique experiences which form future responses.

The extent to which the agent exercises each of the three components (*iteration, projectivity, and practical-evaluation*) is dependent on the situation itself, at that point in time (Emirbayer and Mische, 1998: 792). Agency, although being viewed as complex and multifaceted, is theorised by Emirbayer and Mische (1998) as something possessed by the agent in a fixed capacity. In disagreement with this, Biesta and Tedder (2006) view agency as an achievement dependent on the agent's transaction with their environment, rather than something which is simply possessed. Agency;

*...should always be understood in transactional terms, that is, as a quality of the engagement of actors with temporal-relational contexts of action....*

*(Biesta and Tedder, 2006: 18)*

This allows more scope for theorising why agency is demonstrated in one situation and not another, as each situation is temporally and transactionally different. The transactional element with the agent's environment is what makes this theorisation of agency ecological, it is '...actors-in-transaction-with-context, actors acting by-means-of-an-environment rather than simply in an environment...' (Biesta and Tedder, 2006: 18). In this ecological view

*...the achievement of agency will always result from the interplay of individual efforts, available resources and contextual and structural 'factors' as they come together in particular and, in a sense always unique situations (Biesta and Tedder, 2007: 137)*

This can be different teachers experiencing the same assessment changes but with varied achievement of agency, or a teacher achieving agency in one time and place but not another.

## 5.2 TEACHER IDENTITY

Teachers use their identity to define themselves, with identity itself ‘...defined by personal and social histories and current roles...’ as well as teachers’ ‘...beliefs and values about the kind of teacher they hope to be...’ (Day et al., 2006: 610). Teachers hold both personal and professional identities which ‘...contribute to motivation, commitment and job satisfaction’ (Day et al., 2006: 610), which positions teacher identity as important and influential on teachers’ work (Beauchamp and Thomas, 2009). These personal and professional identities are theorised as dynamic and evolving, and able to enact influence upon one another (Beauchamp and Thomas, 2009; Priestley et al., 2015), for example, Buchanan (2015) located the construction of a teacher’s professional identity from personal previous lived experiences such as their own schooling and teacher education. Identities are also viewed as multifaceted, containing ‘sub-identities’ within them (Beauchamp and Thomas, 2009). Thus, a teacher’s professional identity can have many parts depending on the positions they hold within their job, they may be a class teacher and subject lead or coordinator, or hold leadership positions for phases or key stages.

These identities and sub-identities can be in conflict with one another, for example, a teacher who strongly believes in the value of education but encounters competing demands of preparing children to pass standardised tests which may require teaching practices they disagree with (Biesta et al., 2017). Identities are considered to be subject to both external and internal influences. External influences comprise of policy and policy discourses as well as schools’ mediated enactment of policy (Beauchamp and Thomas, 2009). Buchanan (2015: 704) refers to these external influences on identity as cultural tools, which contribute towards the remaking of identities when ‘...make[ing] sense of and interact[ing] with their local working context.’ Buchanan (2015) focuses on elementary school teachers’ agency in light of years of accountability reforms in America and provides a parallel to this present research through drawing on teachers’ agency under test-based accountability.

Identities are subject to change based on external influence which are, themselves, mediated through internal factors such as values and beliefs (Day et al., 2006), which consequently then inform an individual’s history, forming their future values and beliefs. Policy for example, is viewed as an external influence, shaping teachers’ professional

identities, whereby teachers use their, ‘...pre-existing identities to interpret, learn from, evaluate, and appropriate the new conditions of their work in schools and classrooms’ (Buchanan, 2015: 701) and ‘reform and remake’ their identities as a consequence. Conflict between aspects of identity or sub-identities act as a site for this reform and remaking of identity.

### 5.3 IDENTITY AND AGENCY

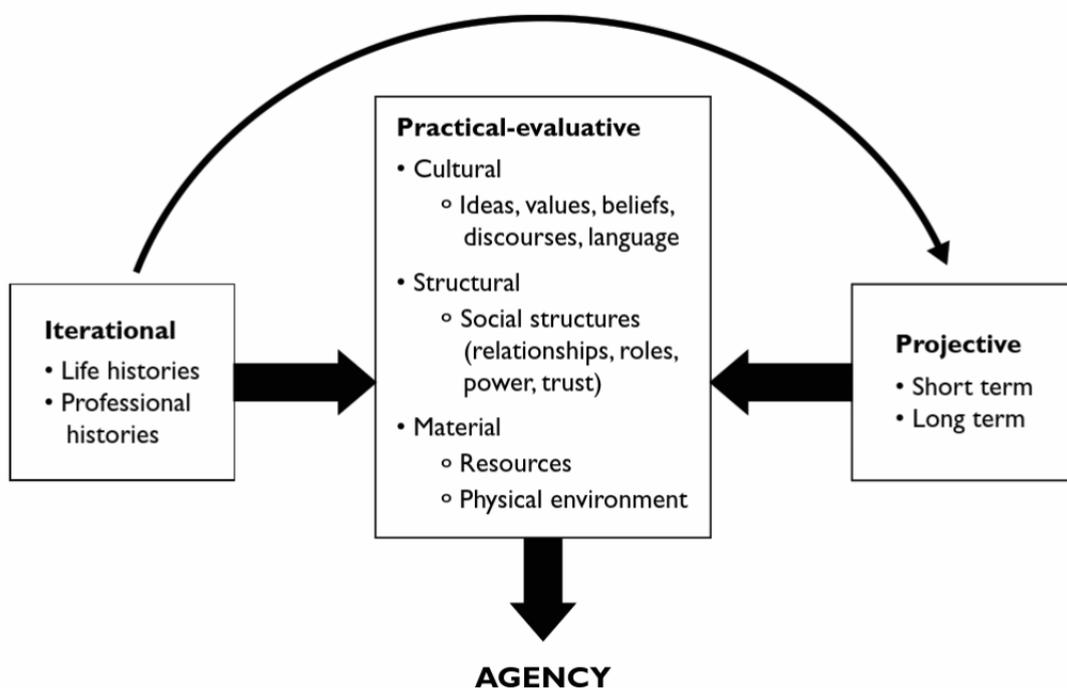
Where contradictions may occur within and between identities, agency can be seen as the expression of how those contradictions are managed (Beauchamp and Thomas 2009; Buchanan 2015; Day et al., 2006). Disruption creates opportunity for agency in the active rebalancing of conflict in a teacher’s identities. A particular site for this agency is the ‘...act of negotiating dominant discourses in the process of identity formation [seen] as a site to locate and investigate teacher agency’ (Buchanan, 2015: 705). As such, Buchanan (2015: 714) described teacher agency as ‘identities in motion’, with agency being an active process. Day et al. (2006: 611) similarly suggests that tensions and contradictions can arise as a result of policy reform as a result ‘...agency is expressed by the extent to which people can live with contradictions and tensions within these various identities...’ It is a common theme in teacher agency literature to look at times of policy reform, theorising changes in education policy as disruptions, and analysing how different teachers respond (Biesta et al., 2015; Priestley et al., 2013; 2015). It is important to note that disruptions and conflicts are not viewed as universally negative or positive, merely as events which impact on an individual’s experience of the world, consequently impacting on a teacher’s identity and providing a site potentially for agency.

#### 5.3.1 Inclusion of teacher identity theory into the ecological model

Bringing this back to the transactional ecological theory of agency (not be confused with Bronfenbrenner’s (1992) ecological systems theory concerning human development), identity to Biesta and Tedder (2007) and Biesta et al. (2017) contributes to an individual’s interaction with an environment unique to that time and place and shape how they respond to constraints within it, with identity referenced in terms of a teacher’s values and beliefs. Similarly, Beauchamp and Thomas (2009) place teacher identity as an influence on teacher agency. Additionally, they suggest that agency can be achieved by awareness of one’s own

self, their teacher identity, and thus are more empowered. Identify in this sense represents an influential factor of agency rather than a driving force. Whereas Buchanan (2015) and Day et al. (2006) emphasise teacher identity in agency, they do not ignore the more ecological context specific influences of agency, rather they frame these transactional components of agency within a teacher’s identity. How this ecological theory of agency can be conceptualised, within literature reviewed, and drawing directly from Priestley et al. (2013), will be considered.

Figure 5-1 Ecological theory of Agency from Priestley et al. (2013: 152)



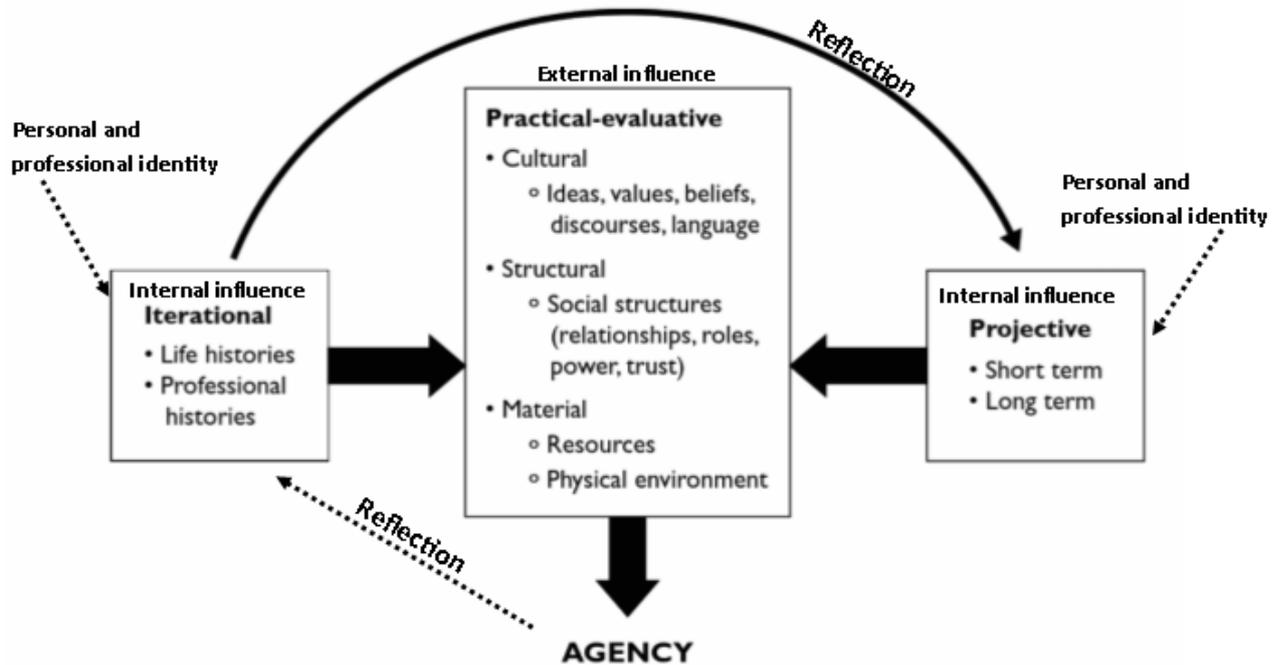
(Priestley et al., 2013: 152)

Reference to teachers’ personal and professional identities will be considered as part of the iterational and projective dimension of the ecological model of agency. As shown in Figure 5-1, from Priestley et al. (2013) the projective domain is fed into by the iterative domain. The life and professional histories in the iterational domain are informed by, and inclusive of teacher values and beliefs (Biesta et al., 2015; 2017; Priestley et al., 2013). This aligns with Day et al’s (2006) stance that values, and beliefs are central to a teacher’s identity.

The importance of future aspirations is highlighted by Day et al. (2006) who described the impact of the projective dimension on the teacher’s achievement of agency, including considerations of reflection upon past experiences. Therefore, life histories and professional

histories are not fixed, but constantly updated with new experiences through life, informing both short term and long-term projective goals. Identity therefore in its foundations of values and beliefs, is as a contributing factor influencing the practical-evaluative domain through an individual's projective and iterational domains.

Figure 5-2 Annotated version of Priestley et al. (2013) model of ecological agency



(annotated from Priestley et al., 2013: 152)

Figure 5-2 is a version of Figure 5-1 annotated by me, showing the place of personal and professional identities, and where they may contribute to the achievement of agency. This could be interpreted as showing the influences of a teacher's constructed identity on agency, or how agency is the result of teacher identities being kept in balance through active reflection and enactment, which adapts the construct of the identity to remain unconflicted in light of a disruption.

#### 5.4 TYPES OF AGENCY

Achievement of agency is categorised in terms of how agents either comply or not with practices required by various sites of power, for example the school or the government, or both. Compliance is not viewed as fixed or inflexible, just as the theorising of the ecological view agency is not. Complying without question or awareness of a choice is not viewed as

achieving agency, as well as 'actors who feel creative and deliberative while in the flow of unproblematic trajectories' (Emirbayer & Mische, 1998: 1008).

There is widespread support in the literature for the notion that policy change in education results in teachers and schools who resist, and others who comply, including active pursuit and promotion of policy reforms. Buchanan (2015: 709) uses the terms 'stepping up or pushing back', where 'stepping up' means not only complying with policy, but actively pursuing it, as opposed to 'pushing back' which is demonstrated by teachers rejecting policy and negotiating its implementation. Resistance can be seen as a search for what is best in response to the upheaval caused by a policy change. For example, teachers may search for pedagogies they feel are more beneficial for students and therefore '...reject or resist key aspects of central policy' (Moore and Clarke, 2016: 667). To Emirbayer and Mische (1998: 1006) when faced with uncertainty some agents may '...hold tightly to past routines' while others may 'seek to imagine alternative futures for a problematic present' Therefore, as schools and teachers were experiencing the transition to AwLs the iterational and projective domains may hold useful clues in determining sites for agentic resistance.

According to Priestley et al. (2016), who examined what they called creative mediation, compliance does not automatically indicate low agentic achievement. They found Cypriot teachers were able to mediate policy creatively when given more curriculum control and benefitted from a less pressurised performative and attainment driven accountability system. Strategic compliance is another term used to describe some agentic achievement while on the surface complying with what is required. A principled form of strategic compliance was categorised by Hoyle and Wallace (2007: 19) as '...principled infidelity: infidelity because they do not slavishly adhere to expectations, and principled because they seek to sustain their professional values'. Teachers however who were not motivated by their principles were characterised by Priestley et al. (2016) as not achieving a high amount of agency, as their behaviour was driven by 'habitual reproduction of former modes of practice' (Priestley et al., 2016: 195), using previous curriculum material and adapting it to suit new requirements. This is not to say these teachers were not achieving agency, but their reproduction of past practices was a limitation. When analysing teacher agentic achievement, both motivation and strategies chosen, need to be considered.

For some teachers, championing and actively pursuing policy was driven by a desire to further their career prospects (Ball et al., 2011). Others, who demonstrated agency by resistance due to tensions between competing personal and school values, potentially missed out on career progression opportunities because they openly questioned national policy implementation and changes (Priestley et al., 2016: 194). Priestley (2011b: 18) stated that 'In such a scenario, agency resolves this tension...' where '...genuine agency is exercised as a reflexive response to contradictions within the individual's immediate social world'. However, being critical of policy may have negatively impacted the aforementioned teacher's future achievement of agency by denying them access to the resources a promotion could have brought. Although Ball et al. (2011) views such critical challenges as keeping alternative discourses alive, when policy only provides a narrow set of discourses that can, for example, influence the development of new teachers' identities who are reliant on policy.

## 5.5 INFLUENCES ON AGENCY AND IDENTITY

It is useful to explore literature around teacher agency and identity and draw attention to factors which impact on this. As referred to above, teachers mediate policy through their own identity, and how this occurs is tied to their agency, therefore enabling factors and constraints on teacher agency are important considerations for teachers experiencing the shift towards assessment without levels. Discourses discussed in [Chapter 4](#) in addition to school values and professional experience, will be further explored relating to their impact on teacher agency.

### 5.5.1 Discourses

Discourses influence identity '...because they are part of the official structures and cultures within teachers work...' (Biesta et al., 2017: 40) or they are unofficial, but currently popular within the education community, for example, teachers as facilitators rather than educators (Biesta et al., 2017). Neoliberal, marketisation, performativity, and accountability discourses are evident in, and have influence on policy (see [Chapter 4](#)). Buchanan (2015: 702) states,

*...a discourse can be seen as a shorthand way to acknowledge how a point in time influences the way its members view the world and understand themselves within it.*

Therefore, how teachers view what it means to be a 'Teacher' and construct their professional identity is affected by the political and social climate of a time.

According to Biesta et al. (2015: 636), incremental changes in education reform lacking a clear educational philosophy for schools, have resulted in education discourses becoming '...a mishmash of competing and vague ideas...' where teachers' '...goals often seem to be short-term in nature, focusing on process rather than longer-term significance and impact', as teachers struggle to '...locate their work within deep consideration of the purposes of education'.

Teachers are at risk of unknowingly incorporating contemporary dominant discourses into their professional identities, for example, internalising the neoliberal marketised influences of education policy (Moore and Clarke, 2016). This implies that age and teaching experience may be a variable in the different identities teachers' hold and thus how teachers interpret education reform.

#### 5.5.1.1 Accountability

Agency has the potential to be '...shaped and enhanced by policy that specifies goals and processes, enhancing the capability of teachers to manoeuvre between repertoires, make decisions and frame future actions' (Priestley et al., 2015b: 144). As presented in [Chapter 3](#), a justification for removing levels from assessment in Primary schools was to give schools more autonomy over their non-statutory assessment. However, the education reforms of 2010 emphasise the discourse of test-based accountability, which Biesta et al. (2015: 638) argues can '...value certain modes of action over others', confusing the educational priorities and goals of a school which risks inhibiting these policy aims through a continued focus on statutory test outcomes. Viewing education policy in England with a lens of irony, Hoyle and Wallace (2007: 14) argue that what education policy aims to do can be inhibited by policy itself, as,

*...neither governments nor parents have abandoned the social, moral, spiritual and aesthetic goals of education, despite the fact that reformist policies have made their achievement even more difficult.*

For example, Priestley et al. (2016) observed a recent trend in curriculum policy in Scotland, less prescriptive and aiming to give more autonomy to schools in curriculum design. However, these changes have been accused of giving with one hand and taking with the other, by decreasing input regulation while increasing output regulation in the form of monitoring and accountability pressures (Priestley et al., 2015b, Wilkins 2015).

*...it is problematic for policy to demand that teachers exercise agency in their working practices and then simultaneously deny them the means to do so, effectively disabling them (Priestley et al., 2016: 189)*

Increased autonomy in a continued presence of a dominant test-based accountability discourse is seen as a contradiction, referred to above by Hoyle and Wallace (2007). This does not mean to say that increased school or teacher autonomy will never result in higher achievement of agency, only that it is too simple to conflate one with the other without considering the wider impact of accountability pressures Priestley et al. (2016) refer to.

How teachers, especially those in Yr6, respond to accountability demands is of concern, with some literature strongly suggesting that test-based accountability is becoming normalised and thus internalised within teachers' professional identity (Buchanan, 2015; Moore and Clarke, 2016). Buchanan (2015: 713) argues that the dominance of this view of accountability is imbedded in teacher practice and pedagogy through an internalisation of the discourse.

*Over the last decade, teachers have altered their instruction in order to adhere to the accountability demands...for many, it is an unexamined influence. The use of external, standardized test measures to guide and assess teaching has become a habit, a social practice that teachers use in order to understand their position in relation to students and curricula.*

Elementary teachers in Buchanan's study disagreed with test-based accountability but felt validated by student test data. Initially considered as a contradiction within the teacher's views of test data, it was then theorised to be the internalisation and normalisation of accountability practices defining a 'successful' teacher- it had become part of their professional identity. This has implications for Primary school teachers now assessing

without levels where test-based accountability persists and in the absence of non-statutory assessment guidance.

#### 5.5.1.1.1 *Performativity*

Measuring teacher performance is essential to current accountability measures in place and is usually centred around pupil assessment data. This practice has been accused of producing a performative culture in schools as discussed in [Chapter 4](#) (Wilkins et al., 2012). The identity of the teacher is therefore not only bound up in their own performance but also in their students' performance. This can produce inauthentic teaching practices to meet performance targets (see [Chapter 2.6](#)).

Performativity is also viewed in regard to Education policy enactment where, it '...must be seen to be done, that is reported as done and accounted for' (Biesta et al., 2015: 629). Proof of action is argued as being more valuable in some cases than action (Hoyle and Wallace 2007). This is not to be understood as schools pretending to teach children and documenting it, rather this could be result of balancing policy enactment with the suitability of policy to the school, as Hoyle and Wallace (2007) state,

*...in order to meet the perceived needs of their pupils, Head teachers and teachers have had to adapt national policies while, at the same time, appearing to implement these policies with fidelity. Hoyle and Wallace (2007: 17)*

For teachers, these performative behaviours come at a cost, participation in practices they do not see value in, subsequently taking time and energy away from practices they do find worthwhile, may conflict with their professional and personal identities (Wilkins et al., 2012). How much of an impact this has on a teacher's achievement of agency is unclear. Wilkins et al. (2012: 68) state that,

*Performative discourses can be construed as denying personal agency through their emphasis on instrumentality. Yet teachers are not merely passive conduits of policy; they mediate, interpret, resist and subvert policy imperatives, bringing their own values to bear on the implementation of performative objectives.*

However, Moore and Clarke (2016: 672) argue,

*As individuals, we might become consciously aware of discourse and the strength of its influence – on ourselves and on others – but this does not necessarily make it any easier for us to resist it or to act outside it.*

Within this view, performativity is engrained in policy and the workings of a school, mediating its effects does not remove the risk of internalising and normalising the impact it can have on a teacher's professional identity and in turn their achievement of agency. This is further explored in the section below 'Age and Teaching Experience' (Ball et al., 2011; Buchanan, 2015).

#### 5.5.1.2 Professionalism

Government reforms and discourses within education impact on how teachers are or are not defined as professionals by others and themselves, which has implications for teacher identity and agency (Moore and Clarke 2016; Wilkins 2011; 2015).

The introduction of Professional Standards for teachers (TDA, 2007) has been accused of de-professionalising teachers by creating a check list for teachers as technicians, carrying an instrumentalist view of teaching (Buchanan, 2015; Priestley, 2011b). An increase in monitoring and accountability measures combined with reduced autonomy from the introduction of the national curriculum has resulted in a loss of trust in teachers, and a high expectation of compliance confining the space available for agency by defining rigidly what it is to be a 'professional' (Priestley et al., 2013; Wilkins, 2011). This consequently leaves anything outside the rigid definition of 'professional' open to accusations of being 'unprofessional' (Moor and Clarke, 2016). This is not to say that there is no benefit, or there has not been successes to the monitoring of schools and introducing a common curriculum for England, or that teachers were unprofessional before, or are unprofessional now, only that as a consequence what it means to be a 'professional' has changed. This impacts on how teachers define themselves as professionals and construct their professional identities (Buchanan, 2015).

For example, teachers' professionalism is viewed as quantifiable and measurable against the Professional Standards (TDA, 2007), as Wilkins (2011: 395) argues,

*Professional Standards attempts to define not simply what a teacher does, but what/who a teacher is; to provide a framework for assessing not just their skills, competency and knowledge, but their values and attitudes.*

To Wilkins (2011), the Professional Standards impact directly on a teacher's professional identity, contributing towards a performative culture where compliance to Government policy is measurable. The impact of this is argued by Wilkins (2015: 1153) as performatively motivated teachers whose own educational values and aims are confined by 'functional competencies'. Not only is the teacher being defined by a narrow concept of 'professional', their professional identities are at risk of being dominated by this, and also influenced by how they are measured against it.

Narrowing what can be conceived as a professional teacher to the standards, risks excluding other concepts, and withholds teachers the opportunity to define themselves, for instance, as professional 'autonomous agents' within a community of practice (Wilkins, 2011: 395). The normalisation of these discourses through active rebalancing of identities (see [Identity and agency](#)) may be why teachers internalise aspects of this professional discourse, Moore and Clarke (2016: 674) caution the result of this may be that;

*Professionalism in this more fixed and limited articulation is thus likely to involve little more than getting on with the simple business of preparing students to pass standardised memory-based tests and examinations, 'managing behaviour', and keeping students quiet and 'on task' in the classroom. Other ambitions that the teacher might have – or have had...may become pathologised within the same policy discourse as potentially 'damaging' to students, even if valued and retained by the teacher.*

It can therefore be the case that some teachers' own discussion of professionalism differs little from policy discourses, thus limits the opportunity for such teachers to critically evaluate and find alternative concepts of a professional (Biesta et al., 2017: 52). More experienced or older teachers may have access to differing discourses of professionalism through a longer career or being taught under different education policies (Ball et al., 2011). This, as discussed below, is of particular concern for newly qualified teachers.

### 5.5.2 School values and ideologies

The values and culture of a school have the potential to impact on a teacher's professional identity and therefore their agency (Biesta et al., 2017; Priestley et al., 2016). A teacher's own values and beliefs, which contribute towards their identity, may be at odds with, or complement aspects of the school's values and beliefs.

Biesta et al. (2017: 48) found that teachers were able to achieve agency when they felt strongly about their views and showed conviction, however, this was strategic and often at classroom level. Teachers were shown to strategically comply with schools' demands they disagreed with, while putting extra effort in to fulfil their vision of how to teach. In line with findings from Buchanan (2015), Biesta et al. (2017) who attributed 'tensions' between teacher views and beliefs and those of the school, to constraining achievement of agency, even when a teacher felt conviction about their educational views. This stresses the role of school context in the achievement of agency, as highlighted by Priestley (2011b: 20) who encouraged schools to promote ecological conditions for teacher agency by knowing '...when not to act as well as being decisive in action' and to '...not deny or repress the agency of classroom teachers as their agency is key to the success of innovation...'. Further research by Priestley et al. (2016) supports this view in regard to findings of a higher achievement of agency among teachers who felt supported enough to take risks in a culture of shared practice and trust.

*The culture of a school, the way it defined successful teaching and learning, and the ideological and pedagogical positions it made available for teachers mediated how teachers experienced and reacted to accountability policies...The interaction between teacher identity and school culture both enabled and constrained teachers' agency. (Buchanan, 2015: 714)*

As Buchanan (2015) states above, school culture is a crucial aspect of a teacher's achievement of agency. It interacts with a teacher's identity, which in turn impacts on how the teacher is able to achieve agency through reflection and projection of what is possible. In the ecological model of agency a school's values and culture is located in the practical-evaluative domain in the structural and material aspects. This domain is mediated by iterational and projective domains, therefore teachers may have long term education goals,

informed by their life histories, which are enacted or not depending on the material and structural position of the school.

#### 5.5.2.1 School mediation and translation

AwLs has meant schools are required to construct their own non-statutory assessment practices with the school's values and ideologies able to influence how AwLs reform was implemented in the school. This mediation of policy can be seen as a '...filtering out and selective focusing done by Head teachers and their SLTs...' (Ball et al., 2011: 626), where teacher experience of policy is mediated through the culture and educational ideologies of a school (Buchanan, 2015: 71). The policy enacted then may not resemble the policy envisioned by those who design it through schools' translation of policy for it to work in the school's context, managing a '...delicate balance between making policy palatable and making it happen...' (Ball et al., 2011: 326). To Ball et al. (2011: 367) schools are '...continually disrupted or faced with contradictory expectations, but this is an incoherence that can be made to work, most of the time'. In making policy work, unintended consequences of policy can be mitigated making the policy reform seem more successful than it perhaps would have been if the policy was followed to the letter (Hoyle and Wallace, 2007). Hoyle and Wallace (2007) suggest the success of Government education policy is down to schools and teachers taking it upon themselves to work harder to ensure the education of the children is not negatively affected by reforms not conducive to the successful education of the pupils. Biesta et al. (2017) found evidence of this in their study with a teacher going beyond the time-consuming test preparation of their students to ensure their vision of a more complete education, being to some extent successful. Teachers as Ball et al (2011: 366) state '...are creative and sophisticated and they manage, but they are also tired and overloaded much of the time', suggesting that teachers and schools are able to cope, but at a cost.

#### 5.5.3 Age and teaching experience

At the start of their career, aspects of teachers' personal identity inform their early professional identity, such as their own experience as a student, and their motivation and aims when entering the profession (Buchanan, 2015; Wilkins et al., 2012). Initial teaching experiences may conflict with these early concepts of who they will be as a teacher, such as educational goals being constrained by school practices, or by unanticipated demands

(Wilkins et al., 2012). For example, Buchanan (2015) found conflict in teacher identity through teacher training, when theory did not match school placement experience. To Buchanan (2015) this is an example of agency as ‘identities in motion’ with early career experiences particularly powerful in constructing professional identity. Similarly, Wilkins et al. (2012: 70) found early career experiences to be particularly formative, as new teachers enter into a new environment with established social and cultural practices and need to develop ‘...a sophisticated understanding of a new institution in an effort to gain accepted membership of it’. Additionally, there is caution of the effects of limited and dominant discourses on education, which have been normalised within notions of what makes a ‘successful’ teacher (Buchanan, 2015; Wilkins, 2011; 2015).

How new teachers mediate policy compared to older more experienced teachers is a concern. Having access to only current policy and a reduced frame of reference compared to more experienced teachers, new teachers are at a higher risk of dependency on policy, seeing limited room for interpretation (Ball et al., 2011; Biesta et al., 2007). For example, Ball et al. (2011: 366) found evidence ‘...of different discursive generations, trained or educated differently as teachers, with access to different discursive archives’ Similar observations were made by Biesta et al. (2017) who found that both age and generation had an impact on how teachers talked about education and education practice. A teacher who had taught longer could recount how ideas in education had repeated, such experiences provided access to ‘a discourse that gives...a ‘bigger picture, and put[s] things in perspective’ (Biesta et al., 2017: 44), compared to a newer teacher who had limited experience of different policy ideologies, thus providing ‘less opportunity to develop...discursive ‘resources’’. In a similar fashion, according to Ball et al. (2011: 634),

*...new teachers have themselves been educated and done teacher training within a particular regime of accountability and ‘design’, and may not be able to imagine a different way of being a teacher. They have no ‘field of memory’ of ‘different times’ on which to draw.*

If new teachers, when forming their professional identities, are exposed to the same or similar dominant discourses as their own education as a child, they may not be aware that there are other ideologies in education (Ball et al., 2011). For Buchanan (2015), a

persistence of discourses such as test based accountability over the past 30 years has contributed to the creation of teachers who have internalised these demands as part of their professional identity.

*Accountability discourses have reshaped the landscape of teacher professionalism...For many teachers, this is not only the paradigm that has been in place for most (or all) of their career, but was also the climate of their schooling experiences when they were students....as a result, it is likely that more and more teachers' professional identities and constructions of agency will be formed within this discourse...as fewer and fewer teachers will have prior experiences that offer alternate conceptions of professional teaching.*  
(Buchanan, 2015: 716)

Therefore, new professional identities are felt to be vulnerable and susceptible to the compliance of 'accountability' and 'marketisation' agendas in education (see [Chapter 4](#)). This matters because of the potential for a reduction of agency, through limiting what can be envisaged as possible to that which comes from limited exposure to differing educational discourse (Biesta et al., 2015; Moore and Clarke 2016; Priestley et al., 2013). In this case agency is not being achieved, as the aspirational views (the projective domain) of the teachers are narrowed from their focus on policy discourse. Based on the previous theorising of how identity disruption provides space for agency, if these concepts do not conflict with aspects of the teachers' personal or student identities there may be reduced opportunities for achievement of agency (Biesta and Burbles, 2003). More recently qualified teachers may not have experienced teaching with NCLs and thus would not have access to assessment experienced from a Government centralised non-statutory framework. More experienced teachers would have taught with NCLs and may have trained at a time where test-based accountability did not dominate, or experienced this as a student. Therefore, the agency and identity theory discussed in this chapter so far provides a framework for analysis of teachers views of AwLs, relating to their experience as a teacher.

#### 5.5.4 Is agency desired?

It may be natural that the achievement of teacher agency is a focus for education research and practice when it has subsequently become harder for teachers to achieve with an

increase in both curriculum and teaching prescription after the introduction of the National Curriculum in the 1980s (Priestley et al., 2016). From this perspective, agency is promoted and encouraged in the literature, specifically the conditions which encourage teacher agency (Biesta et al., 2015; Priestley et al., 2016). More recently global education policy discourses are centring teachers as 'agents for change', as Priestley et al. (2015b: 135) say,

*...policy demands that teachers exercise agency in their working practices, then simultaneously denies them the means to do so...*

For Priestley et al. (2015b), studying teacher agency identifies ways to increase and enable teacher autonomy when schools are required to increasingly develop their own curriculum citing public policy as aiming to increase a teacher's capacity for autonomy. Furthermore, agency of tomorrow is influenced by experience of the past, therefore factors which enable and restrain agency now will influence future achievement of agency. They identify the situational contexts of schools of being the site for this development. Teacher agency in this regard is both, an interesting lens to analyse policy through, but also an opportunity to open spaces for teacher agency when assessment autonomy has increased in primary schools.

Discovering ways teachers can be agentic is a common aim in education agency research reviewed in this chapter. A lack of teacher agency is generally reported as an issue (Moore and Clarke, 2016; Priestley et al., 2016; Wilkins, 2015). When teachers are required by national or school policy to comply with practices felt to inhibit student learning, teacher agency is an opportunity for educational values and beliefs of teachers to persist, for example; Moore and Clarke (2016: 675) who wished to encourage 'unhappily' compliant teachers to find 'active forms of resistance' to ensure the children they taught could succeed. However, achievement of agency is not viewed, unequivocally, as beneficial (Priestley 2011b). Schools seem to respond positively to teacher agency as policy entrepreneurs and enthusiasts (Ball et al., 2011), and when 'stepping up' (Buchanan, 2015) to actively promote, enhance and enact policy in the school. On the other hand, resistance to policy reform and contemporary dominant discourses has not been favourable, even when motivated by teachers' educational values, rather than personal gain (Biesta et al., 2017; Priestley 2011b; Wilkins, 2015). Although Moore and Clarke (2016) encourage teachers to find ways to resist, teachers could be taking risks with their career when voicing

opposition to policy reform, resulting in being overlooked for promotion (Priestley, 2011b), or at the extreme, even at risk of losing their position if viewed as disruptive and preventing the school from fulfilling their values (Wilkins, 2015). Wilkins (2015: 1152) argues that compliance in schools is embedded and normalised, and where a teacher's 'dissent restricts their [leadership's] 'right to manage'', those teachers can be 'purged' from the school. Tied in with this is how the discourse of managerialism has influenced SLT and head teachers' management of schools and teachers using accountability procedures to measure performance against a defined notion of what makes a 'good' teacher (Wilkins, 2015). Hoyle and Wallace (2007) hypothesise that this discourse has been internalised by head teachers and teachers resulting in over-management. Regarding teacher identity, this now '...has to be understood in the context not only of the varied cultures within the school (students, parents, teachers), but also of schools' managerial practices and cultures.' (Wilkins et al., 2012: 67).

This indicates that agency from a teacher, in the form of resistance, could be viewed as a threat and differences in a teacher's educational values and those of the school could impact on their career prospects through the internalisation of managerialism of the SLT (Priestley et al., 2016). Discourses, especially when limited, may reduce the achievement of agency by restricting the projective domain within the ecological view of agency, as it reduces a teacher's opportunities to '...envisage different futures, and through denying them the language with which to engage critically with policy' (Priestley et al., 2013: 157). Priestley et al. (2013) found teachers' values and aims can be limited to discourses found within policy, and mostly focused on short term instrumental goals, indicating the discourses in policy had become dominant. This is concerning when teachers are inexperienced and are limited in approaches to education pedagogies (Ball et al., 2011). The limitation of discourses to those in policy therefore limits the teacher's projective domain, and may prevent them envisioning different approaches to education, other than those dominant in policy, at the expense of aspirational educational goals like social justice (Priestley et al., 2013).

## 5.6 CONCLUSION

Teacher agency and identity will be used as a theoretical lens and analytical tool to consider agency of participants to analyse factors which enable and constrain agency with AwLs. As can be seen from the complexities outlined in this chapter agency and the way it manifests itself in teachers is a highly complex construct impacted by policy and the discourse policy influences. Therefore, teacher agency is a fundamental consideration in addressing both research questions in this study.

- What has been the impact of the assessment reform 'Assessment without Levels' on teaching and assessment practice?
- What are the Policy into Practice implications of the assessment reform 'Assessment without Levels' for Primary school teachers?

It will provide insight into the policy into practice implications of the policy change AwLs as well as capturing the impact through teacher's agency, or lack of, over their teaching practices. In Chapter 9, (Teacher Agency analysis and discussion) I explore how the teachers in my study were not always aware that they had agency but instead felt strongly compelled to adopt a compliance stance with that which they were asked to do by the school in and in turn by government policy.

## 6 METHODOLOGY AND METHODS

---

### 6.1 INTRODUCTION

The aim of this research was to explore the impact AwLs had on teaching and assessment practices in primary schools and the policy into practice implications.

The next section will go into more detail about the paradigms selected that reflect my ontological and epistemological beliefs, which Guba and Lincoln (1994) state is a fundamental aspect of research. This research will approach paradigms with the perspective that they are ‘...human constructs...’ defining ‘...the worldview of the researcher as-interpretive-bricoleur’ (Denzin and Lincoln, 2005: 183).

In brief, I have adopted an ontology of interpretivism, where ‘...there is no single “truth”[and] All truths are partial and incomplete...’ (Denzin and Lincoln, 2005: 89). Constructivism has been adopted as an epistemology, in both social and radical (von Glasersfeld’s radical constructivism) terms, along with Dewey’s pragmatic approach outlined in section Pragmatic research design. The paradigm of critical theory has also been drawn on for analysing education policy in terms of origins and consequences of policy on teachers’ AwLs (Green et al., 2012).

Methodologically, a mixed method approach has been undertaken to gather the views and experiences of participants working as Primary school teachers, interviews in two West Midlands schools and a broader quantitative survey. For each data collection method, detail is provided addressing the relevance and format as well as the ethical considerations made as part of the research design. A brief profile is presented for interview participants alongside background information of the two interview settings. Demographic information of survey participants is also provided from survey questions which gathered data on, for instance, the year teachers taught in or if they had taught with National Curriculum Levels (NCLs).

The analysis procedure for both sets of data is addressed in their own section, referring to underlying theory where required. As part of this section reflections will be made from the Pilot study undertaken to trail the data gathering methods and the analysis of interview

data. Attention will be given to areas of the piloting which resulted in changes made to the main study interview questions and data analysis.

To conclude this chapter, the validity and reliability of the research will be discussed. This sets out procedures and safeguards to ensure the research is trustworthy and credible.

## 6.2 RESEARCH QUESTIONS

It is appropriate to restate the research questions at this point;

- What has been the impact of the assessment reform 'AwLs' on teaching and assessment practice?
- What are the Policy into Practice implications of the assessment reform 'AwLs'?

## 6.3 MULTIPLE RESEARCH PARADIGMS

As a doctoral researcher, I fell into a trap of looking for the 'right' paradigm to fit my research and world view. In search of the 'right' paradigm I discovered two things. The first, that 'right' is subjective and dependent on context, and the second, that there is no 'right' paradigm when what is 'right' is subjective and dependent on context. That theories and perspectives were named and categorised led me at first to believe they were distinct and contained within each category, selected and applied to research which suited them.

Historically, however, paradigms have shifted, changed and ended being '...replaced by new paradigms' (Grenfell and James, 1998: 8). There is also no universal agreement of the vocabulary used for the discussion of 'world views', as well as what that constitutes.

Common between all of the different perspectives on the categorisation and vocabulary of 'world views' are the main groupings within social science of positivism, post-positivism and anti-positivist (including constructivism), critical theory, and pragmatism (Cohen et al., 2007; Creswell, 2009; Denzin and Lincoln, 2005; Green et al., 2012; Grenfell and James, 1998).

For social scientists there has been a separation from a positivist 'objective' view of 'reality' and 'truth' to explore the 'multifaceted images of human behaviour as varied as the situations and contexts supporting them'(Cohen et al., 2007: 22). To attempt to select a 'right' paradigm is hanging on to a positivist perspective that knowledge is obtainable through experiment, controlling variables to reveal the 'truth', which in the past dominated

what was thought of as valid research (Cohen et al., 2018). Therefore, I have developed a view that previous methodological theory is not a manual for research with no right way to interpret and use it (Murphy, 2013: 7). Such theory was derived from ideas available at the time, which were adapted, adjusted and added to, to derive knowledge, henceforth, my methodology draws similarly on relevant theory and perspectives today to derive knowledge from my research's current context; research should not be constrained by the theory, '...theory should not precede research but follow it' (Cohen et al., 2007: 22). There is still, however, tension within the research community, with some advocating a blurring of boundaries between paradigms, while others reinforcing those in place (Denzin and Lincoln, 2005). A cautious blending of paradigms is advised by Guba and Lincoln (2005: 197):

*...there is great potential by interweaving of viewpoints, for the incorporation of multiple perspectives, and for borrowing, or bricolage, where borrowing seems useful, richness enhancing, or theoretically heuristic.*

The carefully considered application of aspects of interpretivism, pragmatism and critical theory will be explained further below, with how they will be useful and richness enhancing when answering the research questions.

In explicitly explaining my justification for my methodology, I am making my positionality clear. My own background as a teacher cannot be ignored when interpreting meaning through this research. I acknowledge, just as the participants have subjective constructs of reality, so do I. Outlining my positionality in this chapter contextualises my interpretation of the data and the construction of the knowledge as a result of it. This is so the reader will be able to view the research through my lens, and consider this in my interpretations of my participants' responses.

### 6.3.1 Interpretivist ontology

Understanding the way in which the 'Assessment without Levels' reform has impacted on schools required taking into account how varied and complex teaching experiences can be, depending on a number of factors, including the year that is taught, previous experience with assessment, position in the school, and the mediation of Education Policy through school policy enactment. This will be analysed using the theory of agency set out in [Chapter 5](#). Assessment changes are also theorised as being mediated through the participants

themselves according to their previous lived experiences. Whereas realists believe there is a “real” reality “out there,” (Guba and Lincoln, 2005: 202), ascribing truth to evidence that brings us closer to proving the ‘real’ reality (Saywer, 2000), interpretivist ontology rejects this. Interpretivism as an ontology is appropriate for this research as it takes into account the subjective reality of the participants, and how this alters their experience of the assessment changes. For example, a participant who had experienced assessment with levels will have a different frame of reference compared to a teacher who had not. Therefore, interpretivism allows analysis of the complex nature of reality faced through each individual’s experience of events (Cohen et al., 2007; Guba and Lincoln, 2005).

### 6.3.2 Social constructivism

Assessment practices in Primary schools do not occur in isolation, knowledge and understanding of assessment is constructed in a particular school environment. The topic of this research explores a time where previous knowledge of NCLs had become redundant, and schools were constructing their own frameworks of assessment. Schools are understood as social places, where the teacher participants interact with students as well as other members of staff, it seems appropriate to employ social constructivism, as it accounts for knowledge which is constructed by ‘...interaction with others (hence social constructivism) and through historical and cultural norms to operate in individuals lives.’ (Creswell, 2009: 8). Hence, social constructivism accommodates social and interactive elements of knowledge creation under the assessment reform being explored.

### 6.3.3 Radical constructivism

Radical constructivism views knowledge as neither fixed nor passive. This form of constructivism is ‘...radical because it breaks with convention and develops a theory of knowledge in which knowledge does not reflect an "objective" ontological reality...’(von Glasersfeld, 1984: 24). So the emphasis on the ‘real’ is not on what happens outside of one’s mind but within it (Burger and Luckmann, 1966).

Radical constructivism focuses on subjective organisation of an individual’s experiences. Since everyone’s experiences through life are not identical, every ‘member of some society, has its own reality according to which and in which it lives’ (von Glasersfeld, 1989: 122). As participants are the ones who ‘do’ the assessing, the reality of assessment changes are

viewed as constructed from their experiences. Any notion of the 'real', outside of their interpretation, does not contribute towards their understanding of the world and the knowledge they gain from it. This radical constructivist lens provides a theory which complements the participants' adaptation to changes as new knowledge, providing space to explore variation in their experiences.

One common criticism of constructivism is that its subjective nature allows no constraint on what can be said to be 'true', and that anything can be interpreted from anything (Sayer, 1999). However, von Glasersfeld argued that knowledge needs to be complemented by experience or it will be challenged, and at best seen as questionable (von Glasersfeld, 1984, 1989). For knowledge to be common, and understanding shared between individuals, radical constructivism proposes a concept of compatibility between our individual knowledge constructs, rather than a matching of our individual knowledge to an external reality (von Glasersfeld, 1984, 1989).

For me to analyse data from the participants I do not aim to assume to have identical cognitive structures (knowledge) between us, but ones that are compatible. My interpretation and understanding of meanings derived from my participants is based on attempts to achieve a compatible understanding of research themes. Viewing understanding as complementary constructs of knowledge, accounts for teachers having different views and experiences of assessment changes, even when working in the same setting. For example, previous experiences, the year group they teach in, and their personal history will impact on how they construct their knowledge. However, shared experiences and the social nature of teaching enables construction of compatible understanding. Agency and teacher identity theory will be used within this frame of reference to explore differences in the teachers' experiences and views of assessment changes.

#### 6.3.4 Hermeneutics

Cohen et al. (2007: 27) state 'Hermeneutics focuses on interaction and language; it seeks to understand situations through the eyes of the participants...and premised on the view that reality is socially constructed'. It recognises subjectivity in the interpretations of these constructs and subsequently identifies that there can be 'multiple, correct interpretations of given "text"' (Green et al., 2012: 15), accounting for differing view and responses from

participants. A text refers to what is being interpreted, for example, a text can be an expressed view, a policy document or a dominant discourse. Because it recognises there can be multiple interpretations, they cannot all be correct, hermeneutic interpretations could 'always be wrong, and may be shown to be wrong as new facts appear' (Green et al., 2012: 15). In disagreement with Green et al. (2012), as Murphy (2013) posits, there is no right or wrong way to use theory, doing so implies there is also a 'right' way which echoes a positivist view of reality. Knowledge will not be viewed as fixed but built upon, challenged, or changed (von Glasersfeld, 1983; 1989). Moreover, interpretations are constructed from one's unique experience, therefore each person's interpretation will be unique, but real to them, therefore neither rendering them as right or wrong, but different (von Glasersfeld, 1983). The pragmatic view of knowledge within this research framework (this will be discussed in more detail below) additionally does not see truth or facts as fixed, but constantly reflected upon, and, at times, reconstructed; this could be viewed as discovering new knowledge as referenced above, but in a similar fashion to radical constructivism, this knowledge is individual to the person, therefore varying and changing (Biesta, 2016). It is this view of adaptive and subjective knowledge that the hermeneutic aspects of interpretation will align, acknowledging that interpretations of assessment changes may be different between participants, rather than 'wrong'.

## 6.4 RESEARCH DESIGN

### 6.4.1 Pragmatic research design

Pragmatism goes beyond being a methodology (Morgan, 2014) valuing the inquiry itself over the purity of aligning one's self to either qualitative or quantitative methodologies (Creswell, 2009). I have used the Deweyan view of pragmatism interpreted and used by Biesta and Burbules (2003) as they apply it specifically in terms of education and educational research. The education research context of these perspectives will also provide a more meaningful theory to apply to this research design than wider pragmatic theory.

Knowledge traditionally, and in a constructivist sense, is attributed to something which happens in the mind, 'knowledge is not a passive registration of reality 'out there'' (Biesta, 2016: 84). This pragmatic view of knowledge requires an action, an interaction with the reality of the environment (Biesta and Burbules, 2003: 11). Assessment is an action and

requires an interaction with the 'thing' being assessed, one cannot 'know' something until it is required to be applied through action (Biesta and Burbules, 2003: 12). Knowledge is in the 'knowing' something will or will not work, reflection, conscious or unconscious, is needed alongside the action, forming an intelligent trial and error (Biesta, 2016: 83). For this research, the need for action and reflection to construct knowledge has relevance with the processes teachers and schools have undertaken to adapt to AwLs. An assessment or teaching approach is not 'known' to work until there is action and reflection. Dewey's placement of reality in experience also provides a space for the effect of 'perceived' pressures to be recognised as both subjective and real at the same time. Dewey's pragmatic view on knowledge accepts that we each have our own individually constructed views on reality (Biesta, 2016: 87). Similarly, knowledge formation also does not depend on a conclusion of the existence of a 'real' world.

#### 6.4.2 Mixed methods design

To gain, most fully, teachers' experiences of AwLs mixed methods were adopted. These comprised of:

- Quantitative survey with optional text boxes if participants wanted to expand on their responses, for KS1 and KS2 classroom teachers
- Qualitative interviews with Year 2 and 6 teachers and a member of SLT

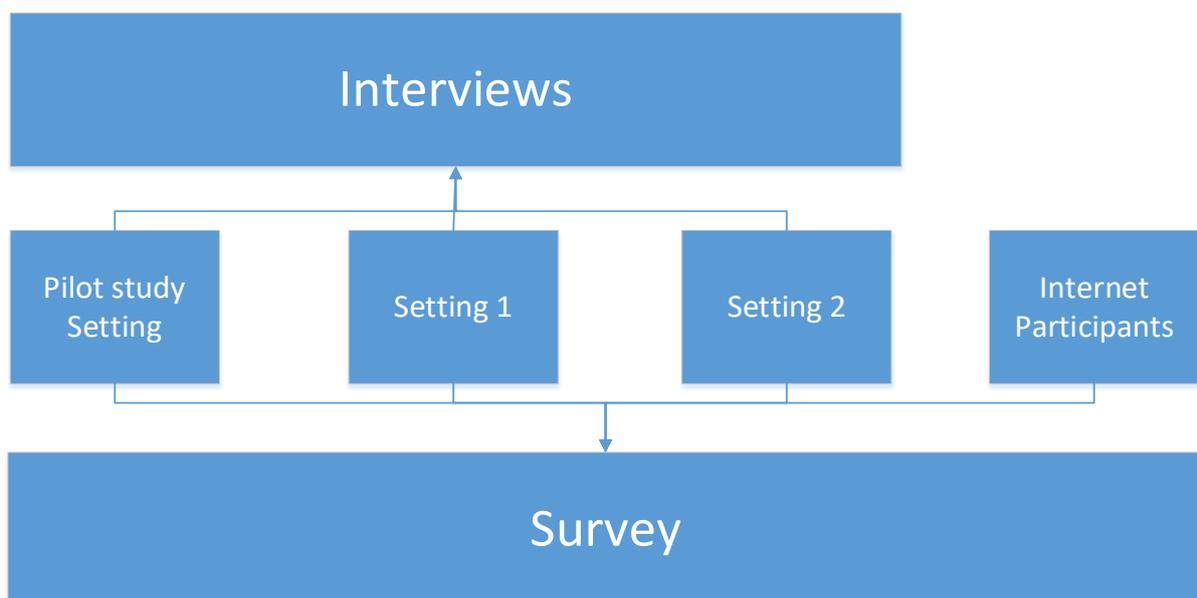
A key reason mixed methods was adopted was to be able to bring both breadth and depth to the topic being explored. Depth was gained from interviews with Year 2 (Yr2) teachers, Yr6 teachers and an assessment lead/member or member of the senior leadership team who were sampled selectively to provide a more in-depth view of experiences of AwLs. Breadth of responses was achieved from teachers across all years in KS1 and KS2, in the form of a quantitative survey to provide a wider scope of views across KS1 and KS2 teachers (Creswell, 2009). Additionally, using mixed methods brings with it the benefits from data triangulation and the added value of differing perspectives which contribute towards interpretations, (Cohen et al., 2018; Creswell, 2009). As such, data was combined with 'Concurrent Triangulation Strategy' (Creswell, 2009), meaning data were collected at the same point during the study but analysed separately before being drawn back together (See Triangulation).

There are theoretical criticisms of the commensurability of mixed methods (Cohen et al., 2018). For example, Sayer (1999) states that resolution needs to be made at the ontological levels to resolve incompatibilities when combining methods and Biesta (2012) questions whether methods can be mixed within research or combined and treated separately. Drawing on multiple perspectives, Cohen et al. (2018: 28) concludes that ‘Researchers need not become mired in the paradigm debate; as long as we know what we are dealing with in MMR [mixed methods research] then this may suffice’. Moreover, according to Gorard and Taylor (2004), combining qualitative and quantitative methods can yield greater strengths than when used alone. They acknowledge pragmatism in their justification for combined methods (used interchangeably with mixed methods), however, they are cautious to label combined methods as pragmatic to avoid a creation of a ‘pragmatic paradigm’. Similarly, I have not used pragmatism as a paradigm, rather as a research design approach which allows me to use a combination of theory within paradigms appropriate to answering my research questions.

#### 6.4.3 Overview of data collected

Table 6-1 below visually displays where the data was gathered for each method used in the research.

*Table 6-1 Diagram to show an overview of the methods and where the participants were recruited from*



#### 6.4.4 Teacher interviews

Interviews were adopted as a data collection method for their advantage of being able to derive depth and insights directly from the participants enabling data to be checked for meaning and relevance as it is being collected (Denscombe, 2014). Interviews occurred in one pilot setting and two main study settings, one for classroom teachers who taught in Yr2 and Yr6, and the second for the HT and/or DHT in each setting. Interviews also enabled meaning to be derived as closely as possible to that intended by the participant as it allows for interaction, prompts and a level of interpretation at the sight of data collection. Semi-structured interviews were chosen as they provided flexibility during the interview process (Cohen et al., 2007; Opie, 2004) allowing participants to expand on certain topics and give greater insight into their thoughts and feelings while still answering core questions common to all interviewees.

##### 6.4.4.1 Interviewees and rationale

Settings were secured for teacher interviews through contact with the University's education department, with invitations sent to partnership Primary schools. Three schools agreed to meet me to discuss the research, but only two maintained contact and became settings for the main study. The third school stopped showing interest at the time when data collection was due to start. It was decided that the study would continue with only 2 schools as the time it would take to recruit an extra setting would impinge on data collection and might have run data collection over the year planned, 2017.

Participants were chosen for interview based on the position they had in the setting and are described in section Research setting and participants below. Yr2 and Yr6 teachers were selected for teacher interviews as these years are subject to external testing and the only years to have received teacher assessment guidance without levels by the DfE. These teachers' assessment and teaching practices, it was felt, were influenced the most by the changes in assessment and/or the head teacher. The members of SLT responsible for assessment were also interviewed. In Setting 1 (S1) the Head teacher (HT) and the Deputy Head teacher (DHT) were interviewed and Setting 2 (S2) the DHT was interviewed. These interviews were felt to be important as they could provide an additional point of view of the assessment changes and the school wide implications for assessment as well as helping to

contextualise the school mediation of assessment policy in terms of the expectation for the teachers.

#### 6.4.4.2 Interview format

The semi-structured classroom-based teacher interviews were split into two to accommodate the limited chunks of time available for classroom based teachers to be interviewed. The pilot study setting and setting one both had separate interviews 1 and 2, however, S2 requested both interviews happen in the same sitting. For continuity the interview schedule remained the same. HT/DHT interviews were done in one sitting in each setting, this interview was not planned as two and consisted of different questions to account for the different perspectives on assessment as a whole school approach adapting policy to fit their ethos and priorities.

Table 6-2: Duration of each interview

		Interview Date	Duration (min:seconds)
Setting 1	Teacher D	Interview 1:12/7/2017	17.57
		Interview 2:17/7/2017	14.50
	Teacher E	Interview 1: 12/7/2017	15.39
		Interview 2: 17/7/2017	26.31
	Teacher F	Interview 1: 12/7/2017	13.06
		Interview 2: 17/7/2017	12.03
HT/DHT	17/7/2017	38.09	
Setting 2	Teacher G	Interview 1 and 2: 22/11/2017	19.09
	Teacher H	Interview 1 and 2: 22/11/2017	43.32
	DHT	Interview 1: 22/11/2017	58.21

A copy of the interview schedules can be found in [Appendix 7 Full Study revised Interview Schedule](#) where more detail can be found on the questions. The teacher interviews aimed to first explore assessment with and without levels among core and non -core subjects as well as drawing on participant experience of the transition between the two. The second interview then focused on teacher assessment compared to test and how AwLs had impacted on their practice. HT/DHT interviews had questions aiming to explore how AwLs

had impacted on school policy, assessment as a whole school and how the transition was experienced. Questions were kept open to help avoid leading the participants into a particular answer (Denscombe, 2014: 191).

Keeping teacher interviews to under 25min was to consider busy teachers' schedules. DHT and HT interviews were 1hr maximum. For Setting 2 where interviews were done in one sitting participants had less time than Setting 1 to reflect on the topic of AwLs between the interviews, which could have impacted on interview question responses. However, participants were still able to add to or clarify their transcripts via participant validation. Creswell and Miller (2000) and Norris (1997) describe this process as showing the participants the interpretations of the data obtained and asking them to confirm it represents their views accurately. However, Cohen et al. (2007: 149) warn that in doing this '...researchers also need to be cautious in placing exclusive store on respondents...' as they cannot always be objective in their own responses. As this research accepts participant subjectivity as inextricably linked with their reality, this caution from Cohen et al. (2007) is not viewed as problematic. The respondent's validation in this research has been to confirm their interview transcripts reflected their view of what they discussed, offering them also the opportunity to clarify, add, or change content if needed to reflect their intended subjective meaning. In doing this participant could clarify their comments and affirm that what was transcribed authentically matched their viewpoint. Only one participant amended their transcript, teacher B From the pilot study setting, who wished to clarify their meaning. All other participants responded via email to confirm they were happy for their transcripts to be used in the research.

#### 6.4.4.3 Interviews: Ethical considerations

Social science research should ethically operate to '...maximise benefit and minimise harm.' BERA (2018). It should leave participants unaffected by their participation in the research (Denscombe, 2007, 2010). Therefore, ethical considerations were central to the research project in design, data collection procedures and analysis are addressed in detail below.

##### 6.4.4.3.1.1 *Equality, diversity in participant recruitment*

All year 2 and year 6 teachers were invited to participate in interviews regardless of gender, ethnicity, sexual orientation or socio-economic background. As participants were teachers,

they already had a good understanding of written and spoken English, however if this was not the case accessibility measures already in place in the school could be utilised. If interviews took part outside of the interview setting the researcher would adapt to accommodate accessibility needs where needed.

#### *6.4.4.3.1.2 Informed consent*

Informed consent was central to data collection. Primarily this took place as a presentation during a staff meeting where the research was introduced. Teachers in year 2, 6 and those in charge of assessment policy were provided with a detailed information sheet providing potential participants, in nontechnical jargon, the aims of the research, how their data would be used and stored, and the risks associated with participation (BERA, 2018; Denscombe, 2009). Participants had 2 weeks following the initial meeting to decide whether they wished to take part in the research and to ask any questions they had about participation.

In addition to consent being informed it was important to emphasise that participation was voluntary, and they had the right to withdraw their data without prejudice up to the point of participant validation (see [Interview format](#)). After the initial contact was made with the teachers through their school all communication continued via my BCU student email. Participants may have felt they were required to take part out of duty to their Head teacher therefore it was important that further contact occurred independent of the school leadership. Leadership would not be informed of the teachers who had or had not taken part. Voluntary consent and their right to withdraw was verbally recounted at the start of the interview to provide another opportunity for questions to be asked or for the participant to choose not to participate. One pilot study interview participant failed to arrange their second interview therefore their data was withdrawn. All other participants maintained contact and confirmed that they happy to keep their data in the research following participant validation of transcripts.

#### 6.4.4.4 Minimise harm

A central consideration for the methods selected was the impact of participation on workloads, 'Researchers should consider the impact of their research on the lives and

workloads of participants,' (BERA 2018: 20). This is why times were kept within 20min and with no preparation required by the participants.

Participants were given flexibility of where they would like to be interviewed for example in a public place, the university or in the school they taught in. All teachers requested to be interviewed in the setting, which arranged a room or space for interview. This was usually a small intervention classroom or office. On one occasion an empty communal area was the only space available. In this case I asked the participant if they were comfortable speaking in that area where they may be overheard by passing colleagues, and another time was offered to maintain the comfort of the participant (BERA 2018), but the participant felt comfortable to continue.

The pilot study school could not accommodate release time within their normal school teaching hours so interviews occurred after teaching time at a time to suit the participants the best. The main study schools provided the teachers with covered release time to allow participation.

Participants were informed of potential risk to themselves through participation of the research. As only three settings were used for the study (1 pilot study, 2 main study) despite creating synonyms for each setting and participant, identifiable information may have been traceable back to a participant. I replaced any identifiable information myself with fictional content as Denscombe (2009: 66) states 'Disguising the identities might involve additional alterations to the details – such as changes to the information supplied about the role of the participant, and possibly their age and sex.' This posed a risk if the teachers made comments which may not have been looked upon favourably by their senior leadership team (SLT). To guard against this, participants were asked to read through their transcripts to validate its conveyed meaning but also to remove any comments they felt would put them at a disadvantage if traced back to them at a later date. However, in doing so 'The better the identities of those involved are disguised, the more difficult it becomes to check the validity of the data.' (ibid: 66). This is an example where,

*Researchers also have a responsibility to consider how to balance maximising the benefits and minimising any risk or harm to participants, (BERA 2018: 8)*

Power relation cannot be ignored and it must be considered that there was an imbalance between myself and the participants. This impacts of the participants agency in participating and the answer they provide.

#### 6.4.5 Online survey

The Primary focus of the survey was to gather views of KS1 and KS2 teachers, providing a data set addressing practicing teachers' views on assessment with and without levels. The survey aimed to provide data which would address the first research question exploring the impact AwLs has had on assessment and teaching from the teachers' perspectives. As the interview data will focus on SLT, Yr6 and Yr2 teachers, the survey sampled teacher views from all KS1 and KS2 years. Survey data was analysed as a whole but also via demographic information for example: if they taught in a SATs year (see Quantitative Analysis procedures and theory).

Figure 6-1 Likert question structure for survey.

	Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	N/A
I feel I can confidently assess WITHOUT levels.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
There has been an easy transition from assessing WITH levels to without levels.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Assessing WITH levels was a better way of tracking children's progress.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
There has been enough guidance from the Government to help teachers assess summatively WITHOUT levels.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
The interim framework helps me understand how to summatively assess children's work WITHOUT levels.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Detailed exemplars for each year and each subject would help improve teacher confidence and accuracy when assessing maths and reading WITHOUT levels.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Please use this space to leave any comments about Teacher Assessment of Reading and Maths.

Figure 6-1 shows an example of the 5 point scale Likert-type questions chosen for the survey and how these were presented to the participants. Likert-type closed questions were chosen to produce comparable focused responses suited to eliciting responses about participants attitudes to a particular topic (Cohen et al., 2007; Williman, 2006). Cohen et al. (2008: 325) states that this method of eliciting responses from a questionnaire can create discrimination in responses when the same words can be interpreted differently with the use of rating scales further creating 'a degree of sensitivity and differentiation of responses'. Williman (2006) agrees that questionnaires are a flexible and useful tool but care must be taken for it to be used in a way that benefits the research which in this case provides a greater breadth of responses compared to the interviews. With surveys there is a chance

that answers given may not be genuine with little opportunity to check; triangulation of survey responses with any text responses and the interview data should help overcome this limitation of survey data collection (Denscombe 2014).

The first 6 questions of the survey collected teacher demographic information including; total years teaching, current school year teaching, previous years taught in, if they had assessed with levels and if they had assessed without levels. Questions 7-9 were presented as shown above. The main content of the survey explored 3 themes, teacher summative assessment of maths and reading, external assessment of maths and reading and summative assessment of non-core subjects. The option of 'non-applicable' was provided for any question relating to AwLs to accommodate those teachers who started teaching after levels were removed. Questions 7-9 had 6 statements. An example of this can be found in Appendix 3. The final question asked if the survey had covered all areas of assessment, they felt were important. If they answered no, a text response was required. In total, 36 text responses were submitted in the open text boxes for each set of questions. These responses were thematically coded and are included within the findings section. Through optional text responses and triangulation with qualitative interview data, the survey also provided qualitative insights, as Denscombe (2014: 29) states, '...surveys can just as easily produce qualitative data, particularly when used in conjunction with methods such as an interview.'

#### 6.4.5.1 Data collection

The survey was originally designed as part of a pilot study to canvas the opinions of all teachers in a setting to help structure the interview questions; it had been constructed on the University's preferred online survey platform, Bristol Online Surveys now Jisc Online Surveys, so it could be accessed by the teachers at a time which suited them. However, although the link to the survey was provided a month in advance and reminders sent weekly, teachers did not engage with the survey before the interviews. Because of this, it was not possible to use the survey data to inform the semi-structured interview questions as originally planned ([see Survey change of purpose](#)).

Survey participants were recruited primarily from the two school settings participating in the teacher interviews. They were provided the survey on paper which was subsequently

entered online by me. This way Jisc Online Surveys could be used as a tool to collate and analyse the data. Only the complete surveys were included from KS1 and KS2 teachers in the school settings to be analysed.

#### 6.4.5.2 Survey change of purpose

The online survey had been previously piloted by four post graduate researchers in my department, to ensure the layout and questions were clear. As referred to above, the survey was originally planned to be part of a sequential mixed methods approach where survey data would be collected and analysed first to inform the interview structure. Low engagement from teachers in the Interview Pilot Setting (PS) meant that the survey was adapted to be an independent source of data, where teachers from Interview settings would be recruited as survey participants. When the survey was reviewed following the piloting of the interviews, the results were at a standard that very little needed to be done to the survey. Therefore, upon consultation with my supervisor, it was decided to keep the data gained in with the main study's online survey data as it produced useful and insightful findings; though a question, Q10, was added to address the assessment of writing, a theme which emerged during the interviews. Due to the research taking place over covid-19 I was unable to gain access to the previous participants located in the PS to enable them to answer Q10. As such responses to Q10, addressing writing assessment have been removed from the data. Also related to survey question design, in hindsight there was opportunity in Q8 to ask if a combination of TA and test was preferable. Asking this would have provided insight into teachers' views of a combined form of statutory assessment.

#### 6.4.5.3 Survey: Ethical considerations

Just with interview data collection, ethical consideration underpinned the construction of the survey. The survey was made as convenient as possible for participants to complete, centring on four main areas and requiring Likert multiple choice selection to '...avoid making excessive demands on them [participants]...' (BERA, 2018: 19).

Participants were informed how data obtained through the survey would be used on the first page of the survey. Consent was required before moving past this page. The survey was available in paper form (manually entered by me) as well as online via Jisc Online Surveys. Although demographic information was requested this only included work related

information related to AwLs, such as experiencing NCLs and current year group. Where text responses could be given, data was handled just as the interview data. Participants had the right to withdraw their survey data up to submission and were informed of this.

To further enhance anonymity responses were given unique codes with survey data being analysed by demographic information provided and not according to where the responses were yielded. This way responses were anonymous with data not being traceable to the setting. A contact email address was given for potential participants to ask for clarification or more information if needed prior to and following participation.

## 6.5 RESEARCH SETTINGS AND PARTICIPANTS

### 6.5.1 Survey participants

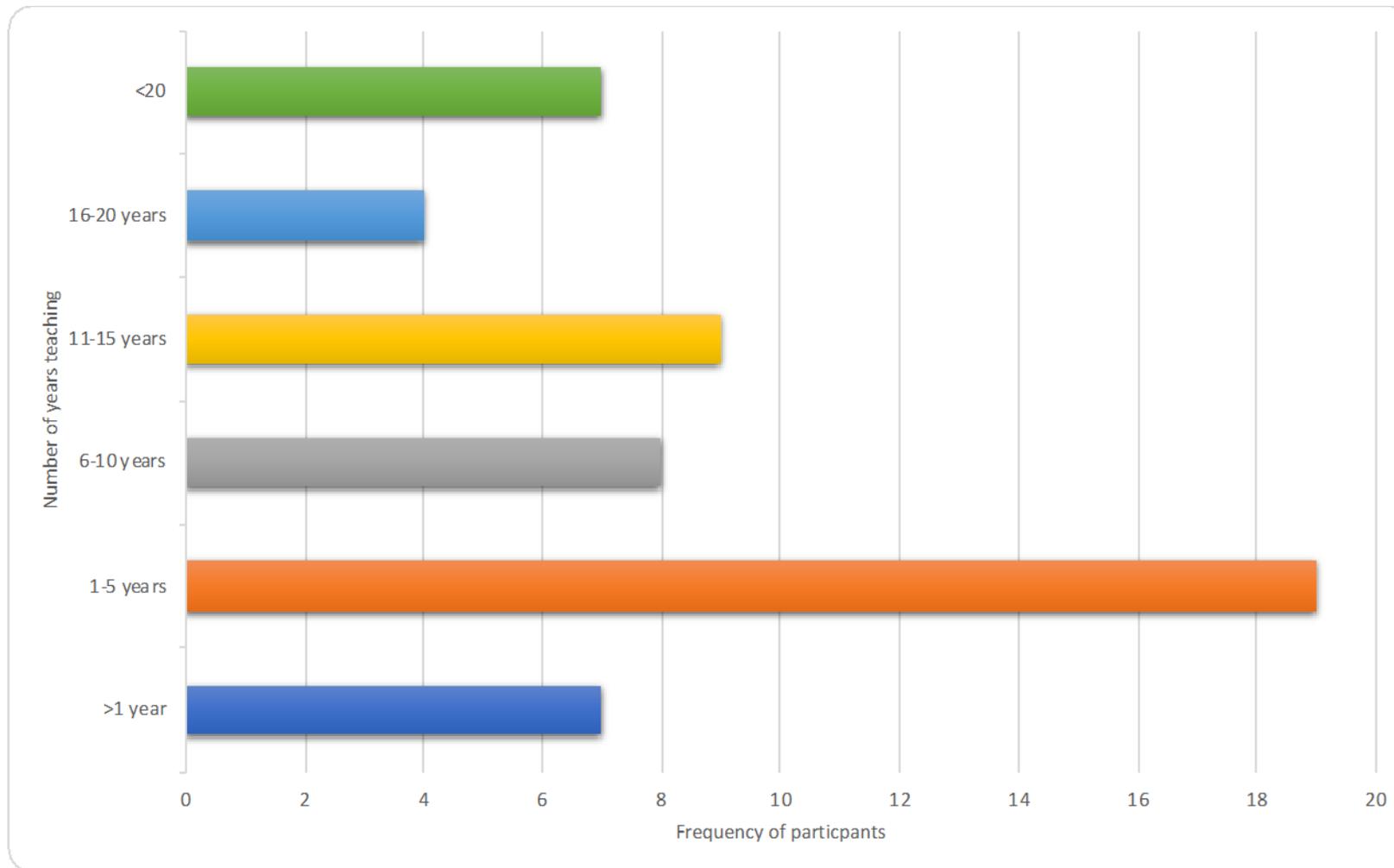
The survey had a total of 54 complete responses. Table 6-3 shows where participants were recruited from for the Online Survey. The setting for piloting the interviews was used to opportunely recruit participants as well S1 and S2 teachers.

*Table 6-3 Survey participant demographics.*

Key Stage		Setting responses				Current Year teaching						Assessed with levels	
KS1	KS2	Pilot	S1	S2	Online	1	2	3	4	5	6	Yes	No
41	14	14	19	16	5	5	10	11	14	13	14	38	16

The majority of respondents taught years 4-6, with the lowest number of respondents teaching year 1. Setting 2, a junior school (only has KS2 year groups), shifts the over representation of participants towards KS2 teachers. Respondents could select multiple years to allow for participants who taught in different year groups.

Figure 6-2 The number of years teaching experience of survey participants.



The spread of participants is weighted towards those who were teaching 1-5 years as seen in Figure 6-2. Overall, 28 out of the 54 had been teaching for 6+ years compared to 26 with up to 6 years' experience.

Figure 6-3 below shows how many participants accessed the survey and progressed to the end. For example; under p.3 the 5 means that many participants completed the survey to page 3 and did not continue.

Figure 6-3 The progress of internet link access to survey



There was a total of 727 links clicked to the survey. Out of these, 54 responses were completed. Participation numbers in surveys is noted by Denscombe (2014) as a challenge which may account for the low response rate from internet promotion of the survey. This may be due to the method of attempting to recruit participants in addition to the study schools though using social media. Twitter and Facebook were used with the hashtags of Primary assessment, Primary teaching, edchat, edchatuk. These are common hashtags used for people sharing tweets about education topics. It is possible that people clicked on the survey and read the information page and decided not to participate. Of these clicks some people may have been researchers, Head teachers, early year's teachers, secondary teachers or students themselves and therefore did not meet the criteria of a KS1 or KS2 teacher but may have viewed the post via the hashtags. Of those who did progress past the first page but did not complete the survey, something else may have become a priority, as teachers are known to have heavy workloads; the survey partially filled in may have been forgotten.

### 6.5.2 Interview participant demographics Setting 1

Setting 1 was a government funded two form entry Primary Academy in the West Midlands with a nursery from an area of low economic wealth. In 2017, students scored below the

national average of 61% to meeting the expected standard in Maths, Reading, and Writing at 53% in KS2 SATs.

#### 6.5.2.1 Teacher D

Teacher D was a Year 2 teacher and had recently trained to be a KS1 SATs teacher assessment moderator for the local authority. They had been teaching in Setting 1 for two years with a total of 13 years' experience the West Midlands local authorities. She had experience teaching in all KS1 and KS2 year groups.

#### 6.5.2.2 Teacher E

Teacher E had 22 years of teaching experience mostly teaching in year 5 or year 6 with some years in KS1. She trained when the national curriculum was being rolled out. She was currently teaching in Year 6.

#### 6.5.2.3 Teacher F

Teacher F was a Year 6 teachers with 12 years of experience. Their experience has been in KS2 only and they had taught in more than 1 West Midland Primary school.

#### 6.5.2.4 Head teacher and Deputy Head teacher Setting 1

The head teacher had been appointed to the school three years previously, subsequently appointing the DHT. The DHT was assessment lead and a lead moderator for the local authority.

### 6.5.3 Interview participant demographics Setting 2

Setting 2 was a three-form entry Voluntary aided (maintained school run by a charitable foundation) Junior school in an affluent area of the West Midlands. In 2017, students achieved above the national average, of 61%, with 77% meeting the expected standard in Maths, Reading and Writing. See Appendix 8 for more pupil demographic information.

#### 6.5.3.1 Teacher G

Teacher G was in her third year of teaching at school 2 following an appointment as an NQT. She had only taught in Year 6, but had experience teaching in Years 2 and 5 during her postgraduate diploma in education. She had never taught with levels, and only had her own

experience of being assessed by them. She was also completing a Master's in teaching and learning during that academic year.

#### 6.5.3.2 Teacher H

Teacher H had been teaching for 7 years, and it was his second career, qualifying later than most of the other teachers in the study. He had taught at one school previously to Setting 2, where he completed his MTL. As a class teacher he had experience teaching in year 2, 4 and 6. He had taught at Setting 2 for 4 years. At the time of the interview he was recently appointed as a maths lead where he works with class teachers to team teach and develop maths across the school, as well as at the associated infant school.

#### 6.5.3.3 Assistant Head teacher and assessment lead Setting 2

AH2 had been teaching for 9 years. The HT was her mentor during her final year of her teaching practice at Setting 2, where she was given a position for her NQT year. She taught at Setting 2 for three years, two of which were in year 5, and one in year 3, before taking a leadership position in English at another local junior school. A year later she applied for the assistant head position at Setting 2, and has held the position for 3 and half years, taking on assessment lead 2 years ago. She currently teaches sets in year 6 for English and maths. She also had a role liaising with the infant school associated with Setting 2 to 'improve' their approach to AwLs.

## 6.6 ANALYSIS PROCEDURE

### 6.6.1 Qualitative Analysis procedures and theories

The procedures and theory adopted for the analysis of the qualitative interview data will be described. This will include any procedures and theory adopted including reflections made following the piloting of the interview data analysis.

#### 6.6.1.1 Pilot study- reflections and adaptations

*Table 6-4 Reflections and adaptations following the Pilot Study*

Process	Delivery	Content
Interviews	More semi structured in nature, less reliant on focusing only on prompts.	Review semi structured improvised questions to include common ones into the full study like moderation and writing.
Analysis	Thematic through repeated trawls then discussed in topic	Coding in thematic form of topics emerged- expanded my analysis to include social theory to discuss the themes that emerged.

Upon carrying out my pilot interviews, I quickly realised how important writing assessment was as an issue but had not planned to include it in my research. Writing was not part of the initial interview schedule as the original aim of the research focused on subjects that were both tested and teacher assessed. With every teacher in the pilot interviews mentioning it outside of my semi-structured questions, main study questions were adapted to include the assessment of writing. A set of questions was also added to the research survey, but the previous participants had not completed these questions so Q10 was omitted from the data. This highlighted an impact of AwLs which I had omitted when originally planning to focus on only the subjects which had a written test. However, pilot study interview participants all referred to the teacher assessment of writing therefore it was added to the main study interviews.

An important part of piloting the interview data was trailing the analysis, it was important to analyse and report the data before main study data collection, as the pilot stage was part of ensuring my main study data was collected and analysed in a systematic, credible and trustworthy manner. In regards to my analysis, the thematic coding of the pilot study

produced a set of themes developed with relationships established between them. Although it was not assumed themes would be reflected in the two main study settings they provided a frame of reference. Importantly, piloting my interview data collection method and analysis enabled me to test out the process, and then go deeper into the transcripts beyond what the teachers said by analysing what it may mean when they say something (semantic and latent themes receptively). For this my theoretical framework lacked the language I needed to discuss the data I was finding. Social theory provided ‘thinking tools’, supplying language in the form of Bourdieu’s notions of *habitus*, *field*, and *illusio*, and from Foucault that of *disciplinary power*. This does not mean that the Primary themes which emerged were redundant, rather that the social theory perspective enabled an extra dimension explaining why Primary themes had the impact they do on the teachers. This produced questions about the agency of teachers in their assessment and teaching practices and why they complied with practices they disagreed with.

#### 6.6.1.2 Main Study: Qualitative analysis procedures and theories

A thematic approach was taken to interview data analysis. Braun and Clarke (2014) draw attention to criticism given for this approach critiqued for lacking of robustness or generalisability. Measures in place to ensure credibility and trustworthiness of this research are discussed in more detail in Validity, reliability, credibility, and trustworthiness.

##### 6.6.1.2.1.1 Thematic Analysis

A reflexive thematic analysis was adopted, centred on work by Braun and Clarke (2006, 2013, 2021). This mode of analysis provided a flexible but systematic approach where themes could emerge from the data; where ‘A theme captures something important about the data in relation to the research question’ (Braun and Clarke, 2006: 82). As themes were formed through my interpretation of the data, my positionality in the research cannot be ignored, thematic analysis acknowledges this:

*Themes cannot exist separately from the researcher—they are generated by the researcher through data engagement mediated by all that they bring to this process (e.g. their research values, skills, experience and training). (Braun and Clarke 2021: 39)*

Similarly, Charmaz (2019: 165) attributes benefits of embracing the researcher’s own positionality, allowing them to ‘...concentrate on what is happening in the research field, acknowledge that they are part of it, remain flexible, follow empirical events, attend to language and meaning...’. Although themes were not decided in advance it is acknowledged that themes were anticipated from the process of conducting the interviews, repeated listening during the transcription process, exposure to emergent themes in the literature and my own experience. However, themes were not confined to those anticipated and no list of themes were created until they emerged from the data during coding ‘...rather than starting with the theory (as opposed to positivism), inquirers generate or inductively develop a theory or pattern of meaning’ (Creswell, 2009: 8)

*As in the doing of qualitative analysis, our subjectivity informs the process of conceptualising and mapping qualitative research. (Braun and Clarke 2021: 44)*

In adopting this systematic approach to data analysis I was able to include reflection in the process.

*Our reflexive approach involves six—recursive—phases of: familiarisation; coding; generating initial themes; reviewing and developing themes; refining, defining and naming themes; and writing up. (Braun and Clarke 2021: 39)*

**Table 1** Phases of thematic analysis

Phase	Description of the process
1. Familiarizing yourself with your data:	Transcribing data (if necessary), reading and re-reading the data, noting down initial ideas.
2. Generating initial codes:	Coding interesting features of the data in a systematic fashion across the entire data set, collating data relevant to each code.
3. Searching for themes:	Collating codes into potential themes, gathering all data relevant to each potential theme.
4. Reviewing themes:	Checking if the themes work in relation to the coded extracts (Level 1) and the entire data set (Level 2), generating a thematic ‘map’ of the analysis.
5. Defining and naming themes:	Ongoing analysis to refine the specifics of each theme, and the overall story the analysis tells, generating clear definitions and names for each theme.
6. Producing the report:	The final opportunity for analysis. Selection of vivid, compelling extract examples, final analysis of selected extracts, relating back of the analysis to the research question and literature, producing a scholarly report of the analysis.

(Braun and Clarke, 2006: 87)

Through transcribing the data in phase 1, I familiarised myself with the contents. Interviews were transcribed by me, an important first step in familiarising myself with the data (Braun and Clarke, 2016). Following the pilot study, the decision was made to edit transcripts for

meaning removing ‘...idiosyncratic elements of speech...stutters, pauses, non-verbals, involuntary vocalizations...’ (Oliver et al., 2005: 1274) where their presence distracted from the inferred meaning derived from the participant. This was because where participants validated their transcripts for meaning, they focused on the ‘idiosyncratic elements’ (ibid) rather than the information they wanted to convey.

Interview transcripts were then read and coded in phase 2 where ‘a feature of the data...appears interesting to the analyst,’ (Braun and Clarke, 2006: 88). Codes differed from themes as they lack the breadth of a theme. For example: codes such as ‘reliance on ITAFs’ and ‘Late Assessment Guidance’ contributed to the theme of ‘Stressful transition’ as it encompasses the codes relationally.

In phase 3 both semantic and latent themes emerged. An example of the more explicit (semantic) themes were ‘Workload’, ‘SATs pressure’, ‘Stressful transition’. The more implicit (latent) themes which emerged were, for example; ‘Lack of Agency’, ‘Performative pressures’ and ‘Education viewed as a commodity’. The examples provided here were viewed as working titles until all transcripts were re-read and analysed as outlined by Braun and Clarke (2006). ‘Hypotheses’ (Denscombe, 2014: 288) were developed with an increasing use of memos, notes and labels used particularly to search for more understanding from the data as ‘...the potential for codes to evolve to capture the researcher's deepening understanding of the data...’ (Braun and Clarke, 2021: 39). For example, questions about assessing what we value or valuing what we measure? If greater depth is so important what is the best way to assess it? A test? Or using teacher assessment? Is greater depth just seen as independent?

Where themes were identified it was important that they recurred, although Braun and Clarke (2006) do not equate how important the theme is to its frequency.

*...the ‘keyness’ of a theme is not necessarily dependent on quantifiable measures, but rather on whether it captures something important in relation to the overall research question. (Braun and Clarke, 2006: 82)*

As such my subjectivity is entwined with the data through its analysis, as Braun and Clarke (2006: 94) argue:

*Thematic analysis is not just a collection of extracts strung together with little or no analytic narrative. Nor is it a selection of extracts with analytic comment that simply or primarily paraphrases their content. The extracts in thematic analysis are illustrative of the analytic points the researcher makes about the data...*

Thematic analysis therefore aligns with the paradigm of research in allowing subjectivity to be part of the knowledge produced.

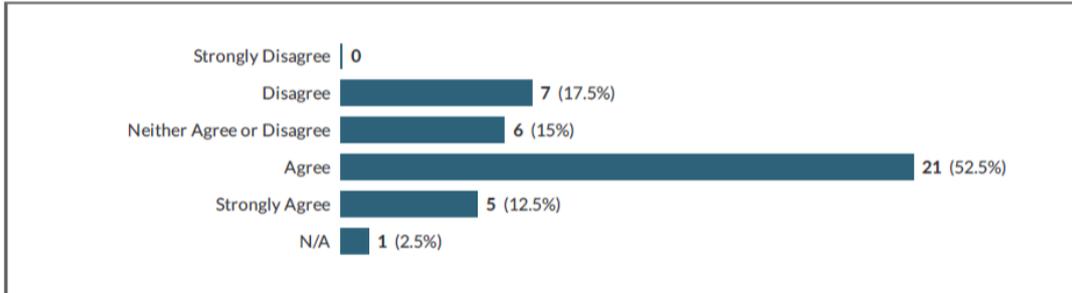
#### 6.6.2 Quantitative analysis procedures

Descriptive statistics have been used to analyse the survey data from all participant responses (all survey data), and then additionally to identify any interesting differences in responses between certain participant demographic groups. All survey data were analysed through Jisc Online surveys, with the number of responses to each item on the scale for statements within Q7-9 converted into a percentage of the total responses for the question. This is shown below in Figure 6-4.

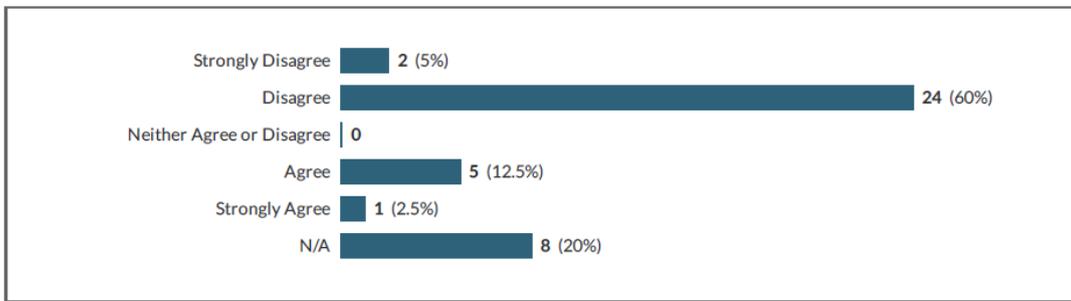
Figure 6-4 Jisc Online surveys presentation of survey response for each statement in Q7-9

**7** These questions relate to Teachers' Summative assessment of Maths and Reading.

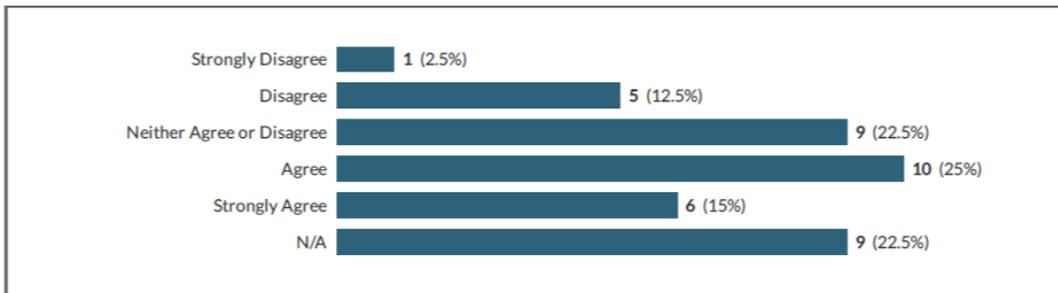
**7.1** I feel I can confidently assess WITHOUT levels.



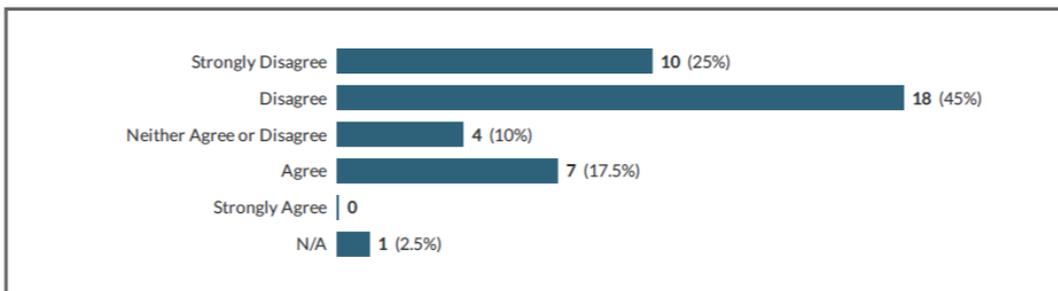
**7.2** There has been an easy transition from assessing WITH levels to without levels.



**7.3** Assessing WITH levels was a better way of tracking children's progress.

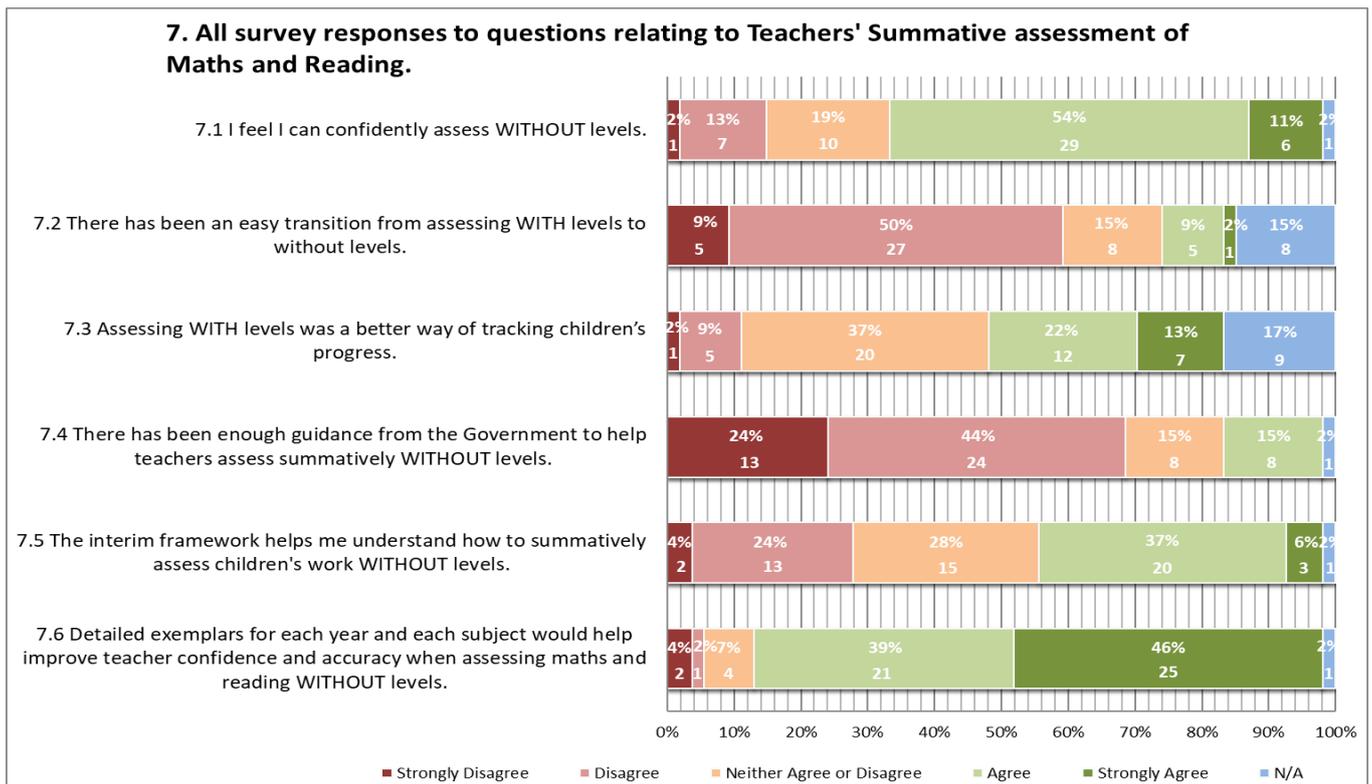


**7.4** There has been enough guidance from the Government to help teachers assess summatively WITHOUT levels.



Data from these responses were then grouped in each overarching question (See [Online Survey](#)). Microsoft Excel was used to condense this data into a bar chart able to display the Likert responses to all statement in each survey question as shown in Figure 6-5. Each bar represents a statement from the question with colour coded and labelled sections directly representing the responses shown above in Figure 6-4.

Figure 6-5 Example of survey data presentation for Q7-9.



Likert-type responses were not treated as interval data as they are ordinal in nature therefore, data from the scales have not been quantified numerically, as Cohen et al. (2008) cautions users may not infer the same strength of difference in feeling between responses such as strongly agree, and agree, and strongly disagree, and disagree. From this analysis of all participant data, findings are reported descriptively in terms how they are representative of the teachers who partook in the survey.

These data were then analysed according to different independent variables based on the participants' demographic information to add description to the responses of all participants. These independent variables (IV) have been analysed in terms of the effect they had on the percentage of agreement responses. Each statement within a question is treated as a dependent variable (DV). The independent variables selected were:

- Experienced assessment with levels: Taught with levels (16) or not taught with levels (38)
- Which Key Stage teachers taught in: KS1 teacher (13) or KS2 teacher (40)
- Teaching in a year with SATs tests: Teaching in a SATs year (18) or Was not teaching in a SATs year (31)
  - Within SATs years a further analysis of the independent variable of SATs used for league tables year: Teaching in Yr6 (14) or teaching in Yr2 (10)

Each independent variable had two criteria as set out above, these are therefore binary pairs of independent variables.

Why each IV was selected is outlined below.

6.6.2.1 What was the difference between responses of participants with experience of NCLs and those without?

The independent variable of participant experience of assessment with levels as well as the independent variable of participants who had not NCLs was selected to explore the relationship these IVs had on survey responses. Analysis of the interview data revealed views suggesting that the familiarity with levels may have created a bias towards views against assessment without levels or a willingness to hold on to an assessment structure they were experienced with.

6.6.2.2 What was the difference between responses of participants who taught in KS1 compared to KS2?

A filter of KS1 and KS2 teachers was applied as this is a common differentiation within schools where KS1 is the 'infant' part of the schools and KS2 is the 'junior'. It is important to note that one of the school settings was a junior school, therefore all teachers from that setting were KS2. This will have skewed the survey data towards KS2 responses. The data is represented in both percentage and raw teacher response numbers so the proportion of each response can be compared regardless of IV participant size. Across England, schools contain 4 years in KS2 compared to 2 years in KS1, so there would typically be a larger amount of KS2 teachers. As previously mentioned, the priority of this data analysis was to

gain the views of all teachers within the settings, therefore it was not the aim to produce equal teacher representation according to national teacher demographics.

6.6.2.3 What was the difference between responses of participants teaching in a year with SATs assessments and those who did not?

Teachers in SATs years have external testing, therefore this variable explored if this affected the views of these respondents. This variable was used to create set of binary variables exploring the relationship between teaching in teacher SATs years and not. SATs in Yr6 weigh heavily for the school in terms of accountability and informs their ranking on school leader boards. Results of these tests additionally influence the frequency and result of OfSTED inspections. SATs in Yr6 only take into account teacher assessment for writing; maths and reading judgments are based on the tests alone, however teacher assessment is submitted. In year, 2 however, a combination of the test result and teacher assessment can be used to form the judgment. This creates two different SATs environments for these years. Within the interview data there is also a disparity in how the Yr6 teachers and Yr2 teachers discuss their views on the SATs tests.

The other binary independent variable is for those teachers who do not teach in a SATs year. These teachers' views were not the focus of the interviews and the survey enabled their views to be canvased. It is interesting to see how these views may differ when they do not directly have the external pressures of government testing.

6.6.2.4 Identifying interesting differences in the binary independent variable responses  
To highlight interesting places to draw on in more detail from the wider data, each binary-variable data was analysed using the Mann-Whitney U test, a statistical test recommended for the analysis of Likert-type data in social science and education (Cohen et al., 2007; Singh, 2007). The Mann-Whitney U test compares the medians of the two demographic groups' responses to each item on the Likert scale for each question (Cohen et al., 2007; Gorard, 2001; Subedi, 2016; de Winter and Dodou, 2010). The test is based on a null hypothesis that assumes no difference between two sets of data, therefore indicating there is nothing significant to note between the responses of each demographic group (Singh, 2007). This hypothesis is rejected if the significance level, the p value, of the difference between the two groups is less than 0.05 which indicates these differences are statically significant

warranting further descriptive analysis (Cohen et al., 2007; Singh, 2007). The aim is to identify which binary variable comparisons resulted in a  $p \leq 0.05$  therefore suggesting data which is interesting to explore and will add to the wider descriptive analysis of all survey responses. All analysis is still held within the interpretivist methodology, therefore there is no assumption that this data represents any teachers outside those who participated in the research (Cohen et al., 2018).

#### *6.6.2.4.1.1 Analysis process*

The Mann-Whitney U test was carried out on an online tool based on the statistical program R (R-Project, 2021). Microsoft Excel was first used to convert the data from tabular form to frequency, providing each response on the scales with a number. This did not mean that responses were assigned an interval value, rather this provided a numerical label for each response on the scale. This scale went from 1 to 5, with 1 being strongly disagree, and 5 being strongly agree. Table 6-5 shows how this was carried out for each question, for each binary IV.



The frequency of the binary IV responses for each question was then entered on Statskingdom (2021) where a p-value was generated. A p value of less than or equal to 0.05 indicated that there was a statistically significant variation in responses as a result of the different demographic groups, providing a justification for further descriptive commentary on these questions. The results of the Mann-Whitney U test will be presented and explained in Section 7.1.6.

## 6.7 VALIDITY, RELIABILITY, CREDIBILITY, AND TRUSTWORTHINESS

Judgements of the quality of research can be made based on arguments of validity and reliability, with how this applies to qualitative research debated (Corbin and Strauss, 2008). In terms of validity there are two forms. One is external validity- that research data is generalisable if applied 'to other examples of the phenomenon' (Denscombe, 2010: 298)- however this contains the remnants of positivist ideologies requiring generalisable results which can be applied universally. The second is internal validity- the extent to which the research collects and presents the data it proposes to (Punch, 2009). Research is also judged by how reliable it is, how likely if it is replicated that it would gain consistent and replicable results (Punch, 2009). Corbin (in Corbin and Strauss, 2008: 301) felt uncomfortable using the terms validity and reliability in reference to qualitative research, citing they come with '...too many quantitative implications (a personal bias).' As discussed in Validity and Reliability in Chapter 2, considerations of validity are not confined to methodology, but are also present in assessment literature. Instead Corbin and Strauss (2008) preferred the word 'credibility',

*...[which] indicates that findings are trustworthy and believable in that they reflect participants', researchers', and readers' experiences with the phenomenon but at the same time the explanation is only one of many possible 'plausible' interpretations possible from data. (ibid: 302)*

Referencing previous work by Glaser and Strauss (1967), credible research should meet certain criteria:

*The first is that there be sufficient detail and description so that readers feel that they were vicariously in the field (thus able to judge for themselves). Second, there should be sufficient evidence on how the data were gathered and how the analysis was conducted (so that readers can assess how the researcher came to his or her findings or conclusions)... Finally, the researcher should specify the kinds of data upon which his or her interpretation rests. (Corbin and Strauss, 2008: 300)*

This criterion has been adopted to judge the credibility of this research and thus its trustworthiness. The first point is demonstrated by a comprehensive review of the literature concerning assessment in Primary school, chapters are included to define and explore

interpretations of policy and teacher practice in terms of theoretical lens, and my positionality as the researcher is made explicit and detailed in the first half of this chapter. The process undertaken to pilot this research and provide traceable data collection and analysis methods supported by relevant theoretical underpinnings provides evidence of their second point. Further, the effect of my own experience and views as a teacher, previously, are explored below. This chapter has set out the data collected of which interpretations will be based, when combined with the results chapter the third point of Glaser and Strauss (1967) is evidenced.

For research to be trustworthy it also requires rigour. Piloting the study before the main data collection enabled me to test out my research methods, technique and analysis. As discussed previously in this chapter this enabled me to; refine my research approach to suit the participants, build experience and confidence when interviewing and develop my use and understanding of theory providing me with vocabulary and concepts for the analysis of my data.

For the quantitative analysis, additional analysis using nonparametric statistics were used to indicate relevant areas within survey responses for further descriptive analysis as set out in Quantitative analysis procedures. This adds rigour to the analysis of the binary IV comparisons of responses to the Likert-type responses, whose purpose was to add to the overall descriptive statistical analysis of all survey participant data (Cohen et al., 2007; Gorard, 2001; Subedi, 2016; de Winter and Dodou, 2010).

#### 6.7.1 Awareness of one's own bias

Research bias is a common criticism of qualitative research (Denzin and Lincoln, 1998), however, as Yin (2009) points out, research from a positivistic point of view, quantitative in nature, can also be to subject bias. An example of this could be omission bias or selection bias; only selecting participants or elements of the research area which are likely to provide a positive result. For qualitative research Yin (2009) recommends a strict systematic procedure to ensure that bias does not have a negative effect on the research.

To minimise the impact of bias I have been explicit about my potential for bias and attempted to accept my subjectivity so that I can best separate my views from those which immerge from the data. This research accepts the subjectivity of one's view of the world

and agrees with Creswell and Miller (2000) that mine, as a researcher, should be made as clear as possible. The second is the use of triangulation in data analysis and the third is the researcher providing a research trail being transparent about the decisions they made and why they made them.

Holloway and Wheller (2002: 8) explain that ‘...researchers are not divorced from the phenomenon under study. This means reflexivity on their part; they must take into account their own position in the setting and situation, as the researcher is the main research tool.’. They point out that the researcher will never be able to remove all bias from their interpretations and research actions as they are naturally going to be part of the ‘phenomenon’ they are researching. Additionally, they highlight the importance of the researcher to identify their potential bias. I have been clear about my practitioner background to facilitate an understand of my positionality in relation to the research (Creswell and Miller 2000; Creswell, 2009; Norris, 1997). It is, however, important to also acknowledge the benefits of my experiential knowledge. My professional experience affords an informed and situated insight into my inquiry of Primary assessment observed from first-hand experience. Miles and Huberman (1994: 17) express that knowledgeable researchers can;

*...see and decipher details, complexities, and subtleties that would elude a less knowledgeable observer. We know some questions to ask, which incidents to attend to closely, and how our theoretical interests are embodied in the field. Not to “lead” with your conceptual strength can be simply self-defeating.*

As an outsider to the interview setting I was aware that I could ‘...create social behaviour in others that would not have occurred ordinarily.’ (Miles and Huberman, 1994: 264). This itself is a threat to the credibility of the research. Any influence created by ‘biological attributes’ such as my gender, age and background practically could not be avoided however I made ‘...efforts to be polite and punctual, receptive and neutral, in order to encourage the right climate for an interviewee to feel comfortable and provide honest answers.’ (Denscombe, 2014: 190)

To minimise the effect I had participants when viewed as an outsider and gain the most authentic responses, I introduced myself as a ‘knowledgeable’ outsider, briefly sharing my

history as a teacher and my recent transition into research. I also followed advice from Denscombe to present myself;

*...in a light which is[was] designed not to antagonize or upset the interviewee (conventional clothes, courtesy, etc.)...[and] remain(s) neutral and non-committal on the statements made during the interview by the interviewee. (Denscombe, 2014: 191).*

Piloting the qualitative part of my study gave me the opportunity to be exposed to an interview setting as an interviewer, building the skills needed to determine when to stay silent, when to spur on and when to prompt (Denscombe, 2014). It also afforded me time to reflect on my assumptions and interpretations of the topic before the main study.

In writing my methodology chapter I have already begun this process of identifying my potential bias (Creswell, 2009; Creswell, and Miller 2000; Norris, 1997). The structure of my PhD also ensures frequent contact with experienced researchers in my field who can challenge my bias and point of view which strengthens my own reflection; this is in the form of critical colleagues, supervisors, feedback from conference presentations and continued reading into the area of research. Norris (1997: 174-175) describe this as using 'critical friends and colleagues' to 'explore their preferences for certain kinds of evidence, interpretations and explanations and consider alternatives'.

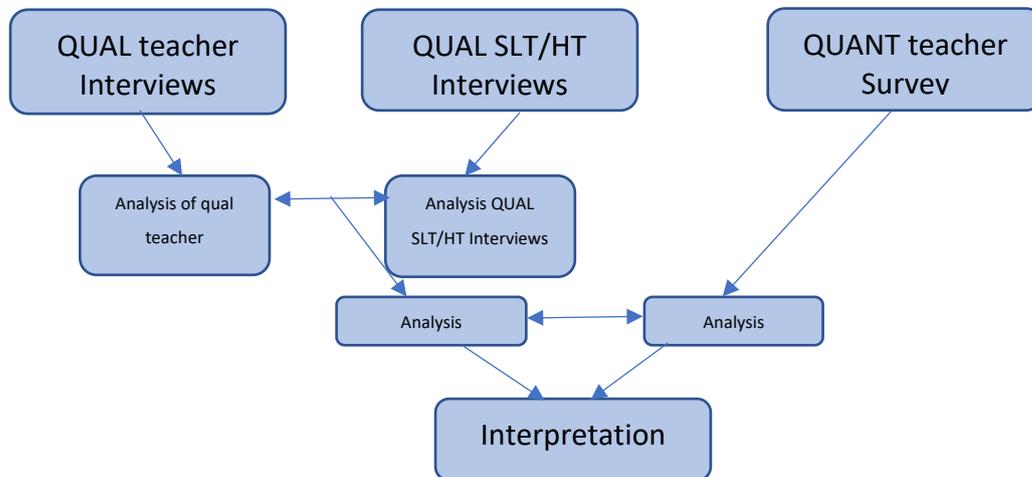
### 6.7.2 Triangulation

Triangulation has an important role in supporting the credibility and trustworthiness of the research. Using mixed methods enables triangulation of both qualitative and quantitative data. A concurrent triangulation strategy has been adopted, the data were collected at the same time but analysed separately, before being drawn back together in the discussion (Creswell, 2009: 213). Yin (2009: 98) states that triangulation of mixed methods (methodical triangulation) ensures more accurate interpretations of data by combining the strengths of quantitative and qualitative research, and that findings are,

*... much more convincing and accurate if it is based on several different sources of information following a corroboratory mode.*

The Figure 6.7 shows at which stages triangulation occurred between the two main data sets of quantitative survey and qualitative interviews.

Figure 6-6 Flow chart showing concurrent triangulation strategy for data analysis, adapted from Creswell (2009)



Denzin (1989: 310) states that this approach ‘involves a complex process of playing each method off against the other so as to maximise the validity of field efforts.’ Miles et al. (2014: 300) says ‘triangulation is not so much a tactic as a way of life.’ This indicates that triangulation should be seen as the analysis and not just part of the analysis.

## 6.8 CONCLUSION

This chapter has set out the methodological design of my research and my positionality as a researcher. Methods have been introduced and justified and critically analysed while considering their appropriateness to answer my research questions for example, adopting an interpretivist methodology which holds the participants’ experiences of AwLs central to their understanding of reality. Ethical considerations were identified, alongside practices adopted to mitigate the risks especially due to the small-scale nature of this research. How I maximised the trustworthiness and creditability of the research through transparency and rigor, has been argued by providing a about my paradigm, methods and data analysis procedure.

The next chapter will be the first of the two data findings chapters.

## 7 SURVEY DATA FINDINGS

---

Core question findings (Q7-Q9) are described and discussed, first as a whole, then in terms of the different binary independent variables effects on responses to questions. Survey text responses are included in this section.

### 7.1.1 Abbreviations and coding of participant demographics.

For text responses where applicable, survey participants have been referenced using a code relating to demographic information:

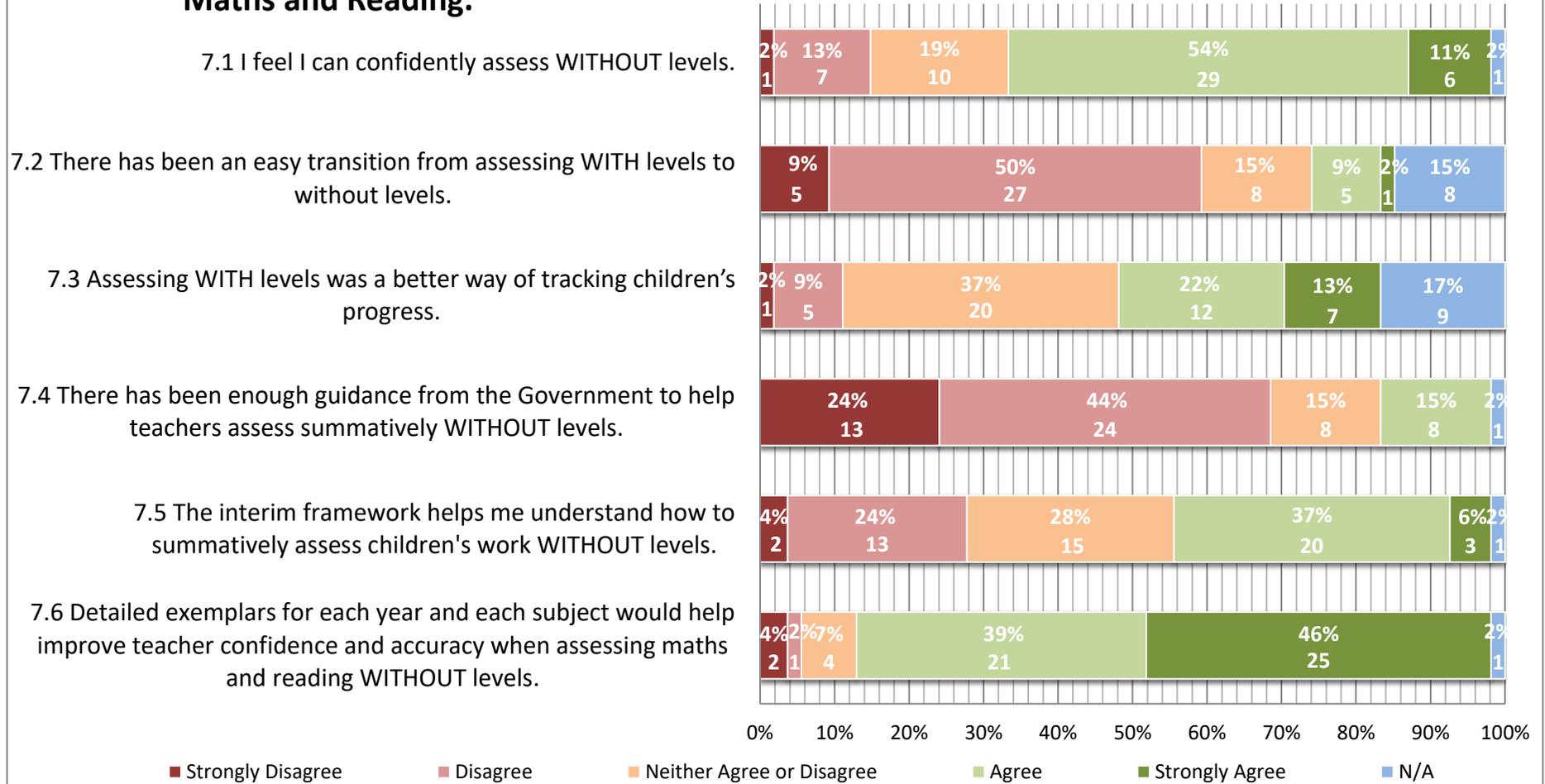
Respondent number: Key Stage teaching in: Year teaching in: NCLs (WL) or without levels (WOL): number of years' experience. For example; 3:KS1:Y2:WL:11-15Y is read as respondent number 3, who teaches in key stage 1 in Yr2, they have assessed with NCLs, and have 11-15 years teaching experience.

All responses are represented in a bar chart, displaying number and percentage of how participants responded. Additionally, I have provided commentary referencing some of the data in the charts directly. For overall ease of discussion, 'agree' and 'strongly agree' categories are combined into (strongly) agree, and (strongly) disagree unless there is an indication that these categories require distinguishing.

### 7.1.2 Question 7- All survey responses to questions relating to teachers' summative assessment of Maths and Reading

These questions explored teacher summative assessment in core maths and reading, aiming to gather teacher opinions on their confidence of using AwLs, the transition of changes and government guidance and support.

## 7. All survey responses to questions relating to teachers' Summative assessment of Maths and Reading.



The majority of participants (strongly) agreed (65%) to Q7.1 reporting that they felt confident with AwLs, however, the majority also (strongly) disagreed (59%) to Q7.2, that it was an easy transition towards AwLs, indicating that although most felt confident now, the process had been difficult. More participants (strongly) agreed (35%), than (strongly) disagreed, (11%) to Q7.3, that levels were a better way of tracking children's progress with 37% responding NA:DA.

Over 68% of participants (strongly) disagreed to Q7.4, that there had been sufficient government guidance for summative assessment of maths and reading without levels. Regarding the guidance in place for end of KS1 and KS2, the ITAFs (interim assessment frameworks), 43% (strongly) agreed to Q7.5, that it was helpful for summatively assessing reading and maths without levels with over 25% (strongly) disagreeing that this was the case.

Overwhelmingly in response to Q7.6, 85% of participants (strongly) agreed said that detailed exemplars would improve their confidence and accuracy with teacher SA without NCLs.

#### 7.1.2.1 Question 7- text responses

There were 9 text responses for question 7. The themes that emerged in relation were confusion, assessment guidance, suitability of assessment approaches and challenges of reading assessment.

Confusion within schools was raised by one respondent, 'Schools seem very confused by the new system' (3:KS1:Y2:WL:11-15Y) however they also stated that 'AwLs fits the NEW curriculum.' Another respondent stated that teachers' summative assessment was 'Very school dependent and not consistent.' (19:KS1&KS2:ALL:WL:16-20Y).

Teacher assessment requirements were described as vague where '...- each requirement covers too much', and that it 'needs to be more flexible.' (28:KS2:Y6:WL:21+Y). One teacher suggested that 'Detailed exemplars for each year and each subject would help improve teacher confidence and accuracy when assessing maths and reading WITHOUT levels-' and that the tracking programme SPTO 'offers an effective framework.' (36:KS2:Y4:WL:21+Y). Another teacher found 'the target tracker levelling... vague in parts, but as I mostly teach Yr6, I feel I can assess them fairly easily using SATs and other information - so I know if

they're where they should be.' (40:KS2:Y5&6:WL:1-5Y). Comments by this participant suggested they felt confident with AwLs, this could be due to the government guidance supplied to Yr6 (and Yr2) that is not available in non SATs years. These teacher's comments indicated that the guidance supplied for teacher SA had not been adequate, and tracking programs were cited by two respondents in relation to SA.

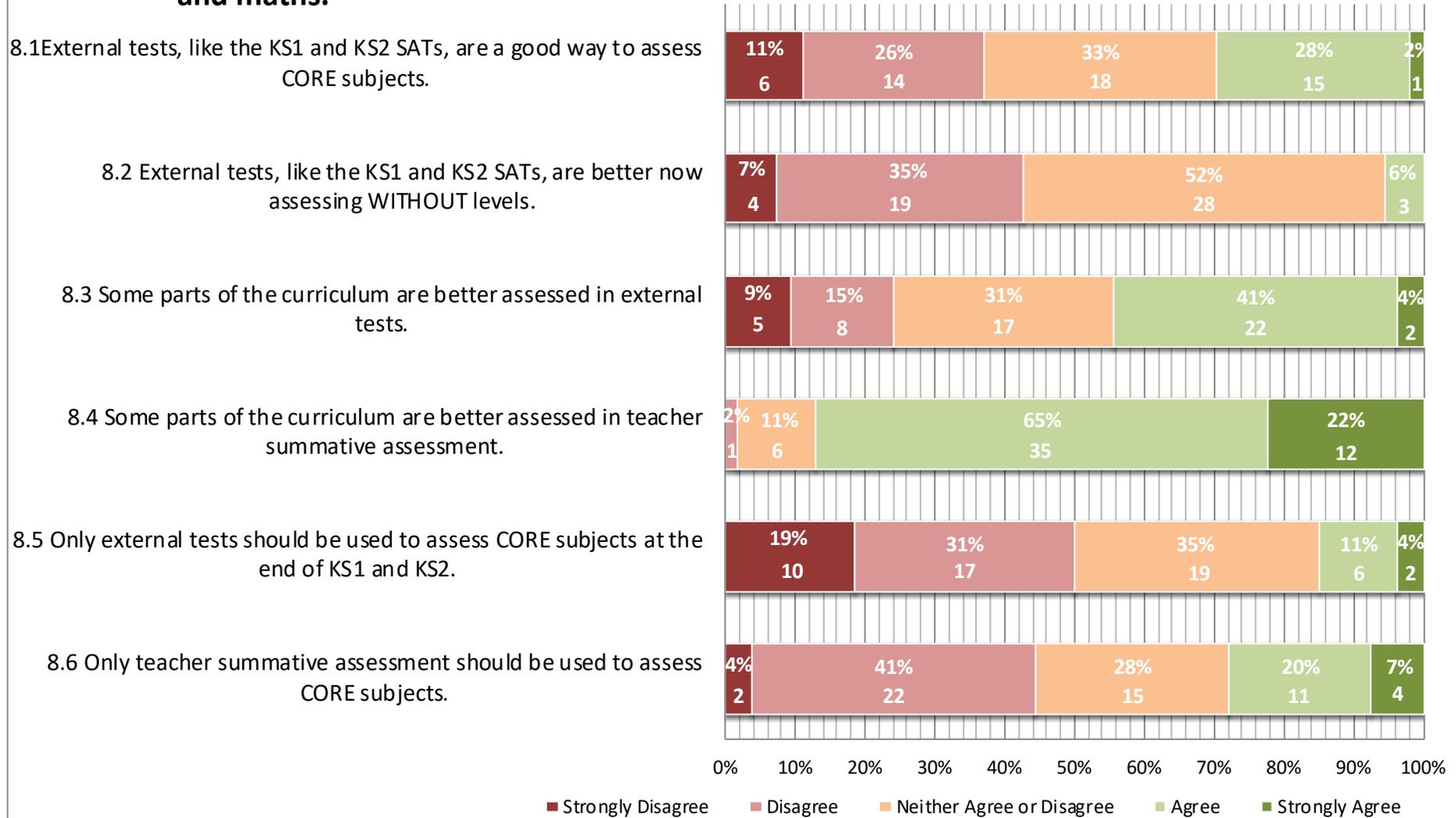
One teacher referred to teacher assessment of writing in this section, commenting it was '...too inexact a science. Teacher SA if it remains needs to be based on comparative judgement and more rigorously moderated. Currently it's a farce.' (21:KS2:Y5:WL:6-10Y). It is unclear if the 'farce' is in relation to the teacher assessment of writing or teacher assessment in general, as there was a previous comment about teacher SA of reading that 'teacher SA of reading is not practical when final summative assessment is a test' (21:KS2:Y5:WL:6-10Y). This teacher's response indicated that they would prefer one type of summative assessment, and do not see teacher assessment and tests as compatible.

Another teacher stated 'Maths is easier to assess than reading' (34:KS2:Y4:WOL:1-5Y), with reading assessment seen as a challenge by another teacher who stated that 'assessing reading accurately can be challenging' (32:KS2:Y6:WL:6-10Y).

### 7.1.3 Question 8- All survey responses to questions relating to External Summative Testing of reading and maths.

These questions relate specifically to the external testing of the core subjects of reading and maths. Participants were questioned about their opinion of external testing compared to teacher summative assessment.

## 8. All survey responses to questions relating to External Summative Tests of reading and maths.



Q8.1 produced mixed responses, with roughly a third responding to each (strongly) disagree, ND:NA and (strongly) agree that external tests were a good way of assessing core subjects, 28% agreeing and 33% neither agreeing nor disagreeing. More respondents strongly disagreed (11%) than strongly agreed (2%) suggesting a stronger feeling against external testing than in support of it.

Tests were not seen to have improved without levels, only 6% agreeing they were better without levels, compared to 42% who (strongly) disagreed. Most teachers neither agreed nor disagreed at 52%. In response to Q8.3, 45% (strongly) agreed that some areas of the curriculum were better suited to external tests with 24% (strongly) disagreeing.

Table 7-1 Cross tabulation of survey responses to Q8.1 and Q8.3.

Some parts of the curriculum are better assessed in external tests.	External tests, like the KS1 and KS2 SATs, are a good way to assess CORE subjects.					No answer	Totals
	Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree		
Strongly Disagree	5	0	0	0	0	0	5
Disagree	0	4	3	1	0	0	8
Neither Agree or Disagree	1	5	9	2	0	0	17
Agree	0	5	6	11	0	0	22
Strongly Agree	0	0	0	1	1	0	2
No answer	0	0	0	0	0	0	0
<b>Totals</b>	<b>6</b>	<b>14</b>	<b>18</b>	<b>15</b>	<b>1</b>	<b>0</b>	<b>54</b>

Responses to Q8.1 and Q8.3 were cross tabulated to compare responses in Table 7-1. This was a feature of the Bristol Online Survey platform (now Jisc Online Surveys) which allowed for responses to two questions to be plotted against each other, enabling a comparison to be made based on the Likert responses. It could be reasoned those teachers who agreed to Q8.1, that external tests are a good way to assess core subjects, would also agree that some parts of the curriculum are better assessed in external tests. The responses mostly follow this pattern, with the exception of one teacher who agreed to Q8.1 but disagreed to Q8.3, and 5 teachers who disagreed that external tests were a good way to assess core subjects but agreed that some parts of the curriculum are better assessed in external tests. This data could suggest that these teachers approved of external testing as part of an assessment judgement, but not the whole judgment.

In response to Q8.4, 87% (strongly) agreed that teacher SA was better for of some areas of the curriculum, with only 2% disagreeing and 11% NA/ND.

The 2% of respondents who disagreed to Q8.5, represented the views of one teacher (4:KS2:Y5:WL:1-5Y). They strongly agreed to Q8.3 that external tests are better for some areas of the curriculum and that only SATs should be used at the end of KS1 and KS2. They also agreed, however, that only teacher SA should be used for core subjects. It could be that this teacher is in favour for the SATs judgments based on external examinations and teacher SA.

Table 7-2 Crosstabulation of survey responses to Q8.5 and Q8.6.

Only teacher summative assessment should be used to assess CORE subjects.	Only external tests should be used to assess CORE subjects at the end of KS1 and KS2.					No answer	Totals
	Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree		
Strongly Disagree	1	0	1	0	0	0	2
Disagree	4	10	4	4	0	0	22
Neither Agree or Disagree	1	4	9	1	0	0	15
Agree	2	3	4	0	2	0	11
Strongly Agree	2	0	1	1	0	0	4
No answer	0	0	0	0	0	0	0
<b>Totals</b>	<b>10</b>	<b>17</b>	<b>19</b>	<b>6</b>	<b>2</b>	<b>0</b>	<b>54</b>

Q8.5 and 8.6 asked if only external tests or only teachers' SA should be used for core subjects at the end of KS1 and KS2. 50% (strongly) disagreed that only tests should be used at the end of KS1 and KS2, compared to 15% (strongly) agreeing. Comparatively 45% (strongly) disagreed to Q8.6 that only teacher summative assessments should be used, but more teachers, 27% (strongly) agreed to Q8.6 than Q8.5, indicating more teachers were in favour of teacher SA only rather than external summative tests for core subjects. As there are only two options when summatively assessing, teacher assessment or test, answering in agreement to one should logically follow with a disagreement to the other. Because of this, these two questions were cross tabulated to compare the responses for each question, shown in Table 7-2 above. When both questions were cross tabulated there are 15 respondents who (strongly) disagreed to both questions. This could suggest they do not prefer a singular type of assessment, teacher SA or external test, but a combination instead. There were 3 participants who (strongly) agreed both forms of SA should only be used, it is unclear what was meant by this.

Table 7-3 Crosstabulation of survey responses to Q8.5 and Q8.1.

Only external tests should be used to assess CORE subjects at the end of KS1 and KS2.	External tests, like the KS1 and KS2 SATs, are a good way to assess CORE subjects.					No answer	Totals
	Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree		
Strongly Disagree	5	2	2	1	0	0	10
Disagree	0	6	7	4	0	0	17
Neither Agree or Disagree	1	4	7	7	0	0	19
Agree	0	1	2	3	0	0	6
Strongly Agree	0	1	0	0	1	0	2
No answer	0	0	0	0	0	0	0
<b>Totals</b>	<b>6</b>	<b>14</b>	<b>18</b>	<b>15</b>	<b>1</b>	<b>0</b>	<b>54</b>

Responses to Q8.5 and Q8.1 were also compared, shown in Table 7-3. Most respondents fitted the general trend that when they agreed or disagreed with Q8.1, they agreed or disagreed with Q8.5 respectively. However, 5 respondents agreed to Q8.1 but (strongly) disagreed that external tests only should be used to assess end of KS1 and KS2 core subjects. This could suggest that even though they approved of external tests for SA, they disagreed that they should solely be used. The 2 respondents who disagreed to Q8.1, and subsequently agreed or strongly agreed to Q8.5 is puzzling. If they disagreed that external tests were a good way of assessing core subjects, it is unclear why they would also (strongly) agree that only external tests should be used when assessing core subjects. This could suggest a case of participant confusion when responding to the question.

#### 7.1.3.1 Question 8- text responses

There were 6 text responses to Q8 with themes identified relating to suitability of assessment approach, rise in curriculum expectations, and test positives and negatives.

A few responses viewed SATs and external testing as appropriate in some circumstance and not others supporting the inference above that teachers may have preferred a mixed approach to SA, a blend of teacher SA and test. 'Agree with KS2 SATs, Disagree with KS1 SATs.' ( 5:KS1:Y2:WOL:>1Y) and 'I am in two minds regarding SATs as some children including myself struggle under that sort of pressure.' (7:KS1:Y1:WOL:1-5Y). One respondent disagreed with a mixed approach to SA in SATs but also felt both could be manipulated:

*KS1 SATS not fit for purpose as is teacher SA while KS2 SATS externally assessed...Should both be teacher SA or both external? Both are open to abuse. Unless a school is audited during external tests or moderated during teacher SA they could easily over aid (21:KS2:Y5:WL:6-10Y)*

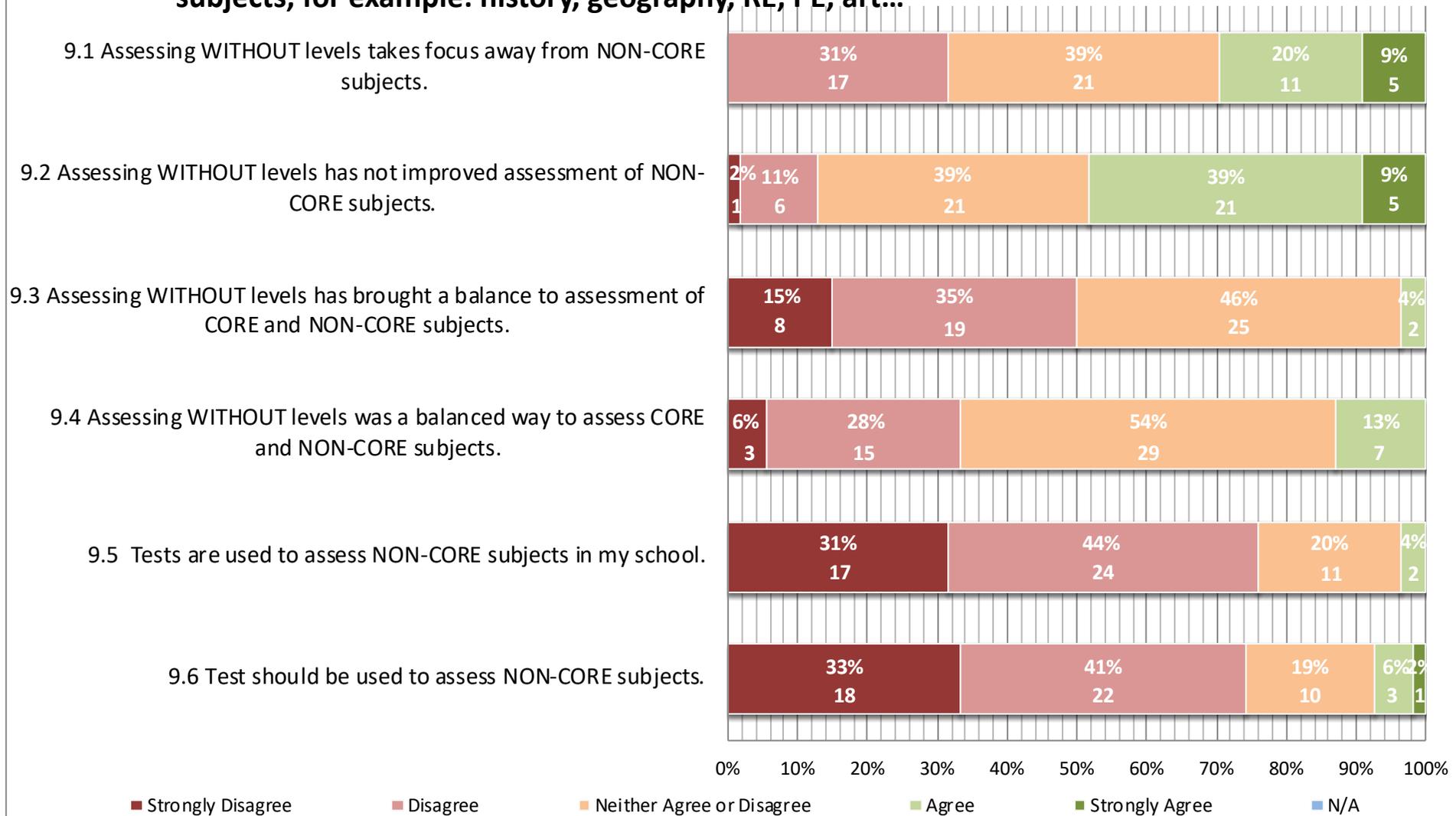
This teacher raises issues concerning high-stakes testing, comparing how 'valid' different ways of assessing are. This suggests they feel that teacher SA and tests assess in different ways, and therefore should not be compared against each other while desiring consistency and shared understanding between schools. They also raise an issue with the validity of

teacher assessment, specifically that accountability pressure could encourage 'cheating', and the rise in floor standards (the raising of difficulty in attainment expected for each year) has increased this. Their criticism of grammar testing was twofold. Firstly, it is inappropriate, and secondly that it does not contribute towards the floor standards attainment judgement. This last point echoes other comments to Q8, that SATs testing has positives and negatives. One respondent commented that 'Tests seem to change the direction of what is being taught in order to teach children how to pass a test'(2:KS2:Y6:WL:11-15Y). Another similarly stated, it 'Can be useful to have external assessment to gauge where they're at, though it does result in a lot of people teaching to the test' (40:KS2:Y5&6:WL:1-5Y), this participant saw both positive and negative implications of external testing. Another participant viewed SATs as useful, 'I think SATs are good prep for further life and show the retention of knowledge' (33:KS2:Y4:WOL:>1Y), this teacher may value SATs for the testing experience it gives children who will be tested throughout their academic career, teachers may also value tests as a more accurate measure of knowledge retention.

#### 7.1.4 Question 9. All survey responses to questions relating to summative assessment of non-core subjects, for example: history, geography, RE, PE, art...

These questions relate to the assessment of non-core subjects. They explore any impact AwLs has had on non-core assessment, and if tests are used to summatively assess as in core subjects

## 9. All survey responses to questions relating to summative assessment of non-core subjects, for example: history, geography, RE, PE, art...



There was a near equal split into thirds in response to Q9.1, with 31% disagreeing, 39% NA/ND, and 29% (strongly) agreeing that AwLs takes focus away from non-core subjects.

In response to Q9.2, 48% (strongly) agreed that AwLs had not improved assessment of non-core subjects. Only 13% (strongly) disagreed, and the remaining 39% of respondents NA/ND. When asked if AwLs had brought balance to the assessment of core and non-core subjects in Q9.3, 50% (strongly) disagreed with only 4% agreeing and a further 46% NA/ND. This indicates an imbalance between the assessment of core and non-core subjects. <sup>1</sup>

---

<sup>1</sup> Question 9.4 will be omitted as I need to be selective about the data being analysed and this question was less relevant than Q9.5.

Table 7-4 The cross-tabulation of survey answers from Q9.1 and Q9.3

Assessing WITHOUT levels has brought a balance to assessment of CORE and NON-CORE subjects.	Assessing WITHOUT levels takes focus away from NON-CORE subjects.						No answer	Totals
	Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	N/A		
Strongly Disagree	0	1	2	3	2	0	0	8
Disagree	0	7	5	5	2	0	0	19
Neither Agree or Disagree	0	7	14	3	1	0	0	25
Agree	0	2	0	0	0	0	0	2
Strongly Agree	0	0	0	0	0	0	0	0
N/A	0	0	0	0	0	0	0	0
No answer	0	0	0	0	0	0	0	0
<b>Totals</b>	<b>0</b>	<b>17</b>	<b>21</b>	<b>11</b>	<b>5</b>	<b>0</b>	<b>0</b>	<b>54</b>

Table 7-4 shows responses to the cross tabulation of Q9.1 and Q9.3. The results show the general trend that those who agreed that AwLs takes focus away from non-core subjects also disagreed that AwLs has brought balance to the assessment of core and non-core subjects. However, 8 participants disagreed that AwLs takes focus away from non-core subjects, but also 8 (strongly) disagreed that it has brought balance to assessment of core and non-core subjects. All but 1 of the 8 who responded this way had taught with NCLs previously. It could be that they viewed assessment in general with or without levels as taking focus away from non-core subjects, and therefore were not attributing this to AwLs as an approach, and did not expect assessment practices to balance assessment of core and non-core subjects.

To keep the survey and its analysis as straight forward as possible the main questions utilised were Likert scales. These disagree and strongly disagree responses will be interpreted as no, tests are not used in their school for non-core subjects and agree and strongly agree that they are. It is unclear if the 'neither agreed nor disagreed' responses were a 'yes' or 'no' or a 'do not know'. Considering the data this way 75% of participant's (strongly) disagreed to Q9.5, with only 4% agreeing. The 4% (2 respondents) who agreed tests were used to assess non-core subjects were not any of the 8% who (strongly) agreed to Q9.6 that tests should be used this way. The majority of teachers, 74% (strongly) disagreed. Q8.5 and Q9.6 were cross tabulated to see how participants responded to both questions concerning testing for core and non-core subjects

Table 7-5 Crosstabulation of survey responses to Q8.5 and Q9.6.

Test should be used to assess NON-CORE subjects.	Only external tests should be used to assess CORE subjects at the end of KS1 and KS2.					No answer	Totals
	Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree		
Strongly Disagree	5	3	6	3	1	0	18
Disagree	4	10	5	3	0	0	22
Neither Agree or Disagree	1	3	5	0	1	0	10
Agree	0	0	3	0	0	0	3
Strongly Agree	0	1	0	0	0	0	1
N/A	0	0	0	0	0	0	0
No answer	0	0	0	0	0	0	0
<b>Totals</b>	<b>10</b>	<b>17</b>	<b>19</b>	<b>6</b>	<b>2</b>	<b>0</b>	<b>54</b>

Table 7-6 Crosstabulation of survey responses to Q9.6 and Q8.3.

Test should be used to assess NON-CORE subjects.	Some parts of the curriculum are better assessed in external tests.					No answer	Totals
	Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree		
Strongly Disagree	5	1	2	9	1	0	18
Disagree	0	6	7	9	0	0	22
Neither Agree or Disagree	0	1	7	2	0	0	10
Agree	0	0	0	2	1	0	3
Strongly Agree	0	0	1	0	0	0	1
N/A	0	0	0	0	0	0	0
No answer	0	0	0	0	0	0	0
<b>Totals</b>	<b>5</b>	<b>8</b>	<b>17</b>	<b>22</b>	<b>2</b>	<b>0</b>	<b>54</b>

Table 7-5 shows the 4 respondents who (strongly) agreed to Q9.6 that tests should be used for non-core subjects. These were not the same participants who agreed to Q8.5, that tests should be used only for core subjects. The teacher who strongly agreed that tests should be used for non-core subjects comparatively disagreed that they should only be used for core subjects. It could be that these participants would prefer a mixed teacher strongly agreed and external testing format.

To explore this further as shown above in Table 7-6, Q9.6 was cross tabulated with Q8.3, which asked if respondents agreed that some parts of the curriculum were better assessed with tests. Of the 24 responses who (strongly) agreed that some parts of the curriculum were better assessed in external tests, 19 (strongly) disagreed that tests should be used for non-core subjects. This indicates that most of the teachers who agreed to external testing for aspects of core subjects do not believe that testing is appropriate for non-core subjects. Only 3 respondents (strongly) agreed to both questions. It could be that teachers are used to core-subjects being externally tested and therefore the practice has become normalised, compared to the assessment on non-core subjects.

#### 7.1.4.1 Question 9- text responses

There were 5 text responses from Q9. Themes emerged regarding over testing and rise in curriculum expectations.

One respondent reiterated the general trend for Q9 that non-core subjects should not be tested, they state 'Children are tested enough on core- I don't agree they should test non-core.' (12:KS2:Y6:WL:21+). Another response questioned the use of a test for just generating a result, 'Testing can be counterproductive if all that is used for [is] a result- teacher assessment must also be part of it.' (18:KS1:Y1:WL:11-15Y), this indicates that this teacher's view of assessment goes beyond summative evaluative uses, and a combination of teacher SA and test should be used. Formative assessment was mentioned by another teacher, who commented 'Teachers should use formative assessment before moving to the next L.I. [learning intention]' (36:KS2:Y4:WL:21+Y). The raised difficulty in curriculum content is to blame for attention being redirected away from non-core subjects, (21:KS2:Y5:WL:6-10Y) for this teacher NCLs has not taken the focus away from non-core subjects, but the higher

expectation in the curriculum with the 'curriculum being pitched beyond maturity of many Primary children' (21:KS2:Y5:WL:6-10Y).

#### 7.1.5 Question 11- any other comments

The last question asked if the survey had covered all areas of AwLs the teachers felt was important. If they answered 'no' they were asked to give a reason. This provided 5 text responses. The majority of these responses related to the impact of assessment on pupils or teachers. Two responses specifically mentioned an emotional impact on pupils, 'I would like to discuss the impact assessment has on the children's experiences in the classroom- stress, pressure, the value of what is taught.' (2:KS2:Y6:WL:11-15Y), a view shared by this teacher 'It[the survey] hasn't covered the emotional [impact] and stress involved for staff and pupils.' (19:KS1&KS2:ALL:WL:16-20Y). This teacher also expressed downsides to AwLs, stating it is 'Too time consuming and does not benefit the children. Needs to be more effective and efficient.' (28:KS2:Y6:WL:21+Y); the impact for this teacher is twofold, the impact on their time and the impact on the children's learning.

This respondent discusses a consequence of a reduced assessment framework;

*Private companies have made a killing trying to fill the void left by the DfE when the ill thought through switch was made. Schools were left floundering. Many of those private companies just re-created levels by a different name completely undermining the whole process. (21:KS2:Y5:WL:6-10Y)*

Describing this re-creation of levels by external companies as undermining the assessment change, suggests this teacher viewed potential in AwLs, which was unrealised because of the 'ill thought out switch' during transition. AwLs was seen as a positive for parents by one teacher's answer, 'Parents- I think parents want to know whether children are where they should be or not. AwLs does this.' (3:KS1:Y2:WL:11-15Y).

#### 7.1.6 Further analysis via demographic information

This section of the analysis provides extra descriptive commentary of survey responses between different demographic groups as set out in [Quantitative analysis procedures](#). This analysis and its discursive elements continue to view the data within the bounds of the

participants, and are not argued in terms of their generalisability to a wider population of teachers. To start with, the results of the Mann-Whitney U tests on the data will be presented to show which question statements for each demographic groups were indicated to be different enough to warrant further commentary and description of the survey data. The results of the Mann-Whitney U-Test will be presented for each question 7, 8, and 9 respectively. As in Table 7-7, p values for each demographic group for each statement are displayed. For ease of viewing  $p \leq 0.05$  are indicated in green, indicating there were differences between responses for this question related to the demographic group of interest. The smaller the p number the more significant the result. It is these results which will then be followed by a summary of what this adds to the complete survey data analysis above.

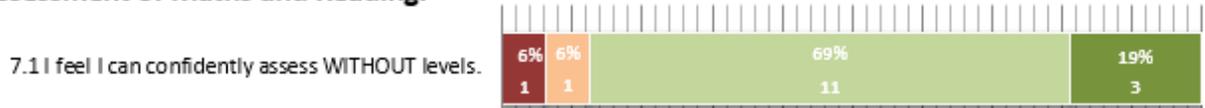
7.1.6.1 Further analysis of responses to Q7 according to binary IV

Table 7-7 Results of Mann-Whitney U test on participant responses from binary IV for Q7

Survey Question	Mann-Whitney U test result for binary variables based on participant demographic data					
	Experience with NCLs compared to no experience with NCLs		KS1 responses compared to KS2 responses		SATs year responses compared with non-SATs year responses.	
	p-value (p≤0.05)	U score	p-value (p>0.05)	U score	p-value (p>0.05)	U score
7.1 I feel I can confidently assess WITHOUT levels.	0.03866	393	0.98140	252.5	0.2238	218.5
7.2 There has been an easy transition from assessing WITH levels to without levels.	0.14320	213.5	0.55460	167	0.7835	209
7.3 Assessing WITH levels was a better way of tracking children’s progress.	0.03352	80.5	0.76230	192	0.4977	215
7.4 There has been enough guidance from the Government to help teachers assess summatively WITHOUT levels.	0.03071	388	0.58930	277.5	0.5014	300
7.5 The interim framework helps me understand how to summatively assess children's work WITHOUT levels.	0.496	330	0.03136	350.5	0.0003	110
7.6 Detailed exemplars for each year and each subject would help improve teacher confidence and accuracy when assessing maths and reading WITHOUT levels.	0.13440	367	0.08855	180	0.1852	319.5

Responses to statements in Q7 yielded the most significant differences between demographic groups who had or had not taught with NCLs, the key stage taught in and whether participants taught in a year with SATs. This provides areas of interest within Q7 general responses related to Q7.1, Q7.3, Q7.4 and Q7.5 as indicated in Table 7-7 by the green p values.

**7. NOT assessed with levels responses to questions relating to Teachers' Summative assessment of Maths and Reading.**

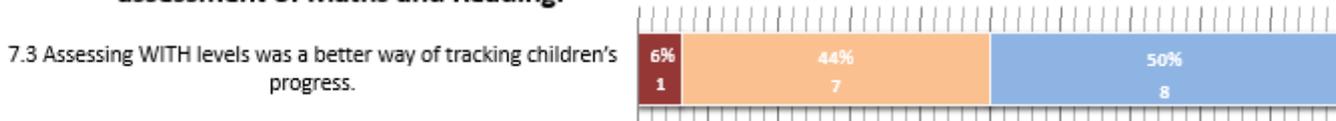


**7. Assessed with levels responses to questions relating to Teachers' Summative assessment of Maths and Reading.**

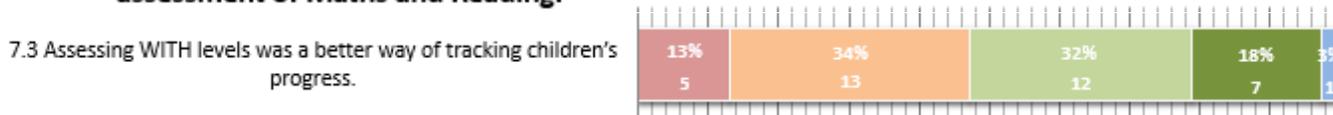


Differences in responses between those with and without NCLs experience were indicated as significant for Q7.1 ( $p=0.03866$ ). Among the participants, those who had not assessed with NCLs expressed more confidence with AwLs where 88% (strongly) agreed to Q7.1, compared to 55% (strongly) agreeing of participants who had assessed with NCLs previously.

**7. NOT assessed with levels responses to questions relating to Teachers' Summative assessment of Maths and Reading.**



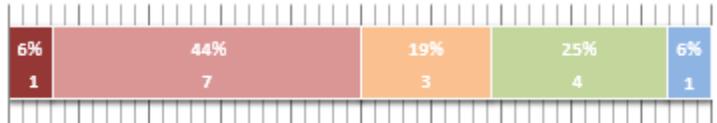
**7. Assessed with levels responses to questions relating to Teachers' Summative assessment of Maths and Reading.**



Participants' experience of NCLs made a difference in responses to Q7.3 ( $0.03352$ ) where 50% of those who had assessed with NCLs (strongly) agreed NCLs was a better way of tracking progress.

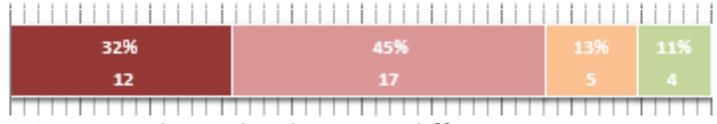
**7. NOT assessed with levels responses to questions relating to Teachers' Summative assessment of Maths and Reading.**

7.4 There has been enough guidance from the Government to help teachers assess summatively WITHOUT levels.



**7. Assessed with levels responses to questions relating to Teachers' Summative assessment of Maths and Reading.**

7.4 There has been enough guidance from the Government to help teachers assess summatively WITHOUT levels.



Whether participants had experience with NCLs was indicated as having a difference on responses to Q7.4 (0.03071). Where participants had experience of NCLs, 77% (strongly) disagreed that there had been enough guidance for teachers to assess without levels, compared to 50% of those who had not assessed with NCLs. 25% of participants who had not assessed with NCLs agreed that there had been enough guidance, compared to 11% of those with NCLs experience.

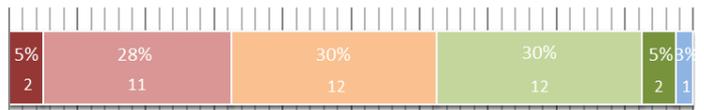
**7. KS1 responses to questions relating to Teachers' Summative assessment of Maths and Reading.**

7.5 The interim framework helps me understand how to summatively assess children's work WITHOUT levels.



**7. KS2 responses to questions relating to Teachers' Summative assessment of Maths and Reading.**

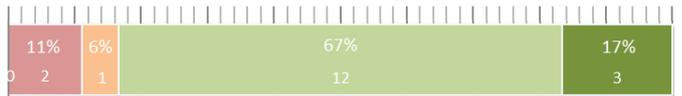
7.5 The interim framework helps me understand how to summatively assess children's work WITHOUT levels.



Data from Q7.5 suggested that key stage participants taught in affected how useful they found the ITAFs (0.03136). This can be seen in the 70% of KS1 teachers (strongly) agreeing, compared to KS2 responses where 33% (strongly) agreed and 35% (strongly) disagreed.

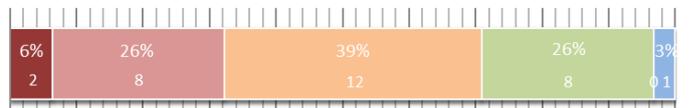
**7. SATs year responses to questions relating to Teachers' Summative assessment of Maths and Reading.**

7.5 The interim framework helps me understand how to summatively assess children's work WITHOUT levels.



**7. Non SATs year responses to questions relating to Teachers' Summative assessment of Maths and Reading.**

7.5 The interim framework helps me understand how to summatively assess children's work WITHOUT levels.



Responses of participants to Q7.5 according to whether they taught in a SATs year or not produced the most significant difference (p=0.0003). This can be seen in a marked

difference in responses, with 84% of participants who taught in a SATs year (strongly) agreeing the ITAFs helped to summatively assess work. This is compared to only 22% agreeing in non-SATs years. Since the ITAFs are guidance for SATs years only, it is not surprising that participants in SATs years agreed they were more useful.

It is also worth noting which questions did not demonstrate significant differences between participant demographic groups. Responses to Q7.2, that there had been an easy transition towards AwLs, was not significantly different between participants regardless of demographic groups. There was also no notable difference identified between NCL experience, key stage or SATs year and the desire for more guidance in the form of exemplars.

7.1.6.2 Further analysis of responses to Q8 according to binary IV

Table 7-8 Results of Mann-Whitney U test on participant responses from binary IV for Q8

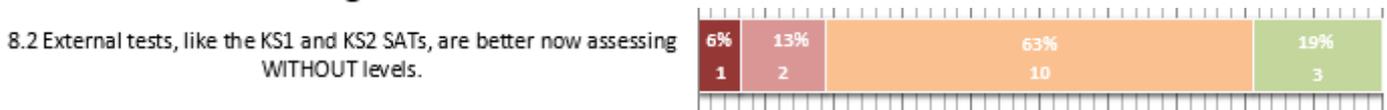
Survey Question	Mann-Whitney U test result for binary variables based on participant demographic data					
	Experience with NCLs compared to no experience with NCLs		KS1 responses compared to KS2 responses		SATs year responses compared with non-SATs year responses.	
	p-value (p>0.05)	U score	p-value (p>0.05)	U score	p-value (p>0.05)	U score
8.1 External tests, like the KS1 and KS2 SATs, are a good way to assess CORE subjects.	0.92150	299	0.09536	337.5	0.897	273
8.2 External tests, like the KS1 and KS2 SATs, are better now assessing WITHOUT levels.	0.00906	428.5	0.65650	279.5	0.5658	304
8.3 Some parts of the curriculum are better assessed in external tests.	0.59620	277.5	0.01364	373	0.5936	303.5
8.4 Some parts of the curriculum are better assessed in teacher summative assessment.	0.00530	428.5	0.03145	171.5	0.09485	210
8.5 Only external tests should be used to assess CORE subjects at the end of KS1 and KS2.	0.88210	296.5	0.20700	318.5	0.8962	273
8.6 Only teacher summative assessment should be used to assess CORE subjects.	0.11060	384	0.03958	165.5	0.7849	291.5

Two demographic groups were indicated, as shown in Table 7-8, as being different enough to warrant further analysis. These were experience with or without NCLs for Q8.2 and Q8.4, and the key stage participants taught in for Q8.3, Q8.4 and Q8.6. The difference between responses were more significant for those with or without experience of NCLs as indicated by their smaller p values.

**8. Assessed with levels responses to questions relating to External Summative Tests of reading and maths.**

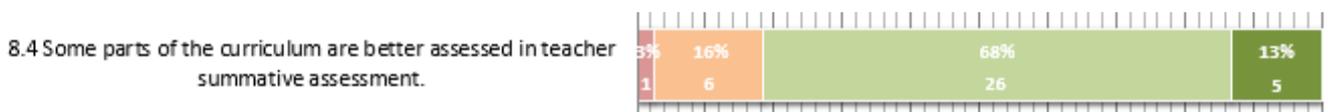


**8. NOT assessed with levels responses to questions relating to External Summative Tests of reading and maths.**



Responses to Q8.2 were shown as significant between those who had or had not assessed with NCLs (0.0906). Those who had assessed with NCLs responded with 53% (strongly) disagreeing, compared to those who had not, where 19% (strongly) disagreed that this was the case. The majority of participants without NCLs experience NA/ND (63%). It is possible that these teachers did not feel they had experience to compare external testing with and without levels.

**8. Assessed with levels responses to questions relating to External Summative Tests of reading and maths.**



**8. NOT assessed with levels responses to questions relating to External Summative Tests of reading and maths.**

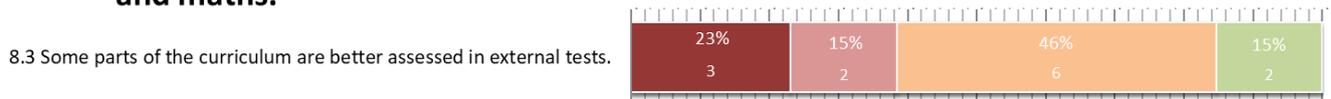


In response to Q8.4 all of those who had not assessed with levels (strongly) agreed that some parts of the curriculum were better assessed using teacher assessment, compared to 81% (strongly) agreed of participants with experience of levels. This data suggests experience with or without NCLs impacted on the strength of participants' response to Q8.4

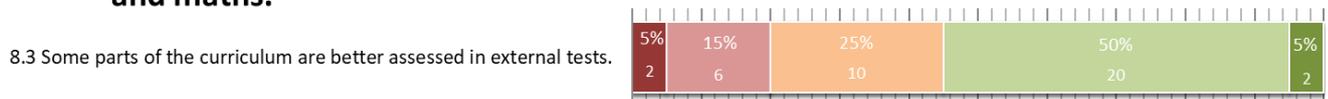
with those who had not assessed with NCLs feeling more strongly that teacher SA was best for parts of the curriculum ( $p=0.00530$ ).

Responses to Q8.3 ( $p=0.01364$ ) and Q8.4 ( $p=0.03145$ ) indicated that KS1 participants supported the use of teacher assessment above the use of tests for parts of the curriculum. Only 15% of KS1 teachers agreed with Q8.3 and 38% (strongly) disagreeing compared to

### 8. KS1 responses to questions relating to External Summative Tests of reading and maths.

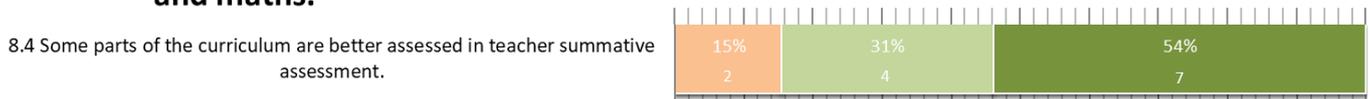


### 8. KS2 responses to questions relating to External Summative Tests of reading and maths.

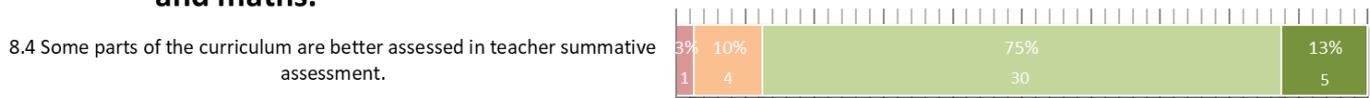


55% of KS2 participants (strongly) agreeing and 20% (strongly) disagreeing, that some areas of the curriculum were better suited to tests. Participant responses to this question indicates that the key stage taught in impacted on the teachers' views of the suitability of external tests for some areas of the curriculum.

### 8. KS1 responses to questions relating to External Summative Tests of reading and maths.



### 8. KS2 responses to questions relating to External Summative Tests of reading and maths.



The majority of participants (strongly) agreed to Q8.4 that some parts of the curriculum were best assessed by teacher SA, however 54% of KS1 participants strongly agreed, compared to 13% of KS2 teachers. This indicates that the key stage of the participants effected the strength of their view that teacher SA was better for some areas of the curriculum.

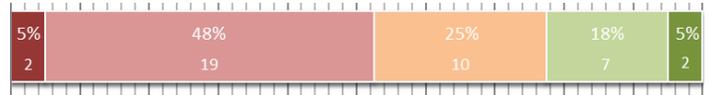
## 8. KS1 responses to questions relating to External Summative Tests of reading and maths.

8.6 Only teacher summative assessment should be used to assess CORE subjects.



## 8. KS2 responses to questions relating to External Summative Tests of reading and maths.

8.6 Only teacher summative assessment should be used to assess CORE subjects.



Responses to Q8.6 suggest that the key stage in which the participants taught affected their views on teacher summative assessment for core subjects ( $p=0.03958$ ). There was a higher proportion, 65%, of KS1 teachers who (strongly) agreed that only teacher summative assessment should be used to assess core subjects, compared to 23% of KS2 teachers (strongly) agreeing. Among the KS2 teachers, 53% (strongly) disagreed that only teacher summative assessment be used compared to 23% of KS1 teachers disagreeing, with no KS1 only participants strongly disagreeing.

The binary IV of the demographic groups compared provided no notable differences for responses to Q8.1 and Q8.5. Both these questions related to views on external testing of core subjects, that they are a good way of assessing core subjects, or that they should be the only way to assess core subjects. All participant responses to these questions were mixed and are not indicated to be impacted by the binary IV explored in this section.

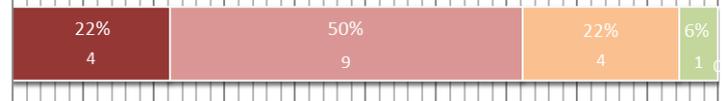
7.1.6.3 Further analysis of responses to Q9 according to binary IV

Table 7-9 Results of Mann-Whitney U test on participant responses from binary IV for Q9

Survey Question	Mann-Whitney U test result for binary variables based on participant demographic data					
	Experience with NCLs compared to no experience with NCLs		KS1 responses compared to KS2 responses		SATs year responses compared with non-SATs year responses.	
	p-value (p>0.05)	U score	p-value (p>0.05)	U score	p-value (p>0.05)	U score
9.1 Assessing WITHOUT levels takes focus away from NON-CORE subjects.	0.79530	291	0.22170	204	0.5182	249.5
9.2 Assessing WITHOUT levels has not improved assessment of NON-CORE subjects.	0.24570	246.5	0.05550	347	0.3615	237.5
9.3 Assessing WITHOUT levels has brought a balance to assessment of CORE and NON-CORE subjects.	0.43500	402.5	0.11200	189	0.04157	370
9.5 Tests are used to assess NON-CORE subjects in my school.	0.28710	251.5	0.66680	279.5	0.2483	331
9.6 Test should be used to assess NON-CORE subjects.	0.16260	234.5	0.75100	245.5	0.1912	338.5

## 9. SATs year responses to questions relating to summative assessment of non-core subjects, for example: history, geography, RE, PE, art...

9.3 Assessing WITHOUT levels has brought a balance to assessment of CORE and NON-CORE subjects.



## 9. Non SATs year responses to questions relating to summative assessment of non-core subjects, for example: history, geography, RE, PE, art...

9.3 Assessing WITHOUT levels has brought a balance to assessment of CORE and NON-CORE subjects.



The only demographic group which indicated a difference worth exploring for Q9 was those who did or did not teach in a SATs year. All statements apart from Q9.3 had responses which did not indicate an area of interesting difference ( $p=0.04157$ ). 72% teachers in SATs years (strongly) disagreed that AwLs has brought balance to assessment of core and non-core subjects. This is compared to 36% (strongly) disagreeing in non-SATs years. The majority of these participants NA/ND. Only 1 respondent in each group of teachers agreed. This data suggests that for survey participants, teaching in a SATs year or not had an effect upon views associated with the balance of core and non-core subjects' assessment. This could be due to the SATs only assessing core subjects, and the high accountability pressures associated with KS2 SATs.

## 7.1.7 Survey data discussion and conclusion

### 7.1.7.1 Question 7- Summary of survey responses to questions relating to teacher's summative assessment of Maths and Reading

To summarise responses to Q7, the majority of teachers felt they could confidently use AwLs for their SA, however, the data shows that the transition was difficult; an experience shared also with teachers who had not used NCLs, although at a lower proportion. Teachers who felt the most confident with AwLs were those who had not used NCLs. Most teachers also disagreed that the government had provided enough guidance, additionally (strongly) agreeing that detailed exemplars were needed to improve their confidence and accuracy for teacher SA. Previously, with NCLs, research indicated that teachers in SATs year already felt the need for more guidance and lacked trust in their own SA (Collins et al., 2010; Marlow et al., 2014). This ties in with teachers in SATs years representing most of the participants who found the ITAFs useful.

### 7.1.7.2 Question 8 Summary of survey responses to questions relating to External Summative Testing of reading and maths.

Overall, this data suggested teachers were in favour of a combination of test and teacher SA for end of Key Stage assessment. There was no definitive preference for only teacher SA, or, only test.

All teachers overwhelmingly agreed that teacher SA was preferable for some areas of the curriculum with less than half of responses feeling the same for external tests. This conclusion is echoed by Harlen and Gardner (2010) who argue teacher SA can capture areas of the curriculum tests cannot. However, this depends on rigorous internal and external moderation (Collins et al., 2010; Johnson, 2013). When comparing KS1 teachers' to KS2 teachers' responses, KS1 teachers felt more strongly about having the option to use their teacher assessment, with 100% of KS1 teachers (strongly) agreeing that teacher SA was better for some aspects of the curriculum. Since KS2 SATs are imbedded in managerial aspect of primary school accountability it is understandable that the KS2 participants may have internalised these discourses more completely into their teacher identity, thus accepting testing as a measure of learning more readily than KS1 teacher (Buchanan, 2015).

7.1.7.3 Question 9 Summary of survey responses to questions relating to summative assessment of non-core subjects, for example: history, geography, RE, PE, art...

Overall, the data indicates that AwLs was not seen to have improved the assessment of non-core subjects, and had not addressed an imbalance between the assessment of core and non-core subjects. The emphasis test based accountability leads to 'measurement-driven instruction' (Gipps, 2011: 27) putting emphasis on core subjects. Some teachers also showed this, to a lesser degree, regarding NCLs, suggesting it was partly an inherited issue where NCLs already impacted on the assessment of non-core subjects. Since a major justification of AwLs was to reduce curriculum narrowing (DfE, 2010a), this data shows for these participants a better balance has not been struck. It was clear that the majority of teachers did not use tests and did not agree tests should be used to assess non-core subjects.

7.1.7.4 What was the difference between responses of participants with experience of NCLs and those without?

The transition towards AwLs was the hardest for teachers with previous experience of NCLs. None of these teachers felt SATs were better without levels, and represented a large proportion of all teachers disagreeing that they were. This supports literature discussed in Chapter 5 relating to how teacher age and experience impacts on their identity, including what it means to be a successful teacher. Those who had assessed with NCLs had developed their identity as a teacher with NCLs measuring their performance as a teacher (Moore and Clarke, 2016). Teachers who had not assessed with NCLs expressed more confidence in AwLs, but regardless of experience with NCLs, the majority of teachers (strongly) agreed with the need for more guidance and exemplars for teacher SA for reading and maths. This indicated that even without experience of NCLs, the majority of participants required more guidance, suggesting that desire for this did not exclusively come from a longing to assess with NCLs, or for the previous assessment guidance. This group of participants additionally responded more favourably to questions relating to teacher SA for example, they represented the larger proportion of participants who agreed that only teacher SA be used for core subjects, moreover 100% of participants who had not assessed with NCLs agreed or strongly agreed that teacher SA was more suitable for some areas of the curriculum. This finding suggests that participants who had assessed with NCLs may have internalised

standardised testing, at least somewhat, into their professional identity (Buchanan 2015; Moore and Clarke 2016). These teachers have taught longer under the dominance of test-based accountability and may ‘...not be able to imagine a different way of being a teacher’. (Ball et al., 2011: 634). Whereas newer teachers, those who had not assessed with NCLs, are still forming their teacher identity, and able to hold onto educational goals that are not yet constrained by accountability driven school practices (Buchanan 2015; Wilkins et al., 2012).

7.1.7.5 What was the difference between responses of participants who taught in KS1 compared to KS2?

All teachers, regardless of their key stage, felt more exemplars in general would be beneficial to their assessment. The majority of KS1 participants found the ITAFs more helpful, this could be due to KS1 only having 2 years compared to KS2 with 4 years. Comparatively, end of key stage assessment guidance may be more relevant for non-SATs years in KS1 than KS2.

More support for external testing came from those in KS2 rather than KS1, with a large majority of KS1 teachers (strongly) agreeing that teacher SA was better for some areas of the curriculum. KS1 teachers can and do use teacher SA to inform their SATs judgement, whereas KS2 teachers cannot apart from writing. This is an important difference in how dominant discourses may influence teachers’ views on appropriateness of teacher SA (Moore and Clarke, 2016). KS2 culminates with SATs, and this pressure may be felt further down the school years. On a similar note, those in KS2 also agreed more that some parts of the curriculum are assessed best via external tests, showing less support for teacher SA only of core subjects.

7.1.7.6 What was the difference between responses of participants teaching in a year with SATs assessments and those who did not?

Regardless of teaching in a SATs year or not, the majority of teachers disagreed there had been sufficient guidance for AwLs. SATs year participants found the ITAFs more useful, understandably, since they are designed for use in those years.

Those who taught in a SATs year felt more that non-core assessment had not improved without levels, and that there was an imbalance between core and non-core subjects. This is likely to relate to the pressure statutory testing in these school years puts on core SATs

subjects and values ‘...certain modes of action other others’ (Biesta et al., 2015: 638), such as the prioritising core over non-core subjects.

## 8 INTERVIEW DATA FINDINGS- MAIN STUDY SETTINGS 1 AND 2

---

Analysis of interview data key themes are discussed discretely within each of the two interview settings starting with Setting 1 which included three teachers and two participants from the SLT, followed by Setting 2 which had two teacher participants and one from the SLT. SLT interview data themes are reported separately within each setting before being compared to teacher interview themes. Themes are then triangulated between settings.

The two theoretical lenses adopted to analyse the interview data are those referenced previously. The first is assessment theory of validity and reliability which incorporates learning theory as a means to conceptualise AwLs impact on learning. The second is the concept of agency as a theoretical lens to analyse and problematize teachers' enactment of AwLs within the context of their schools and National Education Policy. To conclude this chapter, triangulation compares survey data findings to interview themes.

### 8.1 SETTING 1- TEACHER INTERVIEW DATA THEMATIC ANALYSIS

The main emergent themes from setting 1 were:

- Transition difficulties and adjustments
- Different assessment uses and interpretations
- Relationship between teacher SA and test SA
- Evidencing – making teacher SA explicit
- Accountability pressures

Each of these themes will be reported below and supplemented with participant quotes.

#### 8.1.1 Transition difficulties and adjustments

The transition from assessing with levels to without levels was problematic for all S1 teachers. It was described as 'chaotic', 'manic' and a 'grey area' lacking clarity on expectations. Added to this was the notion that senior teachers in the school could not guide them.

*Nobody really [knew] what they were doing. There is nobody to consult because the more senior people don't know what they're doing either...  
(FS1Y6T2F)*

In order to adjust to AwLs there were two major factors, understanding the new curriculum expectations, and how this content would be assessed in the SATs.

Although the new curriculum was introduced in 2013 and made mandatory in 2014, in SATs years previous curriculum content was taught, and assessment continued with levels until the academic Yr 2015/2016. The first year of SATs without NCLS brought with it a sense of urgency indicated by Teacher F to familiarise themselves with the ITAF for writing guidance 'we need to first of all find out what it is, how to do it, and then teach them how to do it and then quickly get it into so many pieces of writing before the end of the year.' Teacher F demonstrated how much they relied on the ITAF to understand the expectation of writing at the end of KS2, with the late arrival of the ITAF resulting in pressure to ensure students had covered the specific content. Teacher E also felt guidance was provided 'last minute' preventing them from knowing how to best prepare their students, additionally echoed in Teacher D's comment that students needed to '...learn this quick, stick this in your work...' Teachers E and D focused on the change in focus as well as an elevation in difficulty compared to the previous curriculum saying

*...the standard and what they expect has gone higher...we're teaching stuff that wasn't taught until secondary schools (FS1Y6T1E)*

Teacher D, did not believe that increasing curriculum difficulty would impact student learning positively, arguing that 'children don't change because we're teaching them harder stuff.' These curriculum changes impacted on teachers' preparedness for SATs without NCLs, first knowing what was expected, and then being able to teach it adequately before the first SATs AwLs.

When describing how they felt now, assessing without levels, responses were all rooted in a better understanding of SATs test expectations, and having teacher SA guidance in the form of the ITAFs for over a year. They described feeling 'mentally prepared' (Teacher E), 'knowing what we're doing' (Teacher F), and having a 'much better understanding' (Teacher D) of expectations. Experiencing AwLs through KS1 and KS2 SATs provided a frame of reference for the new expectations, thus enabling the teachers to adapt. For example; Teacher D referred to using ITAF objectives to plan through the year, '...this year I felt it was because I knew about it [ITAF expectations] from the beginning, I could make sure that through the

year it was all [covered]' and Teacher E discussed how they had reflected on their reading teaching in response to the SATs paper '...we're going to have to do some specific enrichment of vocabulary next year on top of what we've already done this year because this seems to be a huge weighting'. Teacher E felt more reliant on the content of the SATs and ITAFs with AwLs prioritising what they taught more to meet assessment expectations than with NCLs.

*I'm sure it was like that before [with levels], but I think it's even more so now. You can't leave it until January, you've got to start trying to hit some of those [SATs objectives] from September which then narrows the curriculum right down for the year six...(FS1Y6T1E)*

Narrowing the curriculum based on SATs expectations seemed to be a symptom of a need for wider assessment guidance for schools, or a struggle to adapt to the scaled back teacher SA guidance compared with NCLs. All teachers described the pressure they felt when ITAFs were released in March 2016 before the tests in May 2016, describing a rushed and hurried period while trying to prepare students based on this AwLs guidance immediately. Teacher D express this as the ITAFs being '...thrown at us basically'. Even with the ITAFs, Teacher D felt they needed more guidance to teacher assess confidently, stating '...we've only got that [ITAFs] to work on. It isn't a full range of what they [pupils] need to cover.' All teachers felt scope of the ITAFs was limited, with the teachers turning to the internet, referring to YouTube and teacher forums where they found resources and support. Despite government policy increasing autonomy to schools for non-statutory assessment, these teachers when facing wanted guidance which would align their teacher SA with SATs expectations. In addition to online resources, teachers listed AwLs externally purchased resources.

- Ros Wilson<sup>2</sup>
- PiRA and Puma tests<sup>3</sup>
- Grammarsuarus<sup>4</sup>
- Chris Quigley<sup>5</sup>

---

<sup>2</sup> Ros Wilson developed a purchasable Criterion scale for assessing Primary writing without levels.

<sup>3</sup> PiRa and PUMA are purchasable summative test resources for Reading and Maths from the company Rising Stars.

<sup>4</sup> Grammarsuarus is an online purchasable resource provider originally providing resources to support the new grammar requirements of Curriculum 2014.

<sup>5</sup> Chris Quigley is a purchasable online curriculum and assessment tool. Schools can access curriculum objectives categorised in milestones.

In one example provided by Teacher E, summative tests were being purchased instead of non-test based teacher SA.

*As a school we are doing more, different summative assessments and we've had to obviously search around to find the best ones that we feel offer the best way of measuring summative assessment as a school and we're doing those in year six but we're also doing, probably doing, more summative assessments than what we were doing before. (FS1Y6T1E)*

This data displays how challenging S1 found replacing NCLs guidance for non-statutory assessment. This is not to say that NCLs guidance was ideal or idealised by the teachers who referred to them, only that the transition between NCLs and AwLs left a void for the teachers who needed to get AwLs expectations right in order to prepare their students for the SATs without levels.

An attempt to fill the void was seen in an overreliance on the ITAFs and exemplification material, for end of KS1 and KS2. One of the aims of AwLs was to combat teaching to the test, and the wash-back effect of assessment on teaching. From these findings the opposite is indicated, ironically resulting in a reliance on the ITAFs which relies on a narrower framework than NCLs. Teacher D said 'We live our life by the ITAFs, as they are called' suggesting the ITAFs are a major influence on teaching as well as assessment practices. She referred to her observation of other schools' approaches as an external moderation.

*...if you just work to that [ITAFs] then you could miss out a whole chunk of the curriculum and I think teachers are working to that because that's what you're getting moderated against...*

The ITAFs were discussed as checklists and tick lists to assess learning against and inform teaching, demonstrating significant 'wash-back' from assessment to curriculum. Within S1 this was exemplified by Teacher E who described their use of ITAFs like a core curriculum for children working towards 'expected'.

*...some of the children that we knew were going to struggle that's [ITAFs] purely what they focused on whereas I had the higher ones I could branch out and do other things as well. (FS1Y6T1E)*

External moderation of KS1 SATs relied on ITAFs the first year under AwLs according to Teacher D's experience as a moderator. This was suggested by Teacher D as a reason for teacher reliance on ITAFs and exemplars provided.

*Yeah because when you're moderated, and I am, I do go out and moderate other schools, that's [ITAFs are] what you're looking for evidence for (FS1Y2T1D)*

The teachers' perspectives here provide a view of the assessment changes as stressful, not in terms of having to assess differently but in the absence of having timely and comprehensive guidance to know how to assess in line with the new Curriculum. Specifically, these teachers seemed concerned with the weighting and framing of the new curriculum in the SATs assessments relying on the ITAFs and their experience taking their students through the first year of SATs under the new curriculum. This aided the teachers in adapting their practice to better meet end of key stage expectations with AwLs.

#### 8.1.1.1 New policy, same practices

When discussing the start of the transition towards AwLs, Teachers D and E referred to assessment structures and practices which reflected that of NCLs. Teacher D suggested that some schools were constructing assessment practices that resemble that of levels,

*...schools have now made all these, like, pluses and secure pluses because schools need more levels in between to measure progress... (FS1Y2T1D).*

It is interesting that Teacher D reflected on what they had learnt of other school's approaches to AwLs; specifically, that schools have adopted assessment systems that replicate the split level approach reminiscent of sub-levels from NCLs, criticised in 'The Case for Change' (DfE 2010). This critique was not related to their own school the use of colour coding and tiered judgments using the tracker adopted called SPTO<sup>6</sup> and references in other comments that similar approaches had been adopted in their setting with AwLs. As Teacher F commented,

---

<sup>6</sup> Student Pupil Tracker Online is an externally purchased tracking software for assessment data of pupils. It is no longer supported from the end of 2019.

*...[NCLs and AwLs are] kind of the same really because you've basically got a sheet of objectives to assess and you're basically still highlighting what they can and can't do...You're just using a different number at the end of it.*

This could relate to Agyris and Schön's (1974) notion of espoused theory and theory in use, that what is verbalised as practice does not necessarily match the practice observed. Or simply that practices under NCLs had continued. To track non-statutory assessment throughout the school, in a similar fashion as under NCLs, assessment criteria needed to be more comprehensive than the limited AwLs criteria of working towards expected, at expected and greater depth within expected. Therefore, additional criteria were introduced within these judgments to track progress. NCLs were criticised for encouraging the use of split levels assessment (sub-levels) like in APP (assessing pupil progress) which was replicated here. Although the school had autonomy to choose their non-statutory assessment policy and systems without levels, from these teachers' comments in-school assessment was driven, still, by tracking and defining progress.

There was also an intuitive understanding of NCLs as teachers D and E talked decidedly about knowing what a level or sublevel 'looked' like, indicating they could make a judgment without direct reference to the level descriptor. Although there was understandably comfort in the familiarity of NCLs, it was not idolised as a perfect assessment approach, rather there was security in being clear about expectations.

*...you knew what was expected of them [students] to get from that point to that point and what you needed to be able to cover... Still teaching to the test, but there was probably more guidance, and more structure there and as sort of a pathway through from one level to the next. (FS1Y6T1E)*

Having experience of the levels assessment APP rubric may have meant Teacher E felt more confident in their assessments as they knew what needed to be taught in line with SATs expectations. NCLs were also linked directly with TTTT by Teacher E, something they felt had continued with AwLs. NCLs were additionally criticised for limiting and confining student learning, whereas AwLs was praised for encouraging broadening and deepening of knowledge.

*...there's no limit on it [learning] so you know where children are... they might not be able to use capital letters, but they're still trying to do shifts in formality...So that we're not putting a ceiling on what we're asking them to do (FS1Y6T1E)*

This suggests there are ideals within AwLs which resonated with the teachers of high expectations and not limiting learning according to assessment criteria. It also draws attention to teachers feeling NCLs prevented them from having high expectations of their students previously.

#### 8.1.1.2 Still uncertainty of change

A continuing difficulty transitioning towards AwLs was the feeling of instability and uncertainty of future change. AwLs goal posts had changed within the last year, making it difficult to understanding the expectations. Subsequent changes were still being made without adequate time to prepare such as 'moving the goal posts' (Teacher E) for KS2 greater depth. Though Teacher F agreed with this they also talked about expectations that uncertainty is part of working in education. Sceptical of proposed changes to removing mandatory SATs tests in KS1, Teacher D seemed concerned such future transitions would be challenging.

*...they're [government] talking about changing it again and getting rid of the SATs again...I think it just sounds like more change...I think if they're going to change it they just need to make sure it's done properly and everybody knows what's going on and why and in advance because it's all, everything seems to have been done last minute and it's actually the children at the end of the day that really are getting this haphazard education. (FS1Y2T1D)*

For these teachers, AwLs had not been one change followed by stability, but had been two years of uncertainty, uncertainty in how to assess without levels but also uncertainty of if or when future changes would occur.

#### 8.1.2 Different assessment uses and interpretations

The interview data from S1 contained apparent and different conceptualisations of assessment terminologies. Notably there was a blurring between formative and summative

assessment, and a focus on FA as a tool for SATs preparation. Teacher E talked of feeling more attuned to SATs expectations meaning they could focus on SATs content more consciously than with NCLs.

The purpose of FA was often related to SATs preparation. FA judgments needed to be based on independent activities to better assess learning which required no teacher intervention or interaction. For example, Teacher D referred to practice test when discussing their day-to-day FA practices.

*...I think with SATs you do tests and things like that throughout, because it's got to be the independence, that's the key thing at the moment (FS1Y2T1D)*

For the Yr6 teachers, FA was a key part of SATs preparation. They both discussed using FA in terms of driving learning forward but based on end of KS2 expectations. Teacher F felt FA practices had improved by including students more in their 'learning journey', however, this was linked with students knowing what 'level' they were at, indicating a SA judgment. Therefore, this FA practice seems viewed through a SA lens rather than a broader curriculum lens. This risks prioritising specific curriculum content, sampled by the SATs tests and the ITAFs, above other content.

### 8.1.3 Questions of validity

Although not using the word 'validity', accuracy of assessments in measuring learning was questioned. For example, all S1 teachers felt there were unintended consequences of SATs such as, test timing, question styles, and student comfort in a test situation, all of which affected the accuracy of assessing student performance. Teacher assessment too was identified as having drawbacks such as bias.

Teachers felt that passing a test was not solely reliant on learning test content but also on understanding how to succeed when assessed *through* a test.

*But again, it's [I] suppose- can they pass the test? (FS1Y6T2F).*

*...they've got to get that pace and the tricks, for want of a better word, to try and get through a test...Some of it is to do with the environment (FS1Y6T1E)*

This implies that there are elements of the SATs test which the teachers felt impeded some students' performance. Teachers reported aspects of the SATs testing environment and regulations, like time limits and wording interfered with the students' ability to complete tests. Written aspects of the reading test and reading aspects of the maths test were felt to be a barrier for some students. Examples provided were;

*I think [with] maths sometimes the questions in the test, it might be that they don't understand, not the maths behind it, but they might not understand the situation or the wording (FS1Y2T1D)*

*...a reading test relies on children's writing skills and the two are unrelated. (FS1Y6T2F)*

*...the questions...I always feel like they're trying to catch the children out...getting yourself through a test...it's completely alien...if it's a topic that comes up that they can't understand, it will switch some of ours off (FS1Y6T1E)*

Being tested itself was seen to affect some children more than others, with some finding it stressful and upsetting, factors which then inhibited their test performance. Teachers described students having 'an absolute wobble' (FS1Y6T1E) and 'bombing in the test' (FS1Y6T2F) despite working confidently in the classroom environment. The Yr6 teachers particularly demonstrated a need for children to be prepared for testing especially those who found the situation stressful. Students were exposed to sample SATs and SATs like tests through the year which was described as an 'obvious' preparation practice. Although this was for in-school SA they also served a purpose for 'children [to] get used to it [SATs]' (FS1Y6T2F).

In Yr2, Teacher D had a different view; they felt their students were not 'fazed' by the SATs, even so they were still exposed to sample SATs through Yr2 'so it's not a big shock'. Teacher D also felt KS1 SATs tests were 'good preparation for them [students] later on' indicating this experience would help students when tested in the future.

The sampling of curriculum content was an issue for teachers in the maths SATs, where the test was weighted heavily towards to number, therefore did not represent wider maths curriculum learning.

*I think we teach them a wider curriculum than what's on the test...I felt that it was very number heavy again this year, decimals, fractions, percentages ...you get quite a few children who are quite good at the shape space and spatial awareness...there might be one question on it. (FS1Y6T1E)*

This issue that this uneven weighting and the impact that has on teaching practices is explored in [Chapter 2.6](#).

#### 8.1.3.1 Different assessment methods for different subject areas

A combination of teacher SA and SATs tests was desired, depending on the subject being assessed. This was possible for Teacher D who taught in Yr2, but not for teachers E and F who taught in Yr6, where only the SATs judgments are recognised nationally. Teacher assessment for Teacher D allowed for insight into a student's learning more than the tests, expressing that 'the children, really, [need] to be able to explore and explain rather than just put your answer [and] move on' (FS1Y2T1D). Teacher D felt teacher SA was required in addition to the test judgment especially for categories below and above the expected standard but the SATs test provided reassurance.

*I think the greater depth stuff needs to be done through the teacher assessment and discussion but the majority of it can be done in a test and backs up the expected standard...I think it's good that it's not just on the test, I think that it's good that you have the test to backup... (FS1Y2T1D)*

In Yr6, Teacher E preferred teacher SA for reading because of its capacity to let children '...articulate their own answers and talk around things and books in general.' (FS1Y6T1E). Teacher F felt tests were appropriate for some areas of maths but not others, with the example of arithmetic;

*I suppose with maths arithmetic it's fine to measure it with a test because the arithmetic paper doesn't have any words on it so it is just a case of can you*

*multiply these numbers together and you can see where they make mistakes or not. (FS1Y6T2F)*

When discussing teacher SA and the SATs test, teacher SA was felt to be more accurate than the test. Teacher F and E expressed with certainty, that their teacher SA was a better representation of student learning.

*Interviewer: Thinking about the assessment data that you got their share which do you feel represents the assessment of your class better, teacher assessment or test?*

*(FS1Y6T2F): teacher assessment.*

*(FS1Y6T1E): I think teacher assessment, definitely [represent children's learning better]*

Tests were felt to be 'such a tiny window' (FS1Y6T1E) and their accuracy was affected by factors, referred to in the previous section, which did not impact on teacher SA. For example; teacher SA had more opportunity to assess the subject especially in areas which were not represented highly in SATs tests, such as shape and space in maths. As Teacher E argues;

*...you are very limited as to what you can do on a test paper whereas you can do really full on investigations in the classroom in order to assess their learning...that you can't actually do on a test. (FS1Y6T1E)*

Teacher SA was felt to assess children against the objectives they were learning, more so than the SATs, which was not linked to learning so much as test performance.

*...the SATs test is teaching the children how to pass a test and the teacher assessment is teaching the children to achieve the objectives but in a way that suits them...(FS1Y6T2F)*

The teachers' efficacy in their own SA clashed with a discourse which was apparent in interviews, that for teacher assessment to be accurate it should align with SATs testing. This is reported [below](#) in 'Trust in teacher assessment' but links to this theme presented here.

### 8.1.3.2 Secure fit assessment of writing using interim assessment frameworks

A common dislike of AwLs was the secure fit assessment requirement using the ITAFs. It was felt that assessments based on this judgment did not reflect student ability. To achieve an expected judgment in Yr2 and 6, all ITAF framework objectives needed to be met. Teacher D found this all-or-nothing assessment approach particularly difficult in writing, where students could miss out on the expected judgment because of handwriting;

*It's very different now because...it's not best fit, it's secure fit...We had a child last year who had great flare, spark in his writing, he got punctuation, everything was right but he couldn't join his writing... [he] wasn't allowed to have greater depth because of that reason. (FS1Y2T1D)*

ITAFs for writing were not felt to represent creativity, an aspect of writing in the curriculum. Teacher D above refers to a child who had 'great flair, spark in his writing' but could not achieve greater depth due to an inability to do joined-up handwriting. Similarly, Teacher E felt the grammatical focus of the ITAFs directed attention away from creative writing.

*...we've got the interim assessment framework which we are using but I find that that's too heavily weighted towards punctuation and doesn't give the children the opportunities to be as creative as what we want them to be. So they end up going to secondary school without that creativity. (FS1Y6T1E)*

Teachers D and E specifically, disagreed with the secure fit assessment not prioritising creative aspects of writing over handwriting and spelling and grammar.

### 8.1.4 School expectation that teacher SA and test match

There was a strong indication from the all S1 teachers that their teacher SA and SATs tests were expected to match. This links with data from S1 SLT interviews that teacher SA and SATs results should align, which will be discussed in 'Test and Independence Valued' in section 8.2.4. An example, from the teachers' perspectives comes from a comment from Teacher F who traced this pressure back to the school governors.

*...I don't think governors understand that a teacher assessment and a test shouldn't always be the same thing...I'm not sure that's always presented to governors in that way...I'm saying in the classroom this person can do these*

*things. I'm not saying that if you put them in a test they will pass it. It's not the same thing but governors don't necessarily understand that and it's not explained to them either. (FS1Y6T2F)*

There is a clear separation for Teacher F here between student ability and test performance. They also allude to how assessment is 'presented' and 'explained' to the governors which places some responsibility with SLT for governor perspective and highlights how important the mediation of policy is through SLT to school governors. How Teacher F was being held accountable for their practice was strongly influenced by how the governors understood assessment from SLT views. Pressure from governors may be why Teacher D felt 'lucky' their teacher SA matched KS1 SATs tests results, and why Teacher E felt they should have had more accurate teacher SA.

*It [teacher SA and SATs judgement] was quite similar actually, luckily. So, the children that I thought would do well did, the children that I thought didn't. (FS1Y2T1D)*

*I think teacher assessment should have been more spot-on but it wasn't. So, those children didn't perform very well. (FS1Y6T1E)*

Sometimes SATs performance was correlated with teaching effort and input. Teacher E both acknowledged different starting points with different cohorts, and a responsibility to work harder to increase SATs results.

*We got quite a few children to greater depth last year and we knew we couldn't get the same amount of children this year...we did manage to get a few but only by the skin of their teeth and it took a lot of work with them to get them to greater depth...(FS1Y6T1E)*

There is a sense from this comment that Teacher E felt the more teaching input, the better greater depth results would be- suggesting a belief that assessment output should directly relate to teaching input.

#### 8.1.5 Trust in teacher assessment, the need for proof.

There was an emphasis on reliability, or lack of, when discussing teacher SA either from the teachers themselves or implied by evidence required to support it. Teacher F felt a need to

prove their SA and disagreed that this was necessary and that 'teachers should just be trusted to know their children' (FS1Y6T2F). For example, Teacher D described a situation of little trust in their judgments and the need to prove them; the SATs test results provided this validation, and 'proving' their teacher SA was accurate.

*...I suppose, teachers are now not allowed to just know that they [students] can do this and this. You've got to prove they can...I don't know why everyone is fighting against it [SATs] because I think people feel happy that they've done them and backed up [teacher SA]. (FS1Y2T1D).*

However, Teacher E showed distrust in their own assessment, blaming their own teacher SA for not indicating a child would struggle under exam conditions.

*I only had one who didn't pass in my group and she must have just, looking at the scores, gone completely to pieces but she gave me no sign throughout the practice test...she'd grasped everything, she could explain things, she had good reasoning skills and she just completely bombed on the test...(FS1Y6T1E)*

In this situation, Teacher E seemed to feel the test result was the 'correct' assessment, rather than their teacher SA. Ideally, Teacher E felt that lower-stakes tests had potential to support their teacher SA.

*...even some of the tests that we've done ourselves, I think possibly without the pressure of the day, we know that they can answer those kinds of questions (FS1Y6T1E)*

However, pragmatically this is not the case, and Teacher E shows conflict between their discussion-based pedagogy, providing opportunities to discuss and articulate student reasoning, with the need to prepare students for test conditions; potentially feeling that they have disadvantaged students in doing this;

*I suppose in one way we're not doing them any favours for the test but you want them to be able to articulate what they're doing...and sort of reason that out...May be we're doing them a disservice by trying to get them to just do it in their head, I don't know. (FS1Y6T1E)*

There was conflict for Teacher E, between teaching how they felt was right, and adjusting pedagogy to better prepare students for test conditions.

The need to be evidencing learning and teacher SA judgments, was argued to have increased since moving to AwLs.

*...the question is always where is the evidence, show me the evidence and I always think, come into my classroom and watch the lesson and there is the evidence. (FS1Y6T2F).*

An increase in 'book trawls' or 'scrutinies'<sup>7</sup>, and marking requirements felt excessive, expressed by Teacher F as needing to '...highlight everything under the sun to prove it [teacher SA].' (FS1Y6T2F). There was added pressure to ensure even practical lessons are evidenced with AwLs increasing teacher workload as a result.

*I work every night, there's not really a night I don't work...you've got to evidence everything, that's more written things so that's more marking. (FS1Y2T1D).*

Although Ofsted made referenced that 'Inspectors will scrutinise pupils' work...looking at examples of pupils' work to seek out evidence for progression in knowledge, understanding and skills.' (Ofsted, 2015: 25), they additionally set out that they do not require teachers to undergo additional work in preparation for inspection for example 'Ofsted **does not** expect to see photographic evidence of pupil's work'. Yet, the implication from the teachers' comments above was that evidencing pupil work and subsequent feedback in anticipation of future Ofsted inspections had become integrated into their practice. It was not viewed as extra but as essential to practice in the event of an Ofsted inspection. The need to evidence and prove TA without levels went hand in hand with the perception of not being trusted. It suggests there is an issue with the surveillance of assessment, and the assessment itself, which raises the question; has learning happened if it is not seen and evidenced?

---

<sup>7</sup> The monitoring and scrutinising of students books to evaluate what has been taught and how, sometimes referred to as 'book scrutinies'.

### 8.1.6 Pupil mental health

As referred to in the previous [section 8.1.3](#) the SATs assessments, specifically the tests were felt to cause students stress and anxiety. This was demonstrated in the Yr6 teachers' comments when discussing how tests themselves impacted on student performance. Although Teacher D felt this happened at a lesser extent in Yr2, the judgment criteria itself was seen to effect children's mental health, students were good enough or not with no in-between.

*...at the end of the day it's these children that are suffering. Their mental health, because it's that pass or fail. Ultimately the children know 'I've passed or I've failed', and to fail at six and seven is massive and I think...you don't put that on the children but as they get older they know don't they... (FS1Y2T1D)*

Teachers who taught in Yr6 felt specific pressures put on their students, resulting in some children 'falling to pieces'. This was compounded with the added pressure on students of only being able to pass or fail now with AwLs. This was described as 'soul destroying' (FS1Y6T1E) for students who did not get an 'expected' judgment. This meant that receiving SATs results was an emotive topic with pressure felt by both staff and students.

*...I think we're under pressure, for health and well-being for the children...You try not to put that pressure on the children but you ultimately feel that pressure from above and I think it sometimes, without you knowing, it can be passed on to the children...I cried at least three times giving results back so I just think it's not very fair on them. (FS1Y6T1E)*

Whereas, Teacher D who taught in Yr2, did not feel SATs had much of an effect on their students and rather that '...it's what schools make of them, really.' (FS1Y2T1D). This could be because SATs in Yr2 form only part of the judgment, and are combined with teacher SA, therefore there is less pressure put on the test result.

### 8.1.7 High-stakes influences of SATs

The use of Yr6 SATs data, and to some extent Yr2 SATs data, meant these assessments came with high-stakes (see Chapter 2: [High-Stakes Summative Assessment](#)). Although not teaching in Yr6 themselves, Teacher D was aware of increased pressure on teachers in Yr6.

*I think it [KS2 SATs] does make schools accountable which isn't always a great thing but it is for some schools...I know it's [SATs] used to judge schools and that's why the pressure's on because...Yr6 SATs results tell you whether you're a good school or not or not, but actually it's not about that and you end up teaching to the test because you've got to get that result because you've got to. You don't want OFSTED on your back. (FS1Y2T1D).*

Teacher D described a punitive environment, with SATs used as a measure to monitor school quality by Ofsted. TTTT practices Teacher D considered as a strategy to avoid consequences resulting from having 'OFSTED on your back'. However, accountability as a performance pressure seemed internalised for Teacher D, who suggested individual moderators could be responsible for all inter-school moderation judgments in the future;

*...we do cluster moderation now but people are just like yep, yep. Whereas I think...you would be more accountable for what you would say about other people's work and I think that was, that was what I had heard would happen. (FS1Y2T1D)*

This would make teachers, who do not know the students work, accountable for moderation judgments on other teachers' students.

The main purpose of SATs seemed to be understood for the regulation and monitoring of schools, rather than providing a summative judgement of a child's learning through school. Its value was lost on Teacher F, who said 'I don't think the SATs are beneficial to anybody personally.' (FS1Y6T2F). The focus on SATs for accountability purposes resulted in conflicting demands for Teacher E where preparing pupils for the SATs interfered with a broader education they wanted to provide.

*...as a school there's more things that we're trying to get covered as well as what we're trying to get covered for the government. (FS1Y6T1E)*

#### 8.1.7.1 Wash-back

The pressure and expectation of good SATs results encouraged wash-back from the SATs onto teaching. This has been well documented with NCLs in the literature (See [Chapter 2](#)). The concern presented in this data is that although wash-back is a legacy issue from NCLs, not

only did it persist in AwLs, but now relied on a less comprehensive assessment framework. This was experienced differently in Yr6, compared to Yr2, where SATs test results were used to inform teacher SA, Teacher D stated they could 'get around that bit' (FS1Y2T1D) because the KS1 SATs included teacher SA. In Yr6, with the exception of Writing, teacher SA did not contribute towards the SATs judgment. As previously reported the teachers felt pressure when preparing their students for the SATs, finding it a challenge to cover the curriculum in its new form with its shifted expectations. Adjusting to the expectations of the new curriculum involved experience of SATs without levels, all S1 teachers felt more prepared knowing how Maths and English were assessed, either through moderation using the ITAFs, or from the weighting of certain subject areas in the tests. This was evident for all of the teachers when discussing the influence of the ITAFs and reliance on any guidance provided by the Government to shape, not only what was taught, but how it was taught. Specifically, there were examples of coaching and reallocation as wash-back from the SATs.

An important part of SATs preparation was elements beyond curriculum content sampled by the SATs. Teachers were aware they were adopting these coaching practices with their students. Test specific attributes and skills were required for the SATs. Teachers recounted test preparation including coaching their students in test technique.

*...preparing children to pass a test is not the same as teaching objectives in my opinion. (FS1Y6T2F)*

*You have to be much more flexible and be able to spot relationships and patterns in order to be able to get through the tests. Therefore, that is driving how the teachers are teaching...you're coming down to test technique rather than actually teaching them. If that makes sense...(FS1Y6T1E)*

Part of coaching was ensuring students were familiar with SATs style questions from the beginning of the academic year through the use of practice tests. The students could experience the test environment they would be in, in advance, though Teacher E did not feel they had as many chances at this to be effective for all students. Encouraging independence in lessons was part of preparing students for SATs without levels. It well felt that pupils needed to be confident applying their knowledge independently in the SATs tests

therefore independence in daily lessons was increased. Even KS1 SATs questions were practiced throughout the year with independent activities being key in preparing students.

*I think with SATs you do tests and things like that throughout, because it's got to be the independence that's the key thing at the moment. Is it independent?  
(FS1Y2T1D)*

Reallocation was seen in the specific inclusion of ITAF criteria and subject areas with a high weighting on SATs tests, strategically from the start of the year. More concerning was the narrowing of the curriculum, in some cases, to only those areas.

As a moderator for Yr2 SATs in other schools, Teacher D described the temptation of replicating the tasks seen in exemplification material. They implied this leads to students being taught a restricted curriculum.

*...you've got the exemplifications which are really good, but having gone and moderated again, people are falling into the trap of, they just do whatever worksheet was done for the exemplification. They are just copying that...it's actually not the whole curriculum but you end up being focused on that because you know that's what they're [students] going to be checked against  
(FS1Y2T1D)*

Similar practices were discussed by the Yr6 teachers, of aligning what was taught to specific curriculum areas, which I think was felt to maximise the chances of a pass in the SATs.

*...in want of a better word, [we] narrowed the curriculum that we taught them because we knew that that was the basics basically that they needed to be able to cover in year six...So, I think the effect of it was to, to narrow what we were doing rather than enhance it so not necessarily positive. (FS1Y6T1E)*  
*...some children at the very, very low ability were drilled on a certain number of objectives which meant that they could, by the skin of their teeth, pass the test (FS1Y6T2F)*

This meant strategically focusing on areas of the ITAF and SATs content which had a higher representation, narrowing the curriculum. This was felt, however, to give students a better

chance of achieving an expected judgment than spending the same time trying to cover more of the maths curriculum. Teacher F defended their approach of not restricting the curriculum, which may have resulted in less students achieving expected in SATs. But this lack of strategic teaching on Teacher F's part may have opened them up to criticism. Their SATs results were lower than the 'very, very low' ability groups who received more strategic teaching. This is a clear example of assessment wash-back, with further examples given for English demonstrated below by teachers E and F. They discussed adapting their teaching according to the focus of the ITAFs.

*...other people were saying you've got to spell the year 5/6 spelling words correctly and then we were thinking, what if we don't use any of them then they can't make any mistakes. (FS1Y6T2F)*

Reallocation created conflict for Teacher E who felt they were compromising teaching a love of reading for preparing students for the reading SATs paper.

*...some of them want to just read David Walliams and you're trying to nudge them away from all those books...And I think we are having to balance that in year six a little bit, trying to get that love of reading as opposed to just answering test questions. (FS1Y6T1E)*

In 'nudging' children away from reading the 'easy' books they enjoy, in order to prepare their students for SATs, may have conflicted with Teacher E trying to 'maintain positive attitudes to reading' as stated in the curriculum. A variety of reading experiences is encouraged by the national curriculum, however, SATs preparation may be diverting effort and attention away from reading and towards test preparation and coaching.

Curriculum narrowing was extended to non-core subjects, where time was diverted away to focus on SATs assessed subjects.

*...you don't feel like you have as much time to assess them [non-core subjects] because you're so busy assessing the core subjects in so much detail. (FS1Y6T2F)*

*I would say personally in year six because of the narrowing of the curriculum I can't hand on heart say exactly what my children have done in other subjects because we haven't been able to cover as much of the breadth...(FS1Y6T1E)*

*It's all English and maths and around SATs time that's all you do...(FS1Y2T1D)*

With curriculum narrowing both outside SATs subjects as well as within, there was a clear pressure to allocate as much time as possible to areas of the curriculum thought to contribute towards better SATs judgments.

#### 8.1.8 Summary

What is evident from the S1 teachers' interview data is that TTTT and coaching are now based on a more stripped back criteria of the ITAFs perpetuated by their use as the sole focus of moderation experienced by these teachers. Far from removing the pressure to teach to the test, the removal of levels and the accompanying guidance and framework has continued this practice. Moreover, these criteria had additionally been embedded into lesson planning from the start of the academic year, ensuring it is explicitly covered, rather than organically taught as part of the national curriculum. Though teachers disagreed with targeting SATs and ITAF content, there was little notion of choice. SATs test results dominated over teacher SA, where teacher assessment was compared to SATs test outcomes. This was compounded by requirements to evidence and prove learning which was largely discussed in terms of ticking and checking against ITAFs.

## 8.2 SETTING 1- SENIOR LEADERSHIP TEAM INTERVIEW THEMATIC ANALYSIS

The themes presented in this section, although similar in places to teacher themes, are influenced by the leadership roles the HT and DHT occupied. Major themes addressed are shown below. How these themes overlap with teacher themes will be handled in the following triangulation section.

### 8.2.1 Transition and adaptation

Setting 1's adaption to AwLs started with the school introducing the new curriculum while still adopting NCLs practices, before trialling their own AwLs framework alongside it.

*...teachers were asked initially to keep, still add levels on the system. We then set up AwLs: emerging, developing, secure system running and try to run them parallel. To be honest with you that didn't work...to be honest nobody was getting any value out of it and it was a case, we either go all for no levels or we stick with levels until we have to stop, basically. (S1DHT)*

It is interesting that the word 'value' is used to describe whether a practice is worth exploring or not, this feels reflective of the commodification of education within the discourse of marketisation (Ball, 1998; Bailey, 2004; Baltodano, 2012). The language used by the SLT adopted to classify assessment was 'emerging, developing, secure' with the addition of a plus for each stage. AwLs was implemented top down from the SLT to teaching staff, with the policy built around the DHT and HT interpretation of AwLs. Staff meetings to introduce their framework were emphasised alongside moderation to promote constancy of assessments. The focus was on ensuring teachers knew how to assess according the SLT interpretation of AwLs.

Constructing their own AwLs practices was challenging, indicated by constant changes and refinement over the past three years. DHT remarked that '...we're still working on it [AwLs]...we have changed it constantly, we're changing again now.' Part of this challenge seems to have been a lack of guidance on 'benchmarks' and expectations for AwLs, what was expected for each year group. This was especially difficult for non-SATs years which did not have ITAFs.

*...it was quite hard and as DHT said there wasn't much guidance around [how] to decide on what your benchmarks might be. (S1HT)*

Inspiration was initially taken from the interim assessment frameworks (for Yr2 and 6). The structure of assessment at the end of KS2 influenced assessment in years without SATs, replicating a secure fit assessment to all years and not just Yr2 and Yr6.

*I think that now when we've come to the stage when we thought 'oh hold on if Yr6 are secure fit and Yr2 are secure fit then the rest of the school have to be secure fit as well' We can't accept best fit for the others. And I think that's probably part and parcel of this as well that we need to iron out. (S1HT)*

This demonstrates the wash-back of SATs and its influence to school years where these statutory assessments did not take place. In attempting to form benchmarks SLT encouraged subject leaders to work backwards from end of year expectations in all subjects.

*...it was a case of...asking them [subject leaders] to look at their subject and say, right, what do you see as the end of year expectations rather than the whole of the curriculum. And then built up basically something similar to the interim assessments but for every subject. (S1DHT)*

There was acknowledgment of a knee-jerk reaction, resulting in the development of 'a tick-list sheet' assessment which the DHT felt had been moved past that. The focus was now on mastery stating that '...AwLs gives us a bit more of, let's slow it down and let's bed it in' (S1DHT). From the comments of the teachers they still felt, particularly with the assessment of writing, they were assessing to a tick list. There was also not much sense of being able to slow down, especially in year six where teachers described feeling time pressured to cover everything that was needed before the SATs.

Resources to support AwLs were purchased from private companies, previously referred to in teacher interview themes, and influenced the way SA were made, for example, the HT and DHT discussed how their non-statutory AwLs strategies were modified to a points based system from the 'Ros Wilson' reading and writing assessment. This system 'literally gave you end of year expectations and points against it.' (S1DHT)

Summative half termly tests were also purchased, though these were criticised for not matching the challenge of the SATs without levels, so instead were used now to inform teaching.

*...And we brought in PIRA and PUMA assessments...they're not challenging enough...we're now moving further from that...it's actually unpicking it more and actually getting teachers to say how could you use this now to feed into your teaching because it's got to be assessment for learning (S1DHT)*

Teacher SA in this case had moved to being based on tests or score based, with testing expected to have a formative purpose.

### 8.2.2 Difficulties with tracking

A further challenge was how AwLs meshed with assessment data tracking and measuring progress. These were important aspects of in school assessment for the HT and DHT. They showed concern that not being able to track progress sufficiently impedes student progress and this had been challenging when transitioning to AwLs. Having some students taught with levels was felt to impede this further with the DHT stating that this mix of with and without levels prevented getting a 'true measure' of student progress.

*...until we have a full round of [students] where levels is wiped out completely and we've got a cohort of children who never ever sat levels in KS1...I don't think you're ever going to get that true measure. (S1DHT)*

Performance management was linked with pupil progress and the monitoring of assessment data, suggesting teacher knowledge of this data is useful.

*...when we did performance management, pupil progress meetings...children are mapped in terms of a matrix so you can see where they are as a cohorts and a year group, what we deliver to teachers in terms of data in people names, so they know where their children are, which ones are falling behind, how many of them, how far away from end of year expectations. (S1DHT)*

For assessment to work- by improving attainment- it required monitoring and targeted outcomes.

### 8.2.3 School's responsibility to meet raised curriculum standards

As a school, they felt committed in trying to get all students, regardless of starting of point or special educational needs, to the expected standard using AwLs,

*... we asked [local authority], we expect a 100% progress and that includes SEN children because they should be able to make that progress from their relative starting point. (S1HT)*

*I think as a school we do want all our children to leave at age related at least, like I said there will be some that won't but we want to push them as best we can to get there. (S1DHT)*

A message from the Local Authority (LA) was referred to as instilling this target of 100% in Yr6 reaching age related expectations, although DHT also acknowledged that they do not feel this would be possible for some children they would still 'push' to try and meet this target.

SATs were seen to fall short of assessing all areas students could make progress like life skills. There was a purpose of ensuring students were ready for secondary schools. Something else the SATs could not measure. The HT acknowledged that SATs 'only gives you part of the picture' showing they did not believe fully that SATs were a valid measure to judge school performance, as SATs cannot capture all that HT felt was important. There was also concern as SATs, now a scaled score across each cohort, would impact on comparisons made with previous and successive cohorts and that these would not accurately show progress the school was making.

*I think there's an argument that they [SATs] can be used to identify trends but then that's got to be offset against the individual cohorts and the nature of those cohorts and it can be quite dangerous to just measure one against the other and assume that what you're seeing is telling you how you need to move forward when you're actually dealing with a different cohort next time...it's not just about that academic progress anyway (S1HT)*

There is justification from HT that they care about the students beyond academic progress, and this may cause conflict when the school is held accountable based on KS2 SATs results.

#### 8.2.4 Test and independence valued

There was a discourse in S1 that, despite the HT's assertions the SATs did not capture all worthwhile learning, accurate teacher SA had to be based on completely independent student work, with accuracy measured by how well teacher SA matched a SATs results. Teacher SA judgments were assumed to be able to predict SATs results, as they were used as an indication of SATs results and thus were expected to match.

*Interviewer: So do you think that their teacher assessment should match the SATs test result?*

*DHTS1: it should be near enough.*

*HTS1: ...there will be some anomalies but you'd expect them to bear some resemblance to each other yes.*

Discrepancies between the two types of SA were said to be the result of inaccurate teacher SA where too much teacher support had been offered to students. Moreover, providing this support resulted in inadequately prepared students for the independence needed in test conditions.

*...we [have] still got work to do with all our teachers. A lot of it is going back to that sort of confidence in making sure that things are independent...(DHTS1)*

*...a prime example of that would be our year six teacher assessments this year which were, and books were indicating where, children were expected to be in the 90s for them...But in the test situation it was a very different picture. But the books, and the teachers were adamant that that was an accurate assessment. So as DHT said, that's around the level of support that they've had and then some of that is withdrawn... (HTS1)*

Not only is teacher SA seen to be at fault here, but also pedagogy which encourages collaboration and discussion between students and teachers. It may well have been the case that students were not prepared for the test, however the format of SATs and the effect they may have on the students was not questioned. This indicated a default position of a lack of trust in teacher SA when compared to SATs results.

Ensuring students could apply their learning under test conditions was important, part of which was ensuring that teacher SA was based on activities that replicated test conditions.

*... in the test it's completely withdrawn right from the beginning to the end whereas throughout the school obviously when they're teaching lessons is the support isn't completely withdrawn. It's getting that balance right and making sure the teachers are really clear about, how much evidence do we need to say that the children secure on it as well...(S1DHT)*

As teacher SA in this case is evidenced by the students' books, the format of the activities in lessons which inform this teacher SA seemed under scrutiny. There was disjointedness between what activities teacher SA is based on and the level of independence required by the SLT to inform teacher SA; it implies that teacher SA should be based on independent written activities aligned with the format and environment of SATs testing.

In-school moderation was thought to help in improving the accuracy of teacher SA.

*S1HT: we need to do more maths moderation*

*S1DHT: and I think that probably will help when teachers are bringing books and we've got the end of year expectations exactly the same suited for writing...*

There was a reliance on how learning was represented in the books. Independence was also sought for formative diagnostic purposes to effectively plan for topics. The DHT did not feel they had adequately embraced the use of small test, pop-quizzes and independent activities to inform future teaching.

*...what we haven't done is and I know the AwLs documentation implicated not just using tests but using other pop quizzes and doing all sorts of things to do your assessment...let's do it at least three weeks to four weeks beforehand [starting a new unit] and then you use that as your basis to say right can they do this independently and it will be more sort of quiz based, mind mapping's, writing things down, more creative way with all the exploring. (S1DHT)*

An emphasis on written and independent activities listed here takes the responsibility of knowing the children from the teacher, and limits group or class discussion in formative exercises. This emphasis on independence and preparing students for test environments could be indicative of the pressures schools are under to perform well in SATs. Both the DHT and HT describe possible negative outcomes of school monitoring either by OFSTED or by a decline in SATs results.

*...we're getting narrower and narrower with our curriculum because that's what we're being judged against, that all the effort that schools put in is for those measures and we refuse to do that here but it is happening that their curriculum is being narrowed down. It's a big concern with OFSTED that foundation subjects are suffering as a result of those measures. (S1HT)*

However, these critiques are not linked with their own practice, but seem to reference experiences of other schools, even though their emphasis on assessment appears to be constructed around KS2 SATs and an increased reliance on testing and independent activities with AwLs.

#### 8.2.5 Labelling or not labelling children

Sharing AwLs criteria was emphasised for reporting student progress to parents/guardians several times through the year.

*...we do tell them [parents/guardians] each half term roughly an average child, where they should be. It's helped I think, some parents who looked a bit and thought 'oh my child is slightly behind here' and now they'll come in and have a conversation, 'what is it that my child needs?'...they are coming in every half term and then they are like 'actually my child is still emerging plus when they shouldn't be' (S1DHT)*

When discussing the reporting of students' progress, there was a specific focus on why assessment labels were needed to ensure parents understood how their child was doing, feeling that discussion of progress alone did not give the full picture;

*...your parent can get a totally skewed view of how that child is actually doing. A teacher can sit and genuinely talk to a parent about what fantastic*

*progress they've made... but you haven't actually told that parent that they are operating below age related expectation...because teachers, I suppose, are reluctant sometimes to share that. (S1HT)*

This suggests a belief that providing a grade has an impact on parental engagement above reporting only child-specific progress. The DHT and HT suggested that teachers may even be reluctant to tell parents/guardians their children are working below age related expectations and instead 'fall into the trap' of focusing on the progress made instead. For SEND children this was linked to getting students as close to the expected standard as possible.

For this school, the switch to AwLs seems to have enabled a more joined up view of assessment and curriculum objectives when sharing assessment judgments with parents. However, it is unclear what held them back from this approach when using NCLs. AwLs labels themselves were not discussed with students though this was with levels.

*We do share with them [students] this is what the end of year expectation is and you're developing yourself within those expectations. Which I supposed to a point isn't that far off from levelling, however, I think we do now focus a little bit more on the individual statements...(S1DHT)*

Shifting from sharing specific assessment judgments with children, to focusing on the 'statements', could refer to curriculum objectives or interim assessment framework judgments. However, this does contrast with their approach not to share AwLs judgments more with parents, and also draws similarities with NCLs, breaking down criteria into smaller parts to be able to measure progress.

#### 8.2.6 Improvements without levels- student dialogue and parental engagement

The DHT and HT felt the policy reform gave them the opportunity to focus on curriculum objectives for progress, enabling more dialogue on learning between teachers and students.

*...children can probably articulate better now by saying, if you ask them, what is it that they're working on? They will say 'I'm working on my addition now I need to make sure I really work on that addition' That's probably...an advantage without having levels. (S1DHT)*

The DHT felt the secure fit assessment judgments made it clearer for teachers ‘...having the end of year expectation saying, right it’s a secure fit, this is what we expect.’(S1DHT).

The change provided opportunities for parental engagement and the HT and DHT felt they could better explain AwLs to parents.

*...we did lots and lots of parent workshops and so it is important that we try...to bring them along with us really. Because research shows if we have parents on board, that understand to engage and support their children in the learning then they make a lot better progress...(S1HT)*

It was a challenge to help parents understand the raised expectations in the curriculum, but it was felt that helping parents understand the curriculum and assessment changes provided opportunities for parental engagement, as well as helping parents grasp the implications of the policy reform.

### 8.2.7 Summary

AwLs education reform gave the school autonomy for their in school non-statutory assessment. They moved away from levels spending the past two years trialling and adapting assessment strategies, some designed internally and others purchased from education resource companies. The choices in S1 were confined to SLT and did not result in autonomy for teachers in how they assessed- the strategy was top down from SLT to teachers. Teacher involvement came from training and moderation sessions to ensure they adopted, uniformly and accurately, the approach chosen by SLT. Trust and faith in SATs as a measure of school and teacher performance was evident from the emphasis on SATs results above teacher SA with teacher SA being at fault when results differed from SATs. This was further reinforced by in-school assessment aligning more with the structure of the SATs, with even formative assessment changing to replicate independent written activities synonymous with SATs testing. However, this SATs driven agenda existed alongside the HT belief that SATs did not measure everything, and that progress was much more than academic attainment measured at the end of KS2.

### 8.2.8 Interview Themes Triangulation Setting 1

Themes from teacher and SLT interviews were analysed according to where they overlapped or conflicted. Additionally, some SLT themes provided context for teacher interview themes, indicating an influence from SLT response to AwLs on the teachers' experience of the changes.

#### 8.2.8.1 Overlapping themes

Teacher and SLT themes showed an uneasy transition towards AwLs, linked to a desire for additional assessment guidance to set the expectations of all year groups. The contents of the ITAFs were influential for both the SLT and teacher adaptation to AwLs. For the teachers, ITAFs were relied upon heavily to inform SA, contributing to a significant amount of curriculum wash-back. SLT data showed that ITAFs were used as a template for non-statutory assessment frameworks, even in non-SATs years. This reliance on ITAFs was linked to the pressure teachers felt for students to meet and have evidenced end of KS2 SATs expectations. Neither the teacher's nor the SLT themes showed a preparedness for the first year of SATs under AwLs. Pressure to achieve 100% age related expectations at the end of KS2 was felt by the SLT from the LA, influencing a commitment to 'push' students to get as close as possible. This pressure may have accounted for some of the TTTT strategies teachers referred to, to maximise the chances of students who were at risk of not meeting the expected standard.

The lens in which formative assessment was viewed was similar for both teacher and SLT themes. Formative assessment was viewed in terms of how it identified gaps in learning beneficial for SATs assessment, so that these areas of learning could be prioritised and thus better applied in test conditions. Criteria from the ITAFs were assessed against formatively by the teachers and used to structure planning for SATs preparation. Formative assessments in this case were being aligned with the format of SATs, holding a dual purpose of informing what needs to be taught through the lens of SATs, as well as reinforcing testing practices.

#### 8.2.8.2 Conflicting themes

Despite the DHT and HT saying they had moved on from an initial knee-jerk reaction to AwLs, where they produced a tick-list assessment replicating NCLs, teachers still described AwLs in practice as this way. This was particularly relevant in writing, where ITAF criteria

were used through the year for tracking and evidencing of learning against the criteria. In addition to this, 'plus' and 'minus' were added to AwLs criteria. The need was felt to track students using assessments which may have contributed towards a mirroring of NCLs to complement a legacy of sublevel tracking from levels.

An area with a substantial divide in opinion was teacher SA. The teachers felt strongly that their SA should be trusted, based on in-depth knowledge of their students afforded through teaching them day-to-day. Conversely, SLT felt teacher SA should be based solely on independent activities, and not doing so was viewed as evidence of low confidence from teachers when assessing. Teachers did not feel trusted by the SLT and disagreed with evidencing requirements for their SA to be trusted. Books were an important aspect of assessment for the HT and DHT who discussed teacher SA and evidence in 'the books' hand-in-hand. There was not one without the other, if it could not be measured it could not be managed (Green, 2011). Whereas, the teachers discussed evidence in 'the books' more as a performative task, one which neither benefited student learning nor added to their SA (Ball, 2012). Teachers were expected to 'act in particular ways that are valued within the school system and make these visible to others.' (Pratt, 2016: 896). This drive for explicitness was reflective of findings from Ward and Quennerstedt (2019), and indicative of practices influenced by a managerial discourse outlined by Green (2011).

Teachers challenged the idea that an elevated difficulty in the standard of curriculum content would result in higher student attainment. They felt they could not teach more difficult content, *and* teach everything to the depth which was required. For the HT and DHT, this was not raised as an issue, and instead was seen as an opportunity to expose all children to higher expectations resulting in higher attainment.

#### 8.2.8.3 SLT impact on teacher themes

Teachers in S1 were not included in the construction of the school AwLs policy and practices. In school AwLs policy was formed largely from HT and DHT interpretation of National Policy. The stressful transition towards AwLs experienced by the teachers, feeling that everyone was in the dark, included their SLT. There was no criticism of SLT for how the change was handled, teachers criticised the Government for the timing of guidance, or the

lack thereof. However, encountering differing assessment approaches over the span of a few years contributed towards uncertainty and instability.

End of KS2 SATs were a major consideration for AwLs in the SLT interview. Although SATs were understood to represent only a part of the progress students made, SATs performance influenced much of the assessment discussion. As previously described, discrepancies between SATs results and teacher SA resulted in questioning of teacher SA, but this was done without the consideration that SATs results themselves might not accurately represent learning. This demonstrates the impact of the neoliberal discourse in primary assessment pertaining to the high value placed on SATs results (Bailey, 2004; Pratt, 2016). There were examples of where teachers themselves doubted their own ability to assess especially where there were differences between test performance and teacher SA. Teacher SA was expected to be in-line with SATs test results. This expectation may be why teachers discussed TTTT strategies adopted to maximise test performance and aligned their teacher SA to SATs weighting and expectation. Likewise, the SLT interview suggested that formative diagnostic assessments should increasingly resemble independent written activities. The impact of this was evident from teacher interviews where teachers were conflicted between providing independent activities in lessons at the loss of interactive and collaborative activities, essential feeling constrained by performance driven priorities in the managerial model (Stevenson, 2010) S1 had adopted of AwLs.

### 8.3 SETTING 2- TEACHER INTERVIEWS DATA THEMATIC ANALYSIS

Themes which arose from the teacher and SLT interview data of Setting 2 (S2) are presented beginning with teacher interview themes. SLT themes will then be explored before comparing themes from both S2 interview sets.

Setting 2 teacher themes echoed those of S1 because both schools experienced the same policy change, these similarities were expected from their shared experience. What was different, and therefore warrants reporting S2 themes independently is how each setting mediated policy changes, the type of school, and pupil demographics were a consideration for presenting themes separately for each setting to acknowledge the differences in perspective and experience of the two Yr6 teachers from S2. Themes identified in S2 teacher interviews are:

- Adjusting to AwLs
- Questions of validity
- Use of formative assessment
- Evidencing to prove teacher SA
- Tracking
- Accountability

### 8.3.1 Adjusting to assessment without levels

The transition between assessing with to without levels was not discussed explicitly.

However, aspects of AwLs were indicated as areas of difficulty when experienced for the first time. For example, Teacher G felt 'lost' when assessing writing for the first year without levels, describing their experience as 'stressful', and Teacher H reflected back to feeling 'way less prepared'. Experiencing SATs without levels made teachers G and H more aware of expectations of AwLs. The maths SATs test informed Teacher H of a shift in focus, namely simpler calculations, but more understanding of mathematical principles, which was linked directly to what Teacher H now needed to teach.

*...the numbers are a lot easier but the structure is a lot more difficult. That's telling you that they [STA] are assessing children's understanding of the underlying mathematical principles...So therefore you must teach that depth of understanding...(FS2Y6T2H)*

The SATs test gave Teacher H clues of what they feel the Standards and Testing Authority (STA) wanted teachers to focus on. It is interesting that it is the SATs test providing this direction for Teacher H, and not the curriculum. Teacher H, as discussed later, does see SATs being used to hold schools to account as a 'game', so learning the expectations of the SATs under the new curriculum helped them to play the game better.

Another example came from experience of writing moderation. Both teachers felt more prepared for SATs writing moderation the second year in. External moderation of Yr6 writing occurred in this second year of the new SATs in S2. This extra year of experience enabled Teacher H and G to feel more prepared, partly because they had had the ITAFs for a year. Teacher H in particular felt they would have been unprepared if moderated for writing in the first year.

*...we'd have panicked because, well arguably, because the guidance didn't come out until halfway through the year...we were sitting there thinking, just praying that we weren't going to be moderated. (FS2Y6T2H)*

Some of the reassurance provided from this experience felt fragile for Teacher G who anticipated more change saying 'I feel like a lot of us are holding our breath for the next change.' (FS2Y6T1G). In 'holding our breath', Teacher G suggests another change is not desired, as they would not be able to rely on the AwLs knowledge they have built.

From teachers H and G, only Teacher H had experience of NCLs, and believed using NCLs for writing was preferable to AwLs of writing, arguing '...with levels, I'm not saying it was great, but compared to the current assessment arrangements which are a dog's dinner.' (FS2Y6T2H). It is important to note that Teacher H was not viewing NCLs as unproblematic but preferred compared to AwLs. For maths, when reflecting on NCLs, assessment was felt to drive what they taught, resulting in SATs test results which did not necessarily reflect the ability of their students.

*...so everything was kind of 'the tail wagging the dog' I felt, if I'm honest. The thing I objected to most really was the way we would do standardised testing based on, well you know, previous SATs papers or QCA materials...for example, in my last year in year two, that was the final year of the last testing regime [NCLs], all of the children in my maths group achieved level 3, every single one, by some margin. But were they level 3? No they weren't. (FS2Y6T2H)*

Teacher H preferred the clarity of AwLs, that students are either at the 'expected' standard or not; compared to NCLs where it was common practice to assess in gradients using sub-levels. It is difficult to discern whether their preference for the clarity of 'met' or 'not met' criteria came from. It could be influenced by a discourse within policy emphasising the high expectations of meeting the expected standard compared to the previous level approach. Or, that Teacher H felt secure in *knowing* how to assess without levels now they had experience of SA with the ITAFs. When discussing teaching, since the AwLs reform, Teacher H echoed Government Policy of a 'mastery' curriculum, criticising previous education policy

for differentiation, which allowed for a simplified curriculum to be taught to some students, limiting expectations and holding back attainment (Koretz, 2008).

*I do not differentiate the lessons, any lesson, I haven't differentiated a lesson in the traditional sense for about four years. The reason for that is...if you look at maths, it's a mastery curriculum, it says all pupils must, must achieve this. That's the aim. I'm not going to help those pupils by giving them work that's below the age related expectation ever. I'll just work slower, we'll work through it slower. (FS2Y6T2H)*

However, it is unclear how going slower through a concept is different to differentiating to build knowledge up from the child's starting point. This also does not address how Teacher H approached groups of students who do not work well at the slower pace, be that higher or lower achievers.

#### 8.3.1.1 Late and/or insufficient guidance so seeking support elsewhere

Late guidance for end of KS2 assessment contributed to Teacher H 'praying' S2 would not be moderated for writing in the first year of the new SATs. Both Teacher H and G, as previously reported, felt more prepared when moderated the second year of AwLs. This suggested the ITAFs came too late for the first SATs with AwLs. Equally, exemplars for writing received similar criticism, with Teacher G feeling this guidance had been released too late, and Teacher H wanting them to be clearer at exemplifying expectations at the end of KS2.

*...I think we'd have benefited from, even more so, a range of writing across, deeming it being at the standard or not the standard and working towards. I feel it would have been useful to have more exemplars and those to have been released a little bit earlier because they were released quite late as well. (FS2Y6T1G)*

Comparative judgment instead was viewed as an option for Teacher H, eliminating the need for comprehensive guidance and exemplars for teacher SA of writing.

Autonomy for schools to build their own AwLs practices allowed S2 to try external providers for assessment resources, with the freedom to change or abandon them. A number of externally purchased AwLs resources and tools were mentioned by the teachers. These

included tracking software and working with a local Maths Hub. However, this new flexibility caused some instability for Teacher G.

*I feel in a similar position as last year because our assessment system is changed again this year, just as we became comfortable with depth of learning tracker<sup>8</sup> it then changed to something different, which can be quite intimidating... (FS2Y6T1G)*

Experiences of the external curriculum and assessment resources and support was mixed for Teacher H. Working with the White Rose Maths Hub<sup>9</sup> was useful for adjusting to AwLs, but an experience with an advisor for writing conflicted with Teacher H's approach to embed the new curriculum.

*...there was pressure put on from an external adviser who came and said 'you're not getting much writing done,' and I was like 'I know it's because we need to teach this stuff first, or at least some of it before we can help to interweave into their writing...(FS2Y6T2H)*

Teacher G desired constancy of an assessment tool, whereas Teacher H preferred advice which supported their ethos as a teacher.

Writing moderation, for both teachers, '...seemed to be assessing children how well they can fill out the checklist rather than how fluent their writing' (FS2Y6T1G).

*...the lady came and sat down, looked at some books for about half an hour, then called us in one by one, opened up her spreadsheet and said show me a semi colon, show me a fronted adverbial, show me this show me that and I did. And it got to the point by the end where she just went 'whatever it's good, you've done it all.' (FS2Y6T2H)*

Neither teacher felt the ITAF based judgments reflected student writing overall. Basing the writing assessment judgment of evidencing of the ITAF objects was referred to as 'soul

---

<sup>8</sup> Depth of learning tracker is an assessment tool available from Chris Quigley Education to track progress and attainment based on their Essentials Curriculum which adapts National Curriculum Objectives into lesson friendly objectives for teachers and schools to follow.

<sup>9</sup> White Rose Maths Hub is part of the Trinity family of academies, schools and initiatives. It provided a range of maths AwLs resources and schemes of work, offering free to paid resources.

destroying...horrendous' by Teacher H who suggested fault with their moderation experience was not with individual moderators but with the training each received and how rigidly they stuck to it.

*It seems to me that this person...had stuck to what they've been trained to do.*

*Nothing, more nothing less...how can I criticise that person... (FS2Y6T2H)*

This experience of moderation also reflected Teacher H's expectation, 'they [moderator] did exactly what we were told was going to happen, in quite a cold robotic way but that's what I had expected' with the view moderators who diverted from this 'robotic' method '...didn't stick to what they were asked to do...' (FS2Y6T2H). Expecting moderation to be formed around the ITAFs provided some clarity in knowing what the moderator would focus on, even though this was not felt to result in better writing and added to their workload.

*...it was really useful because it made clear what the moderator will be looking for but in some ways I think all teachers...were just trying to make sure that all children had included all of the different elements and...in some ways teaching to a checklist...which meant that some children were producing work which ticked off every single element of the checklist that maybe [was not as good] as a child who'd only ticked off a few elements of it. (FS2Y6T1G)*

*I hate it but I can do that because it's very clear, it's very defined on what you want. Will I produce better writers no, will they have more fun no. But will the workload be great yes it will be. (FS2Y6T2H)*

Although Teacher H expected moderation to be based on the ITAF, they agreed with Teacher G that their moderation experience was an outlier after conversations with Yr6 teachers in other schools suggesting not all moderators stuck 'rigidly' to their training. This highlighted inconsistencies between moderators and added to uncertainty of moderation in the future. For Teacher H, another approach to moderation of writing was suggested, comparative judgment was an opportunity to mitigate some of the issues they felt were present in Yr6 writing moderation.

*...this is where I think the comparative judgement idea really works is that, do I need to show a semi-colon to prove that I have a strong control of the English language. It's too prescriptive...(FS2Y6T2H)*

### 8.3.2 Questions of Validity

Assessment was discussed by the teachers from S2 as both a subjective and objective judgment. For example, assessment was discussed as open to interpretation based on the context of different topics with variation in how different students responded to a SATs environment, but also discussed how SATs assessments can be used to show and measure improvements in learning. S2 teachers' views on this, how well SATs judgments accurately assessed the national curriculum it targeted, are essentially views of the validity of SATs assessment.

For some subjects and subject areas, SATs tests were not viewed as capable of assessing all aspects of the curriculum, for example, tests for maths and reading were not considered to be all-encompassing assessments. SATs tests were not felt to assess 'the way that children will make links between different stories and different genres that they've read or... being able to compare and contrast and perform and to read aloud' (FS2Y6T1G). In maths the SATs was not seen as being able to assess '...following lines of enquiry and investigated work...' (FS2Y6T2H). Teacher H and G did not believe SATs testing was a valid representation of attainment representative of the breadth of curriculum content. SATs tests were viewed as suitable for calculations in maths and retrieval of information in text for reading but fell short of being able to assess verbal reasoning and enquiry. Both teachers differed in what they thought *could* assess curriculum areas not represented well in SATs tests. Teacher G felt their teacher SA would allow for verbal reasoning and problem-solving to be assessed, whereas Teacher H felt that teacher SA based on frequent low-stakes testing was needed.

*I think a test is good for testing fluency and a sort of basic understanding of processes like with the four operations in maths for example but for more complex tasks and verbal reasoning, I think that's where teacher assessment comes in. (FS2Y6T1G)*

*Everything is best suited to teacher assessment as long as the teacher is using a test. Now that sounds like a contradiction in terms, doesn't it? But I'm talking about the tests that you do to inform your judgments. (FS2Y6T2H)*

Teacher H's comments suggested their faith in tests was dependent on its purpose. In recalling their experience teaching in Yr2, Teacher H did not feel the tests with levels reflected the ability of their students, 'all of the children in my maths group achieved level 3, every single one, by some margin. But were they level 3, no they weren't' (FS2Y6T2H). This exemplified score inflation indicative of test coaching, with Teacher H clearly aware that their coaching had invalidated the judgment of Level 3.

Useful testing for Teacher H required a formative purpose, to not solely measure learning but improve it as well. When discussing termly maths tests they set as maths lead, they emphasised the tests purpose as formative.

*...the content of the test, which is a termly test, features just the content of what's been taught and nothing else...I go away and do some data analysis. I report back to the teachers with, these are the five key weaknesses (FS2Y6T2H)*

The implication that teacher SA requires a test to be accurate will be discussed in more depth later.

For English assessment, the curriculum was seen as too complex to be assessed from set criteria. Reading assessment was felt to be very subjective and 'tricky' (FS2Y6T1G) to assess or quantify (FS2Y6T2H). For writing, ideally a comparative judgment model would be implemented according to Teacher H, who felt this would provide a better judgment.

*...comparative judgement would be the best because we could have thousands of examples of writing and look at it from that point of view and go, yes that's pretty much, that's there...That's why a comparative judgement model with something so subjective as writing would be great (FS2Y6T2H)*

Differences between their own moderation experience and that of other schools may have influenced this view, that judgments based on the ITAFs did not reflect what either teacher thought was 'enjoyable', 'fluent' or 'cohesive' writing. An area specifically disagreed with was

secure-fit assessment of writing using the ITAF, resulting in some judgments valuing handwriting and individual components of writing, rather than their effective use or overall fluency in writing.

*...it was such a shame where there were particular children who were deemed as being not the standard just because their handwriting wasn't particularly neat or because they didn't think to use a semicolon in certain places, it was a shame. (FS2Y6T1G)*

The issue seems to be the equal importance of all the criteria in the ITAF for writing; with all criteria needing to be met, handwriting and spelling became as important as how language was used at the expense of what the teachers felt represented good writing.

There were certain subject areas S2 teachers felt *could* be tested well, these included information retrieval from text, calculations, and basic number skills proficiency. However, even for these curriculum areas, test outcomes were dependent on how comfortable students were in formal test conditions. Teacher G believed test conditions caused stress for some students, and that test timing unfairly prioritised speed over competency.

*Some children get really stressed and really anxious through formal assessment and as a result they don't perform as well as they can...they are tested on how quickly they do things...instead of how confident they are with a set of methods (FS2Y6T1G)*

SATs results for these students were not deemed to represent attainment of the construct the test was targeting. This is another example of where teacher SA was preferable for Teacher G who saw teacher SA for these students as 'more fair...and not penalise them because they struggle under test conditions'. Comparatively, Teacher H felt that the mind-set around testing was what needed to change rather than finding alternatives to tests for SA.

*...issue with testing isn't the test, it's everything around it and the mind-set issues... (FS2Y6T2H)*

They tried to avoid 'the T word' with their students indicating an association between tests and stress for students. They acknowledged SATs test-based accountability for schools was potentially punitive, and therefore found it challenging to remove negative associations with SATs testing.

*...trying and avoid the T word even though they know that's what it is but, at least by saying that they know that...I don't care about you score...if kids knew that their teachers just needed to know what they could do and what they couldn't do...But it's a punitive system, it's about accountability and that's what's driving everything really. (FS2Y6T2H)*

Teacher H was trying to convey their mind-set around testing which was formative. A solution to stress and pressure impacting on testing was to help children change how they think of tests, at the same time Teacher H highlights the reason why KS2 SATs cause pressure for schools. Teacher H did not consider that high-stakes testing could be changed, or that teacher SA should be used to additionally inform SATs as Teacher G suggested.

### 8.3.3 Evidence- needing to prove assessment judgments

Evidencing teacher SA judgments was consuming in both time and curriculum focus. Preparing for moderation contributed towards this in ensuring there was adequate evidencing of the content of ITAF for writing. All objectives listed needed to be evidenced in student work, in preparation of moderation.

*...last year we had a checklist for every single child based on all the writing they've done in their...books and we tick off each element and which paragraph it was found in, just so when we came to be moderated we were able to say 'oh there's a modal verb in this piece of writing in this paragraph' so it was quite time-consuming...(FS2Y6T1G)*

The workload of this evidencing was 'quite time-consuming' for Teacher G, and described as 'horrendous' and 'horrible' by Teacher H. The evidencing itself was a means to an end, rather than something felt to positively benefit students. Despite this adding greatly to their workload, Teacher H found it produced little useful information. Though evidencing was an

essential component of supporting Teacher H's Maths and reading SA. The requirement to evidence learning was emphasised to enable them to reach their teacher SA judgment.

*...I know that if a child is meeting the expected standard, I need to gain evidence of a child in a given objective...So it's very simple, actually. If a child can do that, if I have, what I believe is sufficient evidence, I'm happy.  
(FS2Y6T2H)*

Without supporting evidence for all areas of a topic, Teacher H would not summatively assess a pupil as being at the expected standard, regardless even of SATs performance. For example, they felt unable to assess a child in maths at the expected level due to a lack of evidence outside of the SATs maths test.

*I had a boy in my class who was on the autistic spectrum, very capable mathematician...he was also in my English group. He would not write for me in English. He was not going to write for me in maths, no way at all. No way. Did he smash the test? Yes he did. In my teacher assessment, could I assess being at the expected standard? No I couldn't because there wasn't enough evidence of reasoning...the government are saying I need fluency, reasoning and problem solving and I didn't have evidence of this one thing. (FS2Y6T2H)*

Although aware that the lack of lesson-based evidence was linked to their student's autism, Teacher H did not feel they could assess the student at the expected standard despite feeling they were a 'very capable mathematician'.

The reliance on evidencing teacher SA extensively was indicative of performative practices, ensuring evidence of learning was clear whether the teacher felt it represented the learning it is intending to or not. It is also indicative of a lack of trust in teacher SA.

At the time of interview, it had recently been announced that the requirement for teacher SA to be submitted to the LA would become optional. This change was not welcomed by the teachers for example, Teacher G felt teacher SA 'allows you to check and test those things that aren't tested by the formal SATs tests' and is needed when '...a child who was expected to get a certain level...didn't...teacher assessment can explain may be why a child didn't reach that level...' Supplementary science tests were used for the end of KS2 teacher SA

judgment. Teacher G felt this enabled them to check student attainment now the science test was non-mandatory implying the presence of a discourse that teacher SA alone was inadequate for SA. Teacher H felt the removal on teacher SA from KS2 SATs was further evidence that teachers' judgments were not trusted, stating 'You [LA] don't care about what I think, and you don't trust what I think...' continuing to say '...and actually I'm not sure that I trust what I think sometimes.' As previously reported above, Teacher H argued that teacher SA is best but only when based on a test. The onus from Teacher H was on *needing* to 'prove' their assessments and tests provided the evidence. Conversely trust in their own judgment, independent of a test, was low. If SATs tests were taken away, they Teacher H said they would still use tests, emphasising that '...I need to know what they [students] know, and I need to know what you [students] can do independently' The independence testing provided was felt to be the only way progress could be measured, justified by the '...huge disparity between teacher assessments.' Disparity between teacher SA was linked to a lack of common understanding of what non-statutory assessment looked like without levels. Even Teacher H's lesson based formative assessment required planning and preparation such as questions from a website, rather than relying on their own professional experience. Teacher G differed and discussed how their FA was based on observation, questioning and marking.

*...observation...through what children are sharing in the classroom and the work they are producing, questioning to check understanding or to poke and prod and check how deep their understanding is and marking of classwork mainly. (FS2Y6T1G)*

#### 8.3.4 Tracking

Using assessments to track progress was challenging. Different progress tracking practices were needed than those previously used with levels as it was difficult to translate measurement in sub-levels to AwLs. There was a danger, according to Teacher H, that schools may attempt to replicate a 'levels' and in doing so end up with an assessment disconnected with the new curriculum. Schools that did this were felt to be responding out of 'A fear of accountability'. Teacher H had heard of a school who did not track progress, and yet were graded 'Outstanding' by OFSTED. This supported his viewpoint that collecting

assessment data was separate from tracking. They distinguished the two because collecting assessment data had a formative purpose for improving teaching in the future.

*...it's not purely for just to stick numbers in a spreadsheet and say 'look at our lovely numbers.' because that's, we don't believe that's going to have much impact for the children whereas the other stuff we do, we put a lot of effort into, arguably does have impact on the children's learning. (FS2Y6T2H)*

Though, there was a closer relation between assessment and tracking for Teacher G, who discussed the two interchangeably, at times.

### 8.3.5 Accountability

The teachers discussed links between SATs assessments, OFSTED and school accountability. They agreed that teachers and schools should be held to account but described the accountability system as 'punitive' and that the current SATs based accountability system puts too much pressure on schools. For example, the suitability of the current measure was questioned by Teacher H who felt it had become a game and did not serve an educational purpose because of this.

*As soon you put numbers on things and targets in that way, it becomes a game...assessment needs to be useful but the government have got its uses, OFSTED's got its uses. I've got my uses which are about the children. (FS2Y6T2H)*

They imply that assessment data for accountability purposes is separated from assessment data for 'the children'. Data was viewed as open to manipulation, to suit purposes beyond student learning.

*...I think what they've [assessment with and without levels] both got in common is they can both be manipulated and both be interpreted in different, to suit the aims of government and the aims of schools...ultimately I think we're just putting pressure, the systems putting ridiculous pressure on itself because the secondary schools don't care... (FS2Y6T2H)*

The SAT's main purpose was believed to enable comparisons of schools. Teacher H disagreed with Teacher G, who felt SATs had an additional role by providing assessment information to secondary schools.

8.3.5.1 High-stakes influences of SATs- Teaching children to educate them or to pass a test? This emphasis on importance of SATs results for school success was evident through the teachers' comments on the importance of, and preparation for, SATs throughout the year. Preparation included the use of practise SATs tests from the beginning of Yr6.

The use of government standardised testing to directly influence what was being taught was normalised for Teacher H, referring to this as the '...tail wagging the dog. It's the assessment basically putting pressure on the schools.' This could explain Teacher H's openness in discussing strategies that they use for TTTT, referring to SATs preparation as a 'game'.

*If I am to succeed in terms of the numbers as a year six teacher then basically, I've got to game it. I'm not talking about cheating, I'm talking about 'right kids this the criteria [ITAF], we've got to do this.'* FS2Y6T2H

The over reliance on the ITAF for writing, and using knowledge of SATs without the levels, as previously discussed, also indicates a 'gaming' of assessment. Efforts were focused on evidencing according to the guidance provided, and teaching reallocation depending on the weighting of curriculum content in the SATs tests.

### 8.3.6 Summary

Both teachers recognised the function of KS2 SATs as a necessary measure of school performance. Teacher H was more vocal both on their objections to the use of assessment data for this, and their ability to 'play the game' as it was referred to. Structuring writing through the year around the ITAF was a major indication of TTTT strategies, neither teacher found it produced writers at what they felt were the curriculum standard.

Teacher assessment was held second to SATs in discussions related to SA. Teacher H's responses were mixed, ranging from an emphasis on testing to inform teacher SA, and to not assigning an 'expected' teacher SA judgment due to a lack of evidence from an individual student, despite them gaining the appropriate grade in the SATs test. Teacher G felt teacher SA was essential for assessing curriculum areas that SATs testing could not but

referred to teacher SA more as a backup for assessment than being an assessment in its own right. The need to evidence teacher SA seemed driven by the use of the ITAFs to assess writing, with their moderation experience supporting this need. Both teachers felt this significantly added to their workload. Clarity of a secure fit assessment was reassuring to both teachers who both found the first year working towards the SATs without levels difficult especially with the ITAF arriving later in the year.

Without directly using the word validity, the teachers were critical of the SATs assessments being able to represent the expected learning from the curriculum. Areas of subjects such as verbal reasoning and in-depth discussion about reading were not felt to be represented well in SATs testing. Additionally, the quality of writing assessed under a secure fit ITAF assessment of writing were not felt to represent writing well. Sitting formal SATs tests was also felt to impact on some students ability to perform well, resulting in outcomes which did not represent student attainment at KS2.

## 8.4 SETTING 2- SLT INTERVIEW DATA THEMATIC ANALYSIS

As assessment coordinator through the shift to AwLs, the DHT in S2 attended the SLT interview. They were able to share their perspective on changes as they had helped shape and construct S2's approach to AwLs. The themes this data produced were:

- Transition and adaptation AwLs
- Test as an objective measure of learning
- Labelling or not labelling with AwLs
- Parental engagement

### 8.4.1 Transition and adaptation

Transitioning towards AwLs was said to be positive for the school. The conclusion was reached in hindsight, that NCLs had 'run its course', with the DHT supporting its removal. The criticism for NCLs felt indicative of the DfE rationalisation for the changes that levels were not fit for purpose anymore and contributed towards a label driven assessment.

*I'm not completely satisfied with this system as it stands...If I'm really honest actually, going back a few years, the levels business had kind of run its course with us anyway. We were using a particular tracker which really was just a-*

*type in the level, type in the sub level-it meant nothing and it wasn't against key areas of learning or specific objectives.*

Though agreeing with the change, the transition still felt overwhelming initially '...further down the road and we've made the transition, I feel as if it's been relatively positive...'

In adapting to AwLs the school utilised different assessment resources from private companies. These external resources featured prominently when discussing the DHT's strategy for AwLs. The DHT felt AwLs enabled them to review, adapt and refine assessment provision to better suit their ethos. A positive outcome from these adaptations was making learning more measurable.

*...we've had to make a lot of changes in a relatively short period of time...they've impacted positively on our children and therefore the progress that they're making is much more measurable and accurate than it once was...*

It is interesting that the DHT used this vocabulary when describing the positive impact of AwLs on students, emphasising how measurable assessments were, comments reflective of a managerial discourse (Gipps, 2011).

Without government guidance for in-school SA the DHT referred to how the National Curriculum had been utilised as criteria to assess against and targeting teaching and student learning based on the outcome. As a result, they felt teachers had gained a better understanding of the curriculum through assessing against curriculum objectives.

*...now when we're assessing against individual objectives and it's much more specific, it means that we as teachers can identify areas of development which informs our teaching. (FS2DHT)*

Teacher SA of curriculum objectives was recorded using SPTO<sup>10</sup> however this was due to change, with the school shifting towards termly tests by NFER<sup>11</sup> tests. Externally supplied tests were seen as more suitable than teacher SA in measuring progress for cohorts as they

---

<sup>10</sup> SPTO (Student Progression Tracker Online) is an externally purchased assessment tracker where curriculum objectives for subjects can be individually assessed.

<sup>11</sup> NFER (National Foundation for Educational Research) provide tests for KS1 and KS2 SATs assessed subject.

progressed through school years. SPTO tracking would remain for teacher FA against curriculum objectives, where teacher assessment was better suited to formative purposes and tests better suited to summative purpose.

The use of tests for SA had increased as part of the DHT's AwLs plan but they felt these tests had a dual purpose, in addition to measuring attainment they were used diagnostically to inform future teaching.

*...I think AwLs has ensured that it's [FA] much more diagnostic...what we are now doing much more of is, when the children sit a summative test...we analyse them against each of the strands or content domain from the curriculum, and then that informs our teaching for the next half term or term...that's much more deliberate now than it ever was before.*

The use of formative assessment based on termly tests privileges learning which can be evidenced in a test and has implications of assessment wash-back. Teachers were required to analyse the termly test data, the DHT commented that this 'gets a groan' but that teachers 'are much more open to the analysis...they do see the worth...and the value in it' suggesting that there may have been resistance to this at some point.

A further change to FA practices was the introduction of 'responsive teaching' in afternoons. Children were given targeted teaching in core subjects, based on misconceptions or mistakes spotted in lessons taking place in the morning.

*...if a child has struggled with something in that particular lesson, on that particular day or maybe they didn't struggle but when you come to mark their book there's clearly a misconception, then the teacher on that day will identify them and ask one of our support assistants to work with them that afternoon...[this] has been so effective particularly in maths because it catches them right up to where they need to be ready for the next lesson...*

Daily 'responsive teaching' was felt to have had a 'massive' impact on learning evident by 'around 20% achieving the expected standard higher than the year before...' at the end of KS2. This inference assumes SATs could objectively measure learning, and although it is possible that standardised tests can capture improvements in attainment through timely

targeted formative assessment practices, the emphasis was on better SATs performance. The DHT did not discuss other factors which may influence SATs results, for example, a different cohort could have different abilities. The school had a larger than national average proportion of children working above age related expectations so there was a focus on 'pushing' children to the AwLs judgment of greater depth. Again, this focused on the assessment judgments rather than the learning itself.

*...so I guess that was our biggest focus, on ensuring that we are pushing through and getting depth from those children.*

#### 8.4.2 Tests objectively measure progress

Tests seemed to be increasingly relied upon as an accurate measure of progress in learning, with a suggestion from the DHT that teachers can be knowingly and unknowingly biased in their assessments. A main reason given for this was a belief that teacher SA was bias '...no matter what your intention, it may not be intentional...but you do have teacher bias.' DHT seemed to feel that teachers and schools are capable of 'manipulating' teacher SA. There was no blame on teachers, and the DHT acknowledged that it was difficult to avoid unconscious bias. However, they did strongly suggest teacher SA was 'tweaked' by schools, referring to this as 'playing the game', adjusting teacher SA 'in line with what they want them to be', therefore tests were said to be more reliable.

As a teacher themselves, the DHT did see situations where '...depending on the children I'm working with, I can argue my teacher assessment is more accurate than a test score' when students '...could just have a really bad day and...may crumble under the pressure'. Although, they felt the school could work to avoid test related stress.

*...part of that [test stress] is your culture as a school around testing and how we present that to our children so that they don't feel stressed at the site of a test.*

#### 8.4.3 Labelling or not labelling children

An effort had been made to avoid labelling AwLs judgments when discussing assessment with students, focusing instead on building a mind-set for learning.

*...we've spent a lot of time and effort around this idea of the learning pit...we talk about...areas that we need to get better at, areas we need to practice, next steps.*

Discussion around learning was influenced by the school's adoption of 'growth mindset' (Dweck, 2015; 2017), as a view of learning.

*...emphasising the success the children are having...oh they're not quite there yet...we use, yet, we're not there yet, or we haven't achieved this yet and the power of yet is that we will at some point.*

This can be seen in the terminology used by DHT with the emphasis on 'yet' and sharing with children that learning comes with challenges. This focus on process of learning rather than outcome was felt to have engendered a more positive attitude towards learning from the students.

*I think not only is it impacting positively on the progress they're making it's also impacting positively on their self-esteem, on the way that they feel about learning, and making a mistake and that's okay...*

#### 8.4.4 Parental engagement

Although, adapting to AwLs was challenging for some parents, the DHT felt this was something which was improving through workshops and school guidance leaflets. They felt parents were used to being given a specific level and sub-level assessment, and the shift towards more conversational assessment feedback was difficult. At the point of interview the DHT felt parental understanding of AwLs was split between, parents who understood 'the objectives that pupils have to achieve and appreciate the detail that we're able to provide' and those who were still seen as struggling to understand. In hoping to further engage parents to help them understand the more recent changes in school assessment, they were planning to take the lead from parents in terms of what engagement would be useful for them.

*But we will, we are planning to engage parent voice to say what they feel, what would be most effective way to do that.*

#### 8.4.5 Summary of SLT S2 interview themes

From the themes analysed from the SLT interview in S2, the shift towards AwLs provided an opportunity to reshape assessment in focusing on learning. Success in SATs was a driving force behind their assessment practice, increasing the formative use of summative tests, which privileged learning which could be demonstrated under test conditions. There was a development of timely teacher FA used to target misconceptions quickly, to ensure students could progress with future learning. Additionally, an emphasis was placed on student attitudes towards learning, with a focus on embedding a 'growth mind-set' towards learning.

### 8.5 S2 TEACHER AND SLT INTERVIEW THEMES TRIANGULATION.

#### 8.5.1 Overlapping themes

Having a higher confidence in testing than teacher SA was a shared view, with the belief that teacher SA was subject to conscious or unconscious bias. The reliability and comparability of SA was prioritised with little consideration of the validity of tests. Testing favoured writing skills and areas of the curriculum which could only be captured in that format (Harlen and Gardner, 2010; Stobart, 2005; Tennent, 2020). Teacher G, who was in their 2<sup>nd</sup> year of teaching, had the strongest voice in support of teacher SA, viewing testing as unsuitable. The DHT shared these views but interestingly only when talking as a 'teacher', and not as assessment lead.

#### 8.5.2 Conflicting Themes

The theme of evidencing teacher SA and the impact that it had on workload was not raised as an issue by the DHT but was by the teachers. Similarly, the DHT discussed the requirement for teachers to analyse termly tests for planning purposes feeling it was worth the 'groan' it got from teachers though did not validate Teacher H's comments of the time this added to workloads. This indicated a disconnection between the DHT and the teachers in terms of the personal cost of adopting the AwLs practices. Teachers as explored in [Chapter 9](#), showed little agency in terms of complying with evidenced required. This was reflective of the doxa of performativity referenced by Pratt (2016: 896) where teachers 'must do accountability work'

### 8.5.3 SLT impact on teacher Themes

The interview with the DHT cast a positive light overall on assessment changes and the impact that they had on SATs results, and therefore represented both improvements in pupil learning, and being able to capture learning in a test environment. Although the teachers agreed that NCLs were no longer relevant, data from their interviews was more mixed. Frequent changes in resources adopted for assessment was unsettling for Teacher G. Both teachers felt that assessment of writing encouraged technically good writing, but that this was achieved at the expense of fluent writing. There were significant downsides to AwLs not presented in the SLT interview. This suggests an amount of cruel optimism (Moore and Clarke, 2016) on the part of the DHT in that, benefits they believed AwLs could bring where from the teachers' perspective undermined some of those benefits such as reducing TTTT.

The increase in test based non-statutory SA outlined by the DHT was supported by Teacher H, however this seemed to be from an internalising of Policy discourse that teacher assessment is biased and no longer trusted because of its subjective nature even in the case of formative assessment (Christodoulou, 2017). Supporting this inference, Teacher H was adamant about the *need* for evidence to support judgments. By increasing the amount of testing, the DHT may be reinforcing this discourse. Teacher G differed, feeling their teacher SA, based on observations and questioning, was far more detailed, and allowed for assessment to be based of a broader sample of the curriculum than tests, favouring the validity of teacher SA (Black et al., 2011; Gardner, 2010).

## 8.6 CONCLUSION

It was concerning that the thematic analysis of interview data revealed a shared criticism of AwLs practices which impacted on teaching. Aside from the obvious requirement of teachers to follow both national and school policy, fundamentally there was extremely limited evidence or awareness of choice when adopting SATs preparation strategies. This was attributed to how tied SATs results were to school and teacher accountability resulting in stress for both teachers and students, for example, teaching to the test (TTTT) at the expense of educational values as beliefs such as providing a broad and balanced curriculum. From these conclusions, the amount of choice and control teachers had over the impact of

AWLs on teaching practices was of particular interest as it could provide insight into how teachers *could* enact AWLs policy without compromising the educational values they held.

## 9 TEACHER AGENCY ANALYSIS AND DISCUSSION

---

### 9.1 INTRODUCTION

This chapter will analyse and discuss teacher interview data using the theory of agency and identity set out in [Chapter 5](#). Using this theory, the analysis will first introduce the [Categories of agency identified](#) from the teacher interview data with reference to relevant literature to provide context to each category. These categories of agency are then presented in detail using teacher interview data and the integrated agency and identity theory from Figure 5-2 in [Chapter section 5.3.1](#). Teacher agency varied according to their engagement with different temporal environments, participants did not just display one type of agency. This means participants are discussed, when applicable, in each category of agency identified. Integrated within this analysis of agency will be the implications of neoliberal discourses explored in [Chapter 4](#) on achievement of agency within these categories; namely these are marketisation, competition, accountability and managerialism. The chapter will conclude with an integration of neoliberal theory into the agency and identity theory from Figure 5-2. Theorising the impact neoliberal discourse had on participants' achievement of agency assessing without levels.

### 9.2 CATEGORIES OF AGENCY IDENTIFIED

Analysis of the interview data using the theory set out in Figure 5-2 produced five categories of agency achievement among the teachers interviewed. These have been categorised as;

#### **Conflicted compliance**

Where teachers complied with AwLs practices but demonstrated conflict with their educational values. Similar to unhappily compliant (Moore and Clarke, 2016), the word 'conflicted' represented how teachers complied with assessment practices, often undecided if this compliance was right, and if their opposition it was justified.

#### **Non-conflicted compliance**

This category was derived from the opposite of conflicted compliance. Compliance was non-conflicted, with teachers showing no resistance. This did not mean than teachers did not

identify draw backs to adopting certain AwLs practices, but they seemed to accept them as part of their professional identity rather than question them.

### *Espoused Theory of Action*

Interestingly there were occasions where the teachers discussed a particular AwLs practice but also provided examples of assessment and teaching practice contradicting this.

### **Passive compliance**

This category represents a compliance but passively with no awareness of choice or alternatives.

### **Strategic compliance**

Agency was seen as compliance but in a strategic manner. Here, the consequences of resistance in openly disagreeing with AwLs practices are avoided through complying in a strategic way while allowing scope for other educational values and aims to be met (Hoyle and Wallace 2007).

### **Agentic Resistance**

The achievement of agency was generally seen in two forms, 'stepping up or pushing back' (Buchanan 2015: 709). Pushing back is agency as resistance, stepping up was demonstrations of agency when adopting a practice required but forming and shaping in a new way. Priestley et al. (2016) referred to this as creative mediation.

Each of these agentic categories will be discussed in more detail below with reference to interview data which exemplify them. At times this will mean reusing quotes from the thematic analysis because in this chapter they emphasise a different part of the analysis focused on teacher agency.

#### **9.2.1 Conflicted compliance**

AwLs impacted on teaching practices which were questioned but still complied with, resulting in little achievement of agency beyond questioning. Although not always viewed as achieving agency, this questioning of the impact of AwLs was valuable because it showed where teachers were able to envision possibilities beyond what they were currently

complying with - it showed where there was strength in the teacher projective domains, a vital component of agency.

At times, the conflict of compliance extended to teachers feeling uncertain as to whether their disagreement was warranted. Examples of this were seen in teachers questioning their own values and beliefs where they conflicted with AwLs practices. This is suggestive of a conflict between their professional teacher identity, and practices now required with agency conceptualised as an active rebalancing of this conflict in response to disruption (Beauchamp and Thomas, 2009; Biesta et al., 2017). Therefore, conflicted compliance could be indicative of attempts to resolve conflict between *previously* constructed teacher professional identity and what it *now* needs to embody in terms of values and beliefs, to align with AwLs practices.

#### 9.2.1.1 Teacher F

Teacher F was critical of KS2 SATs tests, feeling that they inaccurately represented the curriculum they assessed and caused significant stress for students. They argued that teacher SA was a better alternative for KS2 assessment, feeling strongly that as a teacher they knew their students, and that their SA was more valid than SATs tests.

*Teachers know their children, they know whether something is a one-off or whether the children understand things...in the test it's just what that child could do in an hour on that day. And preparing children to pass a test is not the same as teaching objectives in my opinion. (FS1Y6T2F)*

Regardless of Teacher F's beliefs, they engaged in test preparation strategies to maximise SATs outcomes for their students. This is reflective of the competing demands on teaching to educate, and teaching to be seen to be 'successful', evidenced by standardised testing as part of the managerial model prioritising performance driven practices such as test coaching (Biesta et al., 2017; Stevenson, 2010). Teacher F demonstrated this below where they described preparing students for the new grammar SATs test, while pointing out that they felt students would gain little from it.

*...we had to teach them huge amounts of things, extremely difficult things, in a very small space of time for the purposes of passing a test. And things like*

*the grammar they don't need to use it, it's not beneficial, particularly, but it's for passing a test. (FS1Y6T2F)*

As discussed in the chapter Teacher agency and identity, early teaching experiences exert a strong influence on a teachers' professional identity, and these seem to have provided Teacher F with an alternative picture of assessment. There was a strong sense of what assessment *should* be based on in these experiences where teacher accountability is viewed through an auditing lens as a means of being answerable to management in terms of the effectiveness of their teaching (Green, 2011). Requirements to evidence teacher SA conflicted with these early teaching experiences where they felt trusted, compared to now where they did not.

*...if governors see that I've assess somebody working at expected standard in maths and they didn't pass the test the governors think they should match...obviously the question is always where is the evidence, show me the...When I think back to how when my head spoke to me about assessment when I was in NQT and he told me you assessed by your gut instinct and then having tested that out like for writing for example, if I had then gone ahead and done it on a spreadsheet or whatever you would get the same outcome nine times out of 10 (FS1Y6T2F)*

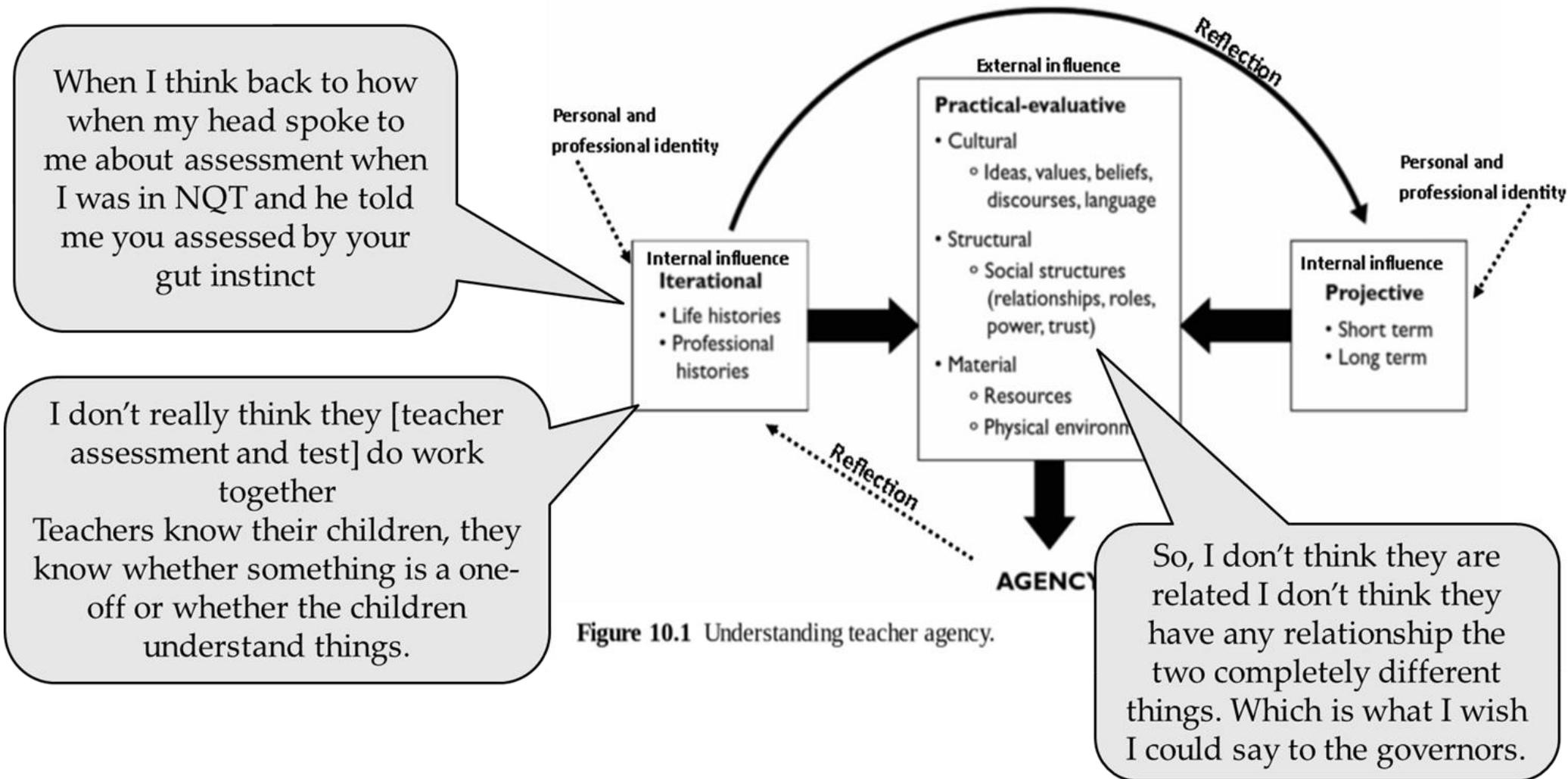


Figure 9-1 Comments from Teacher F used to indicate how agency achievement was enabled and constrained using Priestley et al. (2013: 152) ecological theorisation of agency.

The annotated Figure 9-1 shows where these comments from Teacher F indicated conflict between their iterational domain and their practical-evaluative domain; in the former, early teaching experience were still influencing their teacher identity, in the latter cultural values and belief of governors and SLT did not align with Teacher F's own. Furthermore, the pressure felt to both evidence their teacher SA and for it to concur with SATs test outcomes, suggested structural constraints from neoliberal discourses of managerialism and an audit-based accountability requiring evidencing to be trusted, restricting achievement of agency. Thus, despite teacher SA supporting their educational values and beliefs, Teacher F was constrained by the school's views of assessment and lack of opportunities for teachers to shape school AwLs policy.

#### 9.2.1.2 Teacher E

Teacher E demonstrated conflicted compliance when assessing writing based on a secure fit assessment of the ITAF objectives. They felt this impacted on both the enjoyment of writing for students, and also the creativity of the writing produced, both things they valued highly as a teacher.

*...the interim assessment framework...I find that's too heavily weighted towards punctuation and doesn't give the children the opportunities to be as creative as what we want them to be. So they end up going to secondary school without that creativity...I just, I don't agree with all the drafting and redrafting that we end up having to do and the children, it takes enjoyment out of writing because the children are constantly trying to edit and improve their own writing (FS1Y6T1E)*

Teacher E also described tensions they felt between the required level of independence and how they felt they should teach. In S1, as reported in SLT interview findings, there was an emphasis for classroom activities to be independent and reflect the environment of testing to improve SATs test results.

*I suppose in one way we're not doing them any favours for the test but you want them to be able to articulate what they're doing and reason that out...Maybe we're doing them a disservice by trying to get them to just do it*

*in their head, I don't know. Then you're coming down to test technique rather than actually teaching them. (FS1Y6T1E)*

In providing opportunities for collaboration during lessons, indicating social aspects of learning were valued, Teacher E demonstrated uncertainty in knowing the correct approach for the students. If they provided space for students to articulate their thinking and reasoning, they felt they were not 'doing them any favours for the test', on the other hand, if they shifted to more independent activities, they questioned if they were 'doing them a disservice'. This example from Teacher E demonstrates the conflict which can occur within teachers' professional identities from National policy changes mediated through school ethos and values influenced by neoliberal discourses of SATs based accountability required performative behaviours of the teachers (Beauchamp and Thomas, 2009; Biesta et al., 2017; Buchanan 2015). As a result of AwLs, S1 emphasised the need for explicit and visible teaching along with assessment prioritising independence. This represents mediation following the reform removing guidance and mandatory submission of teacher SA, to ensure student learning can be demonstrated confidently under test conditions. Teacher E felt pressure to ensure lesson activities were more independent, but their teacher training and earlier teaching experiences, which help build professional identity, seemed to be in conflict with this expectation of the school teaching writing without levels.

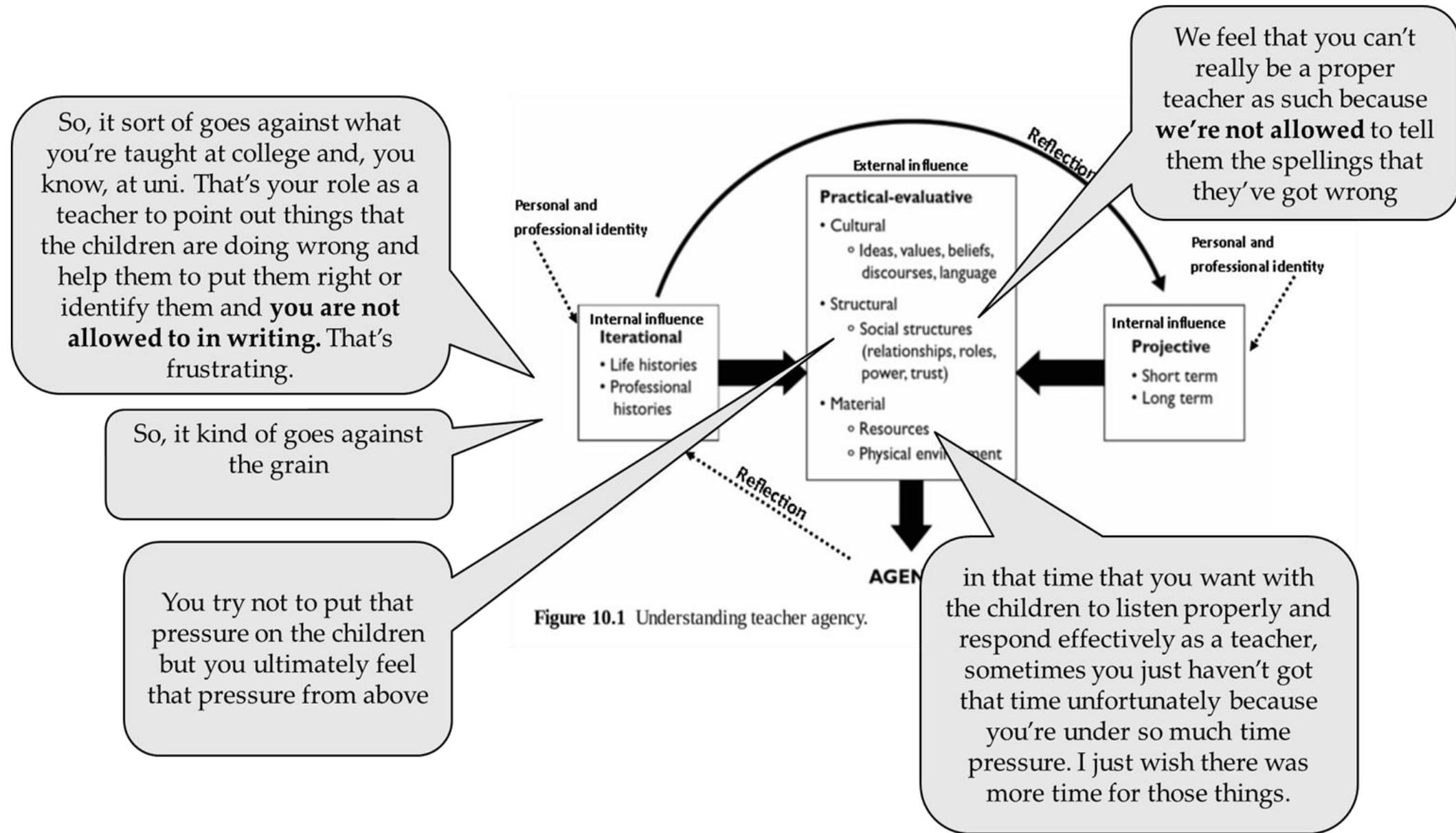


Figure 9-2 Comments from Teacher E used to indicate how agency achievement was enabled and constrained using Priestley et al. (2013: 152) ecological theorisation of agency.

From the annotated Figure 9-2 with extracts of Teacher E's comments, previous life history and professional history, iterational domain, was strong in terms of agency. They were clear in their role as a teacher however, agency achievement seems restricted by their practical-evaluative domain, namely the school's requirement for teacher SA to be based on evidenced, independent activities and aligned with SATs test results for it to be viewed as reliable, and a lack of time to teach anything in depth. Similar to Teacher F, previous experiences informing their professional identity conflict with neoliberal discourses evident from the SLT in terms of auditable accountability, thus evidencable teaching reported by Quennerstedt (2019) as part of AwLs teaching.

The pressure described as coming from 'above' indicates SLT, although they are not directly mentioned. Regardless, there was external pressure felt by Teacher E and their comments suggested that without such pressures they would teach differently. Teacher E's comments indicate some internalisation of the teaching approaches they disagreed with which suggest a rebalancing of their professional identity to align with AwLs practices within the schools emphasising evidencing and TTTT practices. They suggested that in guiding the children in lessons, their teacher SA grade would have been inflated, rather than this guidance scaffolding student learning.

*...we would have guided them too much. I can see where they're [SLT] coming from but it's just so frustrating. It just seems to make a mockery of the whole process that another child can do it [give feedback] for them...(FS1Y6T1E)*

This is itself conflicted, as Teacher E suggests they would be providing the same feedback to students as their peers, and they did not suggest that that would have 'guided them too much'.

#### 9.2.1.3 Teacher D

Similar to Teacher E, Teacher D also felt conflicted about the requirement to evidence. For them this included not only their teacher SA, but any lesson activity. They discussed the added workload in complying.

*My judgement isn't good enough anymore, you've got to prove what you've done...I work every night. There's not really a night I don't work. And because you've got to evidence everything, that's more written things, so that's more*

*marking. You can't do a whiteboard lesson or a practical lesson, you've got to have a photo and a 'next step' in to say you've done this. (FS1Y2T1D)*

This conflicted compliance seems driven by aspects of their teacher identity, that teachers *should* be trusted.

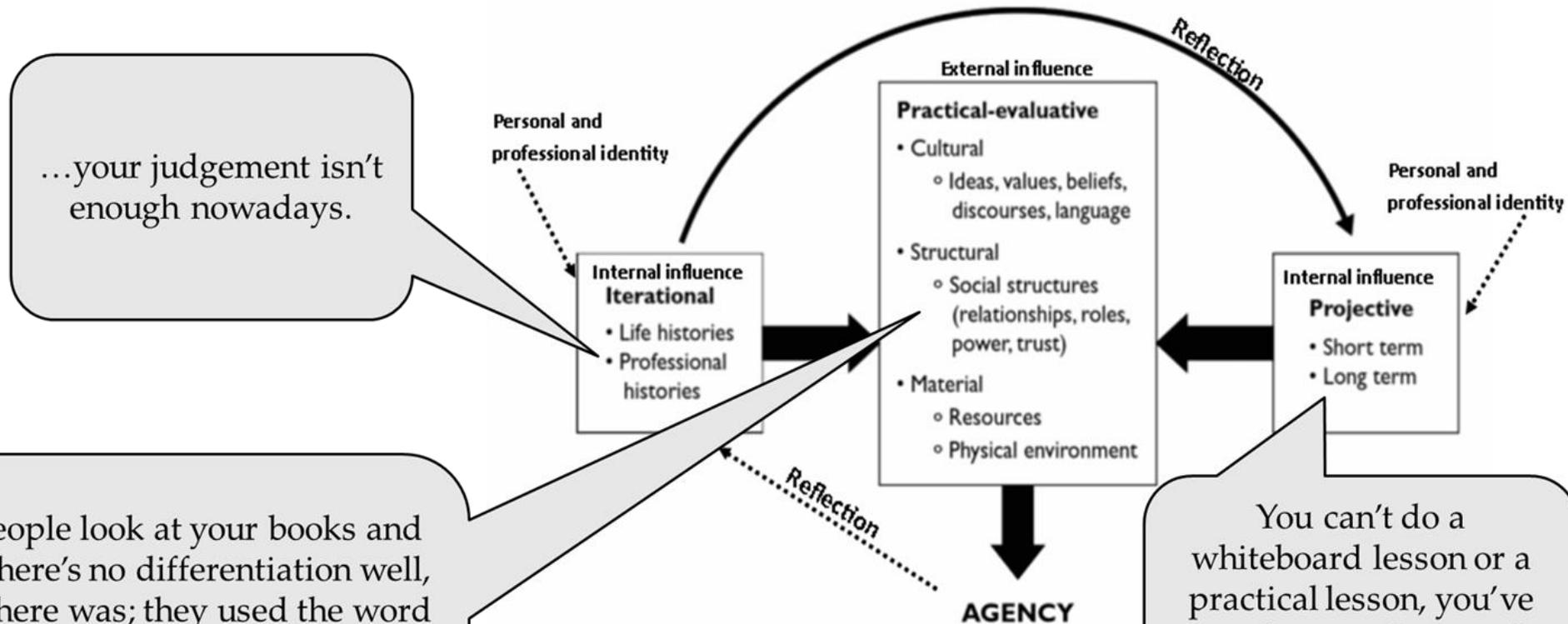


Figure 10.1 Understanding teacher agency.

Figure 9-3 Comments from Teacher D used to indicate how agency achievement was enabled and constrained using Priestley et al. (2013: 152) ecological theorisation of agency.

In terms of any achievement of agency, as shown in Figure 9-3, although their teacher identity informed their iterational domain through their values and beliefs, agency achievement was restricted by their projective domain, as they stated, 'you've got to', 'you can't do', indicating that they saw no alternative practices in the short term. Although not discussed explicitly, the need to evidence may be a result of school policy and school monitoring limiting teacher agency through a lack of teacher autonomy. They acknowledge the pressures schools and teachers face around SATs which may contribute towards the managerial approaches requiring learning and assessment to be explicit and visible, however, this emphasises a critical point about managerialism and education, namely that if something cannot be measured then it cannot be managed, favouring forms of learning and assessment which are both visible and easily evidenced (Green, 2011). Teacher D's comments above refer to how this impact on them through SLT and their comment below displays an understanding that external pressures on SATs impacts the SLT.

*...Yr6 SATs results tell you whether you're a good school or not, but actually it's not about that and you end up teaching to the test because you've got to get that result because you've got to. You don't want OFSTED on your back.  
(FS1Y2T1D)*

Conflicting demands were discussed between compliance in SATs based accountability practices to avoid punitive repercussions and disagreeing that these practices made a 'good school'. Interview data from S1 conveyed the challenge of teaching the elevated difficulty in the curriculum with a discourse embedded that in doing this student attainment could improve. Teacher D showed conflict between this view of learning and their own view that children go at their own speed. They stated '...they've [children have] got to get further now than before so you're pushing them a lot' (FS1Y2T1D) but also that 'I do think it's [curriculum difficulty] pulled it up, but again, children are children; children don't change because we're teaching them harder stuff' (FS1Y2T1D).

There was strong a conflict for Teacher D in complying with the all or nothing 'secure-fit' assessment of writing using the ITAFs and how the government handled the transition towards AwLs in terms of the guidance given.

#### 9.2.1.4 Teacher G

ITAF focused writing assessment resulted in similar conflicted compliance in S2 as well as S1. When discussing the assessment of writing, Teacher G challenged the accuracy of ITAFs to judge KS2 SATs writing. Complying resulted in tension between what Teacher G believed represented creative and fluent writing, and writing which was technically perfect according to the ITAF based judgement.

*...we had our ITAFs that we'd filled out so we knew which children were there, which ones weren't...It seemed to be assessing children on how well they can fill out the checklist rather than how fluent their writing is or how cohesive it is or how they adopt a particular tone in a certain piece of writing...*  
(FS2Y6T1G)

As a recently qualified teacher, Teacher G's teaching experiences comprised their teaching degree placements and NQT year. There was little discussion of alternative options for in school assessment despite a belief that teacher SA should complement SATs results, either in the short-term or long-term. A reason for this may be the constant anticipation of change Teacher G discussed so they were unable to envision what assessment could be in the long-term, while still needing to adapt to the most recent changes. Furthermore, the external pressure of moderation, which focused specifically on the ITAF, was indicative of a lack of resources to assess in their practical-evaluative domain such as alternative discourses.

##### 9.2.1.4.1.1 *Espoused theory of action inconsistencies with theories in use*

At times, assessment and teaching practices the teachers claimed to use differed from those they described using. Argyris and Conant (1985) provide a useful concept for this which refers to two types of 'theories of action' people can present:

*Espoused theories are those that an individual claims to follow. Theories-in-use are those that can be inferred from action. (Argyris and Conant 1985: 81-82)*

When these two 'theories of action' agree they are referred to as consistent, times when they do not agree they are inconsistent. Although aware of their espoused theory of action through their claim to follow it, individuals may not be aware of inconsistencies between

what they espouse, and what they actually do their 'theory in use' (Argyris and Conant 1985: 82).

#### 9.2.1.4.1.2 Teacher H

When discussing summative assessment in two forms, a test or teacher SA, Teacher H expressed confidently that 'Everything is best suited to teacher assessment as long as the teacher is using a test', additionally referring to standardised testing as the only reliable form of assessment due to discrepancies between TA and SATs tests. However, an example was provided by Teacher H where they felt unable to form a teacher SA for a student due to a lack of written or verbal work in lessons, but the student achieved an expected standard on the SATs test. This is an interesting inconsistency where Teacher H espoused that all TA should be test based, but in action they required more in terms of verbal interaction and lesson activities to support their teacher SA. With the context of SATs tests they explained that they felt SATs testing could not represent the entirety of the maths curriculum. Moreover, the submission of KS2 teacher SA judgments to the LA had just been announced as changing to non-mandatory, which could further reinforce this teacher's espoused belief in privileging testing above teacher SA.

*They're [government] not really interested [in teacher SA], they're going to use a SATs test or whatever. Which is sad because national curriculum testing in the context of maths, say national curriculum reasoning, is not the same as standards and testing agency reasoning, it's not. You look at the definition of the curriculum, you look at the content of the test, the curriculum talks about being, following lines of enquiry and investigated work. (FS2Y6T2H)*

#### 9.2.1.4.1.3 Teacher E

Another example of inconsistencies between theories of action could be seen in Teacher E's discussion of a student, who, according to their teacher SA, was expected to pass the maths SATs test, but then did not. Teacher E recalls 'in action' that work in lessons and a practice test had not indicated that the student would struggle, implying through that description that they believed the basis of the teacher SA was sound, therefore they could not have known the student would not pass the test. However they 'espouse' that their teacher SA should be more accurate.

*...she gave me no sign throughout the practice test that we'd done in the lesson...she'd grasped everything, she could explain things, she had good reasoning skills and she just completely bombed on the test...but she didn't make me aware at all unfortunately that she was struggling...I think teacher assessment should have been more spot-on but it wasn't.*  
(FS1Y6T1E)

The school's belief that teacher SA should align with SATs testing could be supporting Teacher E's espoused theory that their SA should have identified that the student would struggle on the SATs test. In action, Teacher E explains that this was not possible with the first indication of struggle coming from the SATs results themselves.

#### *9.2.1.4.1.4 Teacher D*

Teacher D had begun to develop an identity as a moderator for KS1 SATs through their moderator training. This provided them with 'a bigger picture' of assessment than before. When discussing moderation and the assessment guidance comprising of ITAFs and exemplars, Teacher D often referred to the concerns of how other teachers were assessing, for example;

*I think it's just that worry that people are just going to down the ITAF route which I can understand...people are falling into the trap of, they just do whatever worksheet was done for the exemplification...people are just getting tunnel vision because this is the moderation, this is the exemplification, we're going to do this.*

They 'espoused' that other people have relied heavily on the ITAFs, and that their school and themselves have 'seen' a bigger picture. However, when asked about their teacher SA without levels, they stated 'we live our life by the ITAFs' and described how in action their teaching of writing is formed around the ITAFs describing the teaching of writing impacted by the ITAFs producing a competency-based framework, as opposed to a more holistic view of writing.

*Now it's about ticking boxes. So, you are reading a piece of work, you are not looking at has that got flare, has it got spark in it. (FS1Y2T1D)*

#### 9.2.1.4.1.5 Teacher G

Teacher G held espoused beliefs regarding the suitability of SATs tests in representing curriculum content. They expressed beliefs that SATs tests can adequately capture student attainment, albeit for students who were not anxious about testing, this still conflicted with views that such tests could not represent the whole curriculum, and therefore teacher SA was required to represent student learning.

*...you need teacher assessment too, there are certain skills that can't be assessed by a formal assessment. Things like verbal reasoning and certain types of problem-solving skills that a test, that will never...for more complex tasks and verbal reasoning I think that's where teacher assessment comes in.*  
(FS2Y6T1G)

Their concern over the removal of teacher SA at the end of KS2 further showed an espoused belief that teacher SA is an important aspect of assessment practice compared to their previous statement that tests were only impacted by nervous children. The DHT's comments that TA was intentionally or unintentionally biased, may influence these conflicted views on the place of teacher SA.

These contradictions between what the teachers espoused and what they did in action could be the result of the conflict with the teachers' identity and external influences becoming normalised as part of a rebalancing.

As teachers were conflicted about their compliance with teaching practices driven by the marketised pressures on statutory assessments these tensions between what is said to be believed, and what is described as being done, are symptomatic of the conflict between the teachers' internal influences and the external competing demands of standardised assessment on their professional identity. Identity theory, explored in [Chapter 5.2](#), highlights the need for identities to be in balance and not remain conflicted (Day et al., 2006), thus when theories of action are inconsistent with espoused theories this suggests teachers are in the midst of reforming and rebalancing their professional identity, making the struggle visible (Ball, 2003; Buchanan, 2015). The result of this can be seen in the internalisation of performative behaviours as part of what it is now understood as practice as a professional teacher.

### 9.2.2 Non-conflicted compliance

The emphasis on ‘mastery’ of core curriculum knowledge encouraged a move away from differentiation of curriculum content, and instead supported exposure of all core curriculum content regardless of ability. This aspect of education policy reform did not produce much conflict for example, a move away from differentiation of content for ability groupings was supported by Teachers H and E.

*Maths, it's a mastery curriculum so it's, it says all pupils must, must achieve this. That's the aim. I'm not going to help those pupils by giving them work that's below the age-related expectation ever. (FS2Y6T2H)*

*...the way we've done it here is that children are still, they might not be able to use capital letters, but they're still trying to do shifts in formality because we've talked to them all about shifts in formality. (FS1Y6T1E)*

This teaching adaptation seemed motivated by what was best for the pupils by exposing them to the more challenging curriculum content. However, it directly reflected discourses from the reform of heightened expectations for students regardless of starting point and was, by these teachers not contested. Reflecting on the recommendations of the NCR (2011) high expectations for the outcome of better attainment was rooted in practices such as additional support for those who required it and catching students up if content was missed. The emphasis here is on pedagogy rather than exposure. It provides an additional perspective on high expectations that was omitted from the recommendations taken from the NCR (2011) framework. Instead, the policy view of high expectations was taken of exposure, to all students, of a curriculum elevated in difficulty regardless of starting point (DfE, 2010b). With this view from policy makers being driven by the need for education to compete globally as an economic product (Exley and Ball, 2014; Bailey, 2004; Baltodano, 2012; Ball, 1998); with little increase in resources for schools.

#### 9.2.2.1 Teacher D

The purpose of the KS2 SATs as an accountability measure was questioned by most of the teachers, but the attitude towards KS1 SATs was different for Teacher D. They acknowledged the stress and pressure of KS2 SATs, but found KS1 SATs to be less pressured in practice. Although there were still SATs test at the end of KS1, Teacher D was able to

factor in their TA. This could account for less resistance or conflict when complying with AwLs testing, having instead the option to complement SATs tests with teacher SA judgments with less impact from test-based accountability as experienced by KS2 teachers. Other assessment resources were available to Teacher D, thus reducing the conflict for them between belief in teacher SA, and being held accountable with the KS1 SATs tests. Subsequently, KS1 SATs tests were viewed positively as confirmation of their teacher SA.

*I don't know why everyone is fighting against it [KS1 SATs tests] because I think people feel happy that they've done them and backed up [teacher SA].  
(FS1Y2T1D)*

Being recently trained as an external moderator for KS1 SATs may have provided new external influences for Teacher D's professional identity, developing a sub-identity as a moderator (Beauchamp and Thomas, 2009). They described seeing the bigger picture now, 'So, maybe now, I'm actually...seeing the bigger picture as well. So, then it's a bit frustrating that not everybody does.' However, there were indicators of conflict between these identities, that of the classroom teacher identity and that of the new moderator identity. For example, by their own admission, they too referred to adopting teaching practices prioritising ITAF content ensuring its coverage in preparation for KS1 SATs despite their criticism for example of writing assessed using the ITAF as presented in the [previous section](#). This demonstrates how important context is to the varied degree of agency achievement.

#### 9.2.2.2 Passive compliance

Occasions where teachers demonstrated compliance of assessment and teaching practices with no indication of awareness, conflict, or questioning of discourses such as performativity, were considered examples of passive compliance. In particular, teachers demonstrated an internalisation of teaching and assessment practices associated with performativity, TTTT strategies and an acceptance of test-based accountability as a measure of school success. In these situations, there was limited to no effect from the iterational or projective domain assuming aspects of teaching have been internalised into their teacher identity. In particular, Teachers D, E and G seemed to accept the impact of test-based accountability, which they demonstrated through comments implying that it was their

responsibility as a teacher to ensure that students were prepared for SATs without levels, even when unaware or unclear of what that preparation needed to be. For example,

*I think it was March or something it came out, and then it was like, oh my gosh I haven't done this... they were supposed to have done this and they haven't and that's my fault but I didn't know I had to do this. (FS1Y2T1D)*

*I suppose the formal SATs are more to allow comparison between schools and... so that the secondary schools know how children perform under test conditions (FS2Y6T1G)*

Teachers D and E both accepted responsibility for not knowing to target areas of the curriculum which represented in the interim assessment frameworks. The sense of panic these two teachers conveyed demonstrates how normalised and accepted it was to teach according to the SATs assessments.

There was acceptance from Teacher G, in the quote above, that the SAT's function was to compare schools, rather than representing summatively student learning; SATs seem accepted as a performance measurement. As a recently qualified teacher in their early 20s, their own school experience would have been influenced by New Labour and their drive for higher accountability and improvement focusing on the SATs results in Yr6. Teacher G may not have experienced accountability in any other format than high-stakes and test based. This discourse of accountability could have been dominant through their education, teacher education and early teaching experiences, therefore it may be difficult for Teacher G to envision an alternative system of education, as Priestley et al. (2013: 157) argued, they do not have '...the language with which to engage critically with policy', through limited access to educational discourses other than those dominant at the time; which in this case is SATs based accountability for schools.

Curriculum narrowing as a result of SATs based accountability was accepted by Teacher E, who did not question this aspect of SATs preparation, discussing its specific use for preparing 'lower ability' students.

*...I definitely think it [ITAF] does narrow the curriculum down...for those children who we were worried about, are they going to pass or fail it [ITAF], in*

*want of a better word, [we] narrowed the curriculum that we taught them because we knew that that was the basics basically that they needed to be able to cover in year six...some of the children that we knew were going to struggle that's purely what they focused on... (FS1Y6T1E)*

In targeting specific content weighted highly in SATs, and ensuring students have experience expressing subject knowledge in a test environment, Teacher E accepted performative practices, resulting in inauthentic teaching practices to meet SATs performance target (Wilkins 2015). This is referred to by Moore and Clarke (2016: 674) as a 'Trojan horse' of professionalism where there is,

*...an agreement, however, reluctantly undertaken, to the implementation of certain externally imposed curricula and pedagogies that may be underscored and driven more by perceived (and perhaps misrecognised) national or international economic and market demands than by strictly educational ones.*

Thematic analysis of interview data from both settings demonstrated a specific focus on KS2 SATs preparation. Practice testing and reallocation curriculum was embedded in Yr6 teaching. It is not surprising therefore that Teacher E's goals were short-term SATs results focused, rather than long-term, motivated by a 'deep consideration of the purpose of education' (Biesta et al., 2015: 636), thereby limiting the achievement of agency within the projective domain. This is attributed by Biesta et al. (2015: 638) to accountability systems which '...prioritise and value certain modes of action over others'. In this case, teaching practices which prioritised and valued learning which *could* be measured well in SATs were demonstrated, representing an external influence on Teacher E's professional identity which may now be internalised, in other words performative teaching practices derived from SATs accountability pressures have now become part of Teacher E's professional identity. This was further demonstrated through accepting that it was the school's responsibility to prepare children for unintended consequences of test environment. For example, it was the job of the school to mitigate stress and pressure felt, by some students, which impeded SATs performance.

*I think the whole nature of like I say SATs week and being really nervous... We had a couple of children that just didn't do the time very well even though it's something that we've practised and rehearsed and, on the day it just didn't go to plan... Something to think about for us, next year (FS1Y6T1E)*

The importance of SATs in the monitoring of schools may limit the projective domain of Teacher E in being able to envision alternative futures where these pressures do not exist. Moore and Clarke (2016) posit that dominant discourse can become embedded into teacher identities through internalisation of pressures they put onto teaching practice; in essence, Teacher E's longstanding experience with SATs based accountability may mean SATs and associated pressures have become part of what it means for them to be a teacher.

### 9.2.3 Agency

The policy reform enabling schools to have more autonomy over their own assessment practices did not allow the teachers autonomy over their assessment practices. There were few conducive environments that afforded agency. Only 2 teachers were analysed as displaying achievement of agency, Teacher H, and on one occasion Teacher D. Both of these teachers had additional roles other than class teacher, and thus had opportunities to observe and discuss assessment in situations outside of their classroom. This afforded access not only to additional discourses of assessment, but also access to resources and more power within the school structures to shape assessment and teaching.

#### 9.2.3.1 Agentic resistance

Teacher H displayed criticisms of the assessment of writing without levels similar to the Teacher G, the second Yr6 teacher in S2, however they had experienced writing assessment with levels providing an additional frame of reference which Teacher G did not. With experience of assessing both with and without levels, Teacher H preferred the assessment of writing with levels because of the effect AwLs had had on their students' writing. The criteria for assessing writing without levels was felt to be insufficient, with their experience of moderation reinforcing this view.

*This is where if I could go back I'd rather do that...with levels, I'm not saying it was great, but compared to the current assessment arrangements, which are a dog's dinner... (FS2Y6T2H)*

Resistance with Teacher H was not confined to critique and questioning of writing AwLs, but was also indicated as agentic resistance, where they stated;

*I despise the current assessment regime, with a passion. To the point where I actually said I wasn't going to work in year six again because, I just feel so strongly about it. (FS2Y6T2H)*

Recently being appointed as maths coordinator, it was not clear whether they pursued this role as an alternative to teaching in Yr 6, or if the change in position was influenced by a more general career progression motivation.

In regards to writing, Teacher H envisaged improvements if a comparative judgement model was adopted. Comparative judgement was suggested as a more reliable form of assessment for writing by Teacher H, feeling that something so subjective as writing was not suited to being assessed by fixed criteria such the ITAFs.

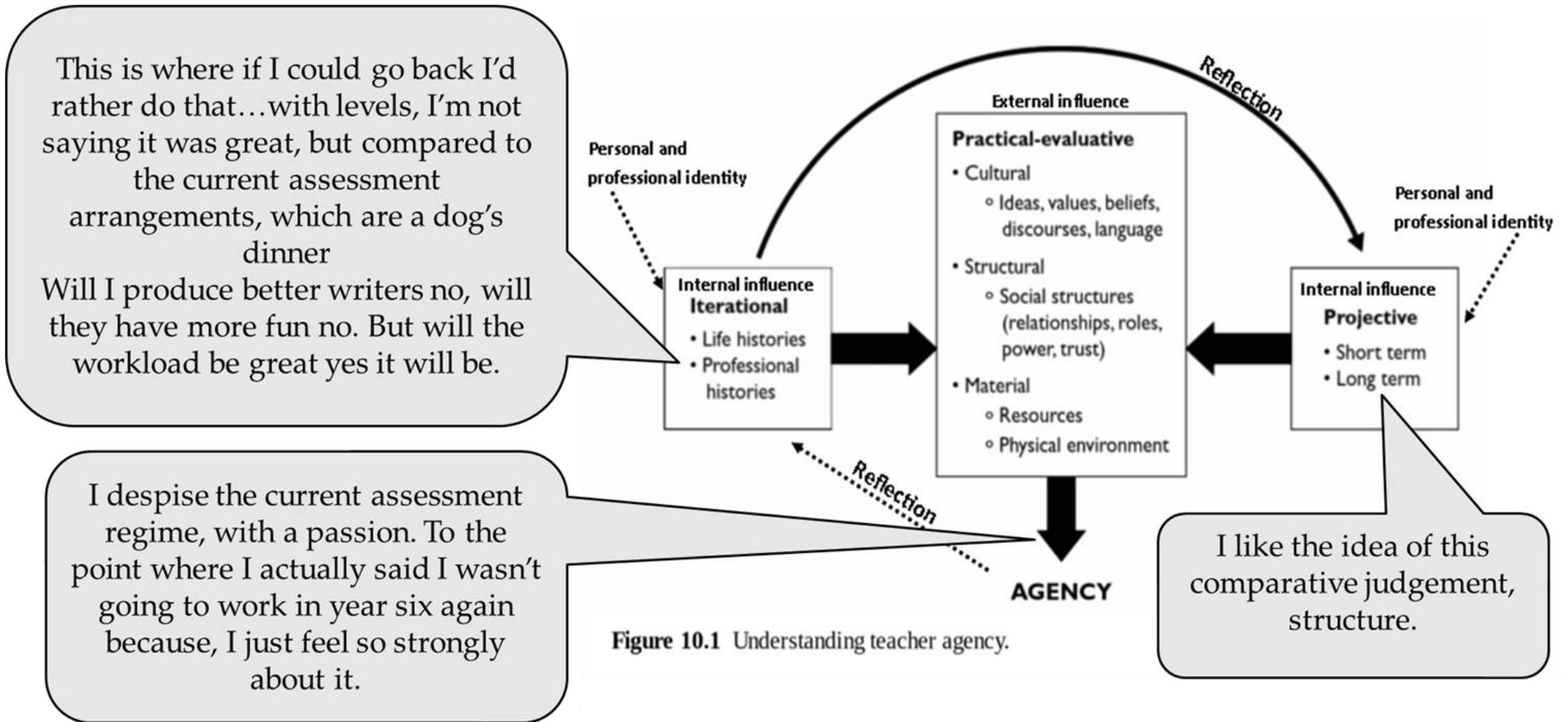


Figure 10.1 Understanding teacher agency.

Figure 9-4 Comments from Teacher H used to indicate how agency achievement was enabled and constrained using Priestley et al. (2013: 152) ecological theorisation of agency.

Teacher H achieved agency by pursuing teaching outside of Yr6 and being appointed as maths lead. The annotated Figure 9-4 displays where Teacher H's comments indicate enablements in all 3 domains, iterational, practical-evaluative, and projective. Teacher H had exposure to assessment of writing outside of AwLs. Their professional history informed by previous experiences of assessment represented a better alternative for the assessment of writing. Central to their teacher identity was the belief that assessment should have a formative role in students learning; they felt that AwLs had a negative impact on the students learning of writing, conflicting with this aspect of their professional identity. Similar conflict was seen in other teachers, but agency achievement there was limited, with little evidence of the teachers being able to envision a better or different future. Teacher H was able to do this, for example, they felt comparative judgement would improve writing assessment without levels. This was also identified in S2 SLT interview, comparative judgement of writing was being trialled in several years. In this case Teacher H's projective domain was supported by enablements in their practical-evaluative domain, where the option for comparative judgement of writing was available as a resource, and with the DHT supporting this as an option. Their objection to teaching in Yr6, as a result of their experience of AwLs in writing, however, seems enabled more by opportunity within school structures in the availability of such a position. Teacher H took advantage of an opportunity to progress in their career, in a role outside of the classroom, as a maths lead hence resolving the conflict they felt through remaining as a Yr6 classroom teacher. As such they may have alleviated some risk of being viewed as a non-compliant teacher and thus a threat to management, as identified by Wilkins (2015). Priestley et al. (2016) also reported agentic resistance in teachers to curriculum change which resulted in a loss of opportunity for promotion in response. This agency is not viewed independent of the dominate neoliberal discourses at play in policy and the school setting. Teacher H displayed beliefs which aligned with managerial and market views of education. They put their 'faith in the numbers', as Pratt (2016; 901) characterised it, when the numbers were based on their strong educational values.

#### 9.2.3.2 Strategic compliance

Agency was categorised as strategic compliance when teachers were able to display some principled subversions to assessment procedures or pressures, they were subject to. This

enabled them to mediate their compliance while additionally fulfilling their own educational goals.

#### *9.2.3.2.1.1 Teacher D*

Teacher D, was able to avoid succumbing to some of the pressure felt from the SLT, caused by measuring teacher performance using statutory testing, resulting in prioritisation of learning and assessment which reflects SATs format and content. For Yr6 teachers this significantly contributed towards TTTT, however for Teacher D this was mitigated by the inclusion of teacher SA in end of KS1 judgments.

*...but I think in year two because the SATs aren't everything...[you can] get around that [teacher performance measured of SATs] a bit. (FS1Y2T1D)*

They did not feel there was the same stressful impact for Yr2 students as those in Yr6, feeling instead that the children were not 'fazed by the tests, so it's not an issue for them... they're quite excited that they're getting their own little booklet and in colour as well.' Teacher D benefited from the KS1 SATs not being linked to school accountability directly in addition to being able to include teacher SA in their judgment. These affordances contributed to feeling in more control of their end of KS1 assessment. They suggest in the phrase 'get around that bit' that they are agentic in their assessment practices through their use of teacher SA to overcome managerial pressures of being held accountable using standardised testing to measure teacher performance. Green (2011: 45) refers to this as the 'new forms of control', which encourages teachers to 'enterprise themselves' through accountability being tied to student performance, which in turn links to Davies and Bansel's (2007) view that the neoliberal citizen is driven by their own individual success to enable success, for example, for the school itself.

#### *9.2.3.2.1.2 Teacher H*

Teacher H seemed able to achieve agency strategically by adopting performative strategies to meet the short-term expectations of SATs, while also prioritising their own long-term educational goals. Because KS2 SATs results were directly linked with school accountability, Teacher H felt that this made them a game.

*I think somewhere along the line, what assessment is for has been lost. I get that we should be made accountable...It's very important [accountability] but*

*it's the criteria that you place on it. As soon you put numbers on things and targets in that way, it becomes a game. (FS2Y6T2H).*

For assessments to be worthwhile, Teacher H believed it needed to impact student learning, this signified a strong internal influence from their professional identity but also displayed the internalisation of neoliberalism with a 'n' (Ong, 2007) where their actions as a teacher determined the successful outcome of their school in terms of SATs results. Compliance with assessment practices they criticised was only possible if Teacher H was able to simultaneously find a way to benefit student learning. They demonstrated this by ensuring that data driven and TTTT practices also fulfilled their educational values.

*...[we] use it [assessment data] for research purposes, as a diagnostic tool...it's not purely for just to stick numbers in a spreadsheet and say 'look at our lovely numbers.' that's [not] going to have much impact for the children whereas the other stuff we do, we put a lot of effort into, arguably does have impact on the children's learning. (FS2Y6T2H)*

Unlike Teacher E, who was unable to envision longer term educational goals beyond SATs pressure, Teacher H, as argued by Wilkins et al. (2012) was capable of mediating and subverting these managerial, marketised pressures from policy, by putting extra time and effort into analysing assessment data, specifically termly test data, to benefit students future learning.

*I go away and do some data analysis. I report back to the teachers with these are the five key weaknesses...It's forming part of a picture that's not about tracking progress, it's about what assessment, I believe, should be for which is: what do they know yet?...What depth did they know it? What don't they know yet? How can I get them to know it and understand it? (FS2Y6T2H).*

However, in fulfilling their own goal of ensuring student data collection also supported their educational values and beliefs, their workload became more demanding (Ball et al., 2011; Biesta et al., 2017). A conversation, referred to by Teacher H with a consultant, provided evidence for Teacher H that schools can have successful Ofsted inspections without traditionally tracking assessment. This may have provided motivation for Teacher H to use SA data formatively despite this adding to workload. In this sense, Teacher H's long-term

goals informing their projective domain, that testing can be used formatively and schools can still succeed in spite of what Teacher H referred to as 'A fear of accountability... I'm HMI (Her Majesties Inspectorate), you need to show me progress.' Teacher H acknowledged that they were in a privileged position, having some control of assessment practices as maths lead.

*I'm in a very privileged position because I'm in SLT here and I get to go around to a lot of schools...(FS2Y6T2H)*

This could provide Teacher H with experiences and information that build upon their professional identity and inform their projective domain as and how they view maths assessment. This experience also provided Teacher H with additional discourses regarding assessment through their observations of maths in other schools. In addition to this, Teacher H had achieved their Master's in Education, increasing availability to educational discourse and mediation of education policy outside of school practice.

*...the outcome of my study and it shocked me, and it did change things at my school and this is before the new curriculum...Everything was driven by the test, by the content of the test. (FS2Y6T2H)*

Teachers can unknowingly internalise and adopt the values of dominant discourses if they have limited availability to alternate policy discourse (Moore and Clarke, 2016). Studying at Master's level, and observing how other schools have approached AwLs provided a number of alternate discourses Teacher H could draw on to enable them to strategically fulfil their educational values and beliefs, while complying with internalised neoliberal discourse emphasising the importance of the self to drive success for the organisation, in this case the school. When it came to data driven monitoring of learning, tests were the priority, Teacher H's resistance to these practices were strategic in terms of how he used them to target future learning, but they were accepted, hence this is viewed as strategic compliance rather than agentic resistance.

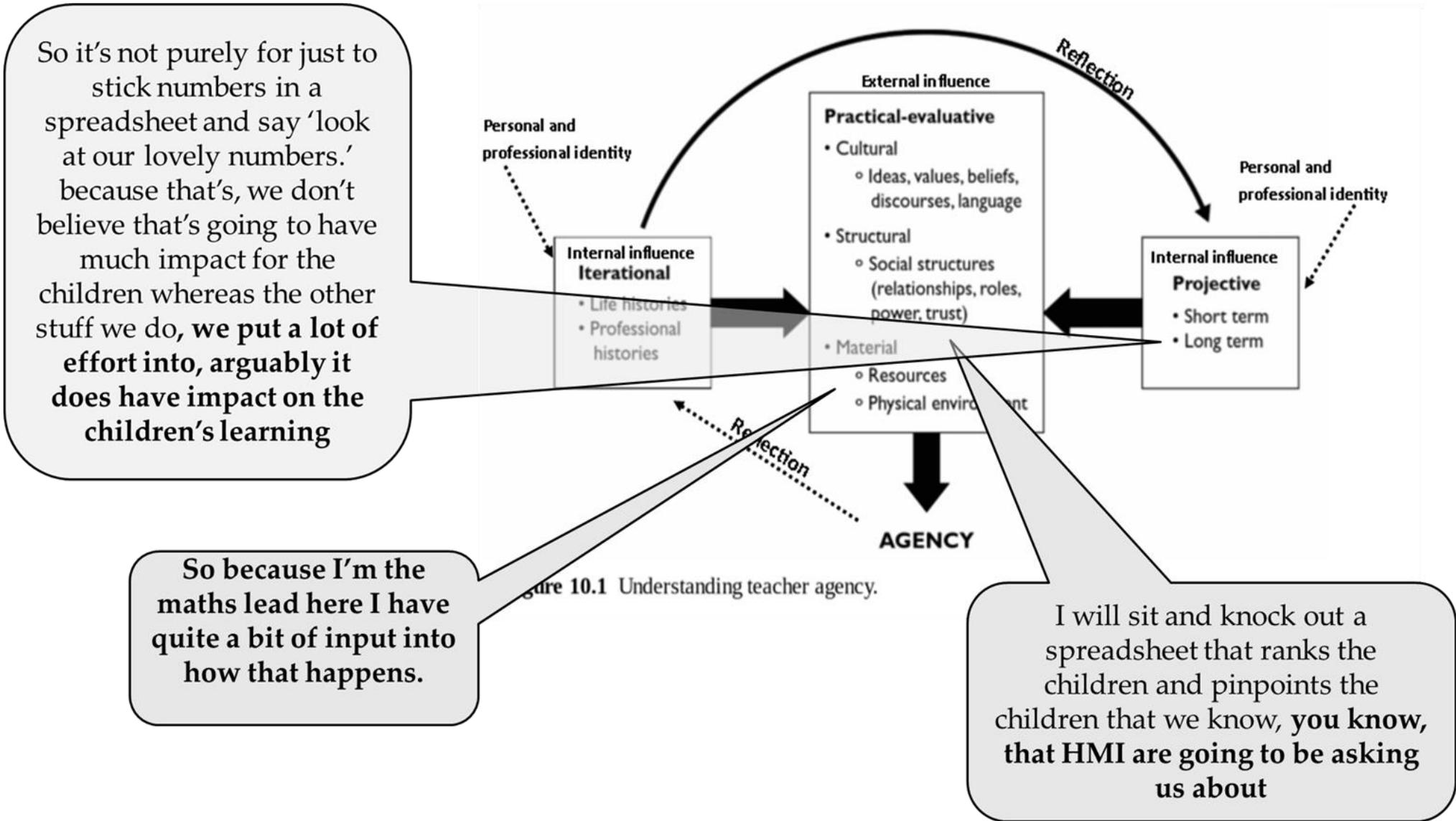


Figure 10.1 Understanding teacher agency.

Figure 9-5 Comments from Teacher D used to indicate how agency achievement was enabled and constrained using Priestley et al. (2013: 152) ecological theorisation of agency.

Unlike other teachers in this study shown in Figure 9-5 Teacher H was less constrained by their practical-evaluative domain, since they had a position to enact change, and had experiences which helped them to envisage change as possible. For example, they disagreed with the structure of some summative tests, including the SATs, and as maths lead they could choose the non-statutory summative assessments used, ensuring that they fulfil their vision of benefiting and accurately reflecting student learning. Their awareness of SATs as a part of the neoliberal marketised management for schools, rather than one of educational measurement, meant that they could view it as a game, working out the rules so that they could also devote time to fulfilling their own educational goals informed by their professional identity. The focus on success by numbers, meaning pupil data, indicated an awareness of a pressure that had an impact on their teaching practices. However, Teacher H coaching students by being transparent about areas of focus to be successful in the SATs, displayed an alternative view of how students can be successful, rather than narrowing the curriculum, instead informing students of the need to 'play the game'. Teacher H's position, with access to different educational discourses and determination for assessment to impact learning positively, avoided the situation in this case which concerned Stevenson (2010: 342), where neoliberal discourses of 'markets, targets and test' for the regulation of education would close off other possibilities.

There was a difference in agency between Teacher H's experiences of assessment of writing and their assessment in maths. It is possible that the external pressure of moderation was a constraint on Teacher H and their ability to adapt how they were teaching in a way that would benefit students. This demonstrates how a teacher can have varying achievements of agency due to environmental constraints. In this example relating to whether they had control over what was taught and how, or if it is imposed from an external source like moderation expectations. As assessment for maths in the SATs is mainly focused on the test, as maths lead for the school they may have more control of teacher SA leading to more agentic achievement than in subjects such as English where they did not have same enablements.

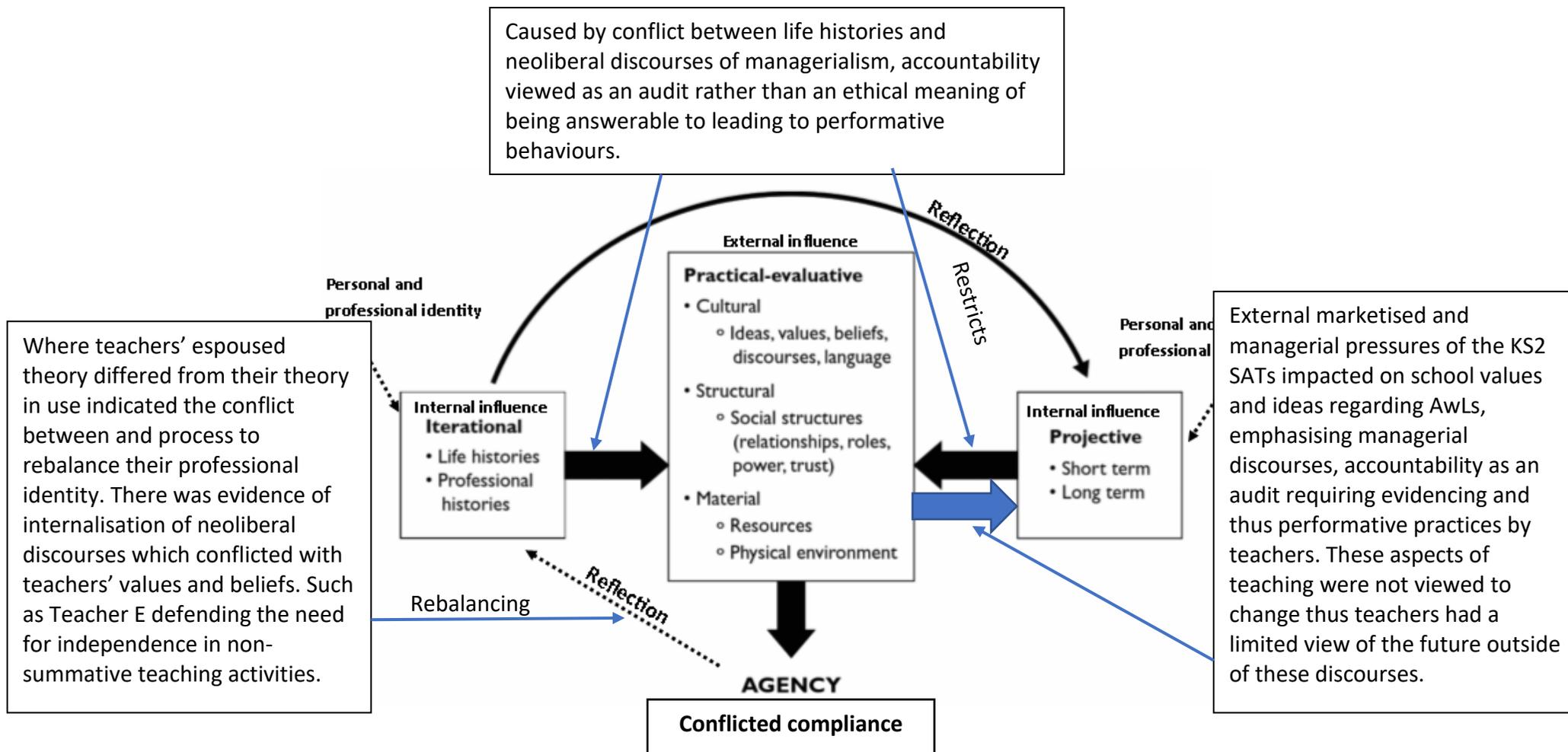
## 9.3 NEOLIBERAL DISCOURSES AND TEACHER AGENCY

### 9.3.1 Conflicted compliance

Discourses evident of managerialism, accountability as an audit of evidence and performativity, conflicted with the teachers' previous educational and professional identities. These discourses felt so dominant that teachers who demonstrated conflicted compliance did not envisage a future where they could change, therefore, their agency was restricted. These discourses had started to become internalised by some teachers who criticised their own practice in favour of dominant neoliberal values evidenced in the schools' required assessment practices.

External influences had an impact on the how the teachers could envisage a future outside of the dominant discourses discussed above. This adds an additional dimension to the annotated ecological theory from Priestley et al. (2013: 12), that the practical evaluative domain feeds into the projective domain in terms of external policy-based pressures. These external influences are additional to the impact of policy, with these discourses embedded in the monitoring and management of State Maintained schools.

Figure 9-6 Summary of Neoliberal impact of conflicted compliance evident in teacher interview data.

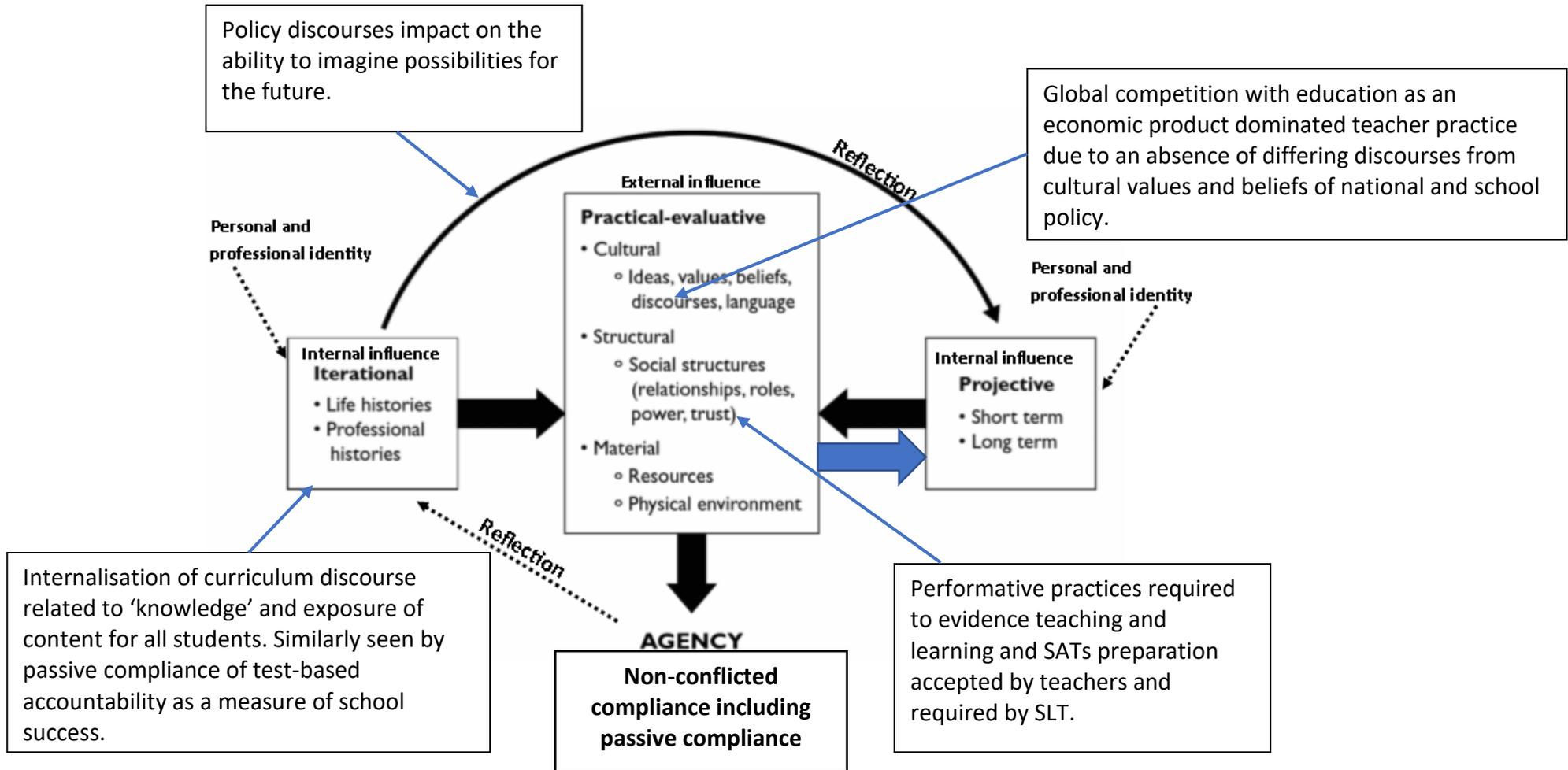


(annotated from Priestley et al., 2013: 152)

### 9.3.2 Non-conflicted and passive compliance

The lack of agency achieved in situations of passive compliance and non-conflicted compliance seemed to be a result of the internalisation of performative discourses, of the perception of education as a commodity, and the managerial structures within which teachers were required to operate. In addition to the iterational domain's impact on agency though carrying a teachers' values and beliefs, external influences of in-school policy impacted on teacher ability to envisage alternative futures. For example, Teacher G had no previous assessment experiences other than AwLs, and saw the main function of KS2 SATs as holding schools accountable for their teaching. The heavy reliance on ITAFs and any guidance supplied emphasised the lack of trust in teachers to assess, but also the internalisation of managerial modes of assessing which allowed for measurement and tracking (Moore and Clarke, 2016).

Figure 9-7 Summary of Neoliberal impact of non-conflicted and passive compliance evident in teacher interview data.

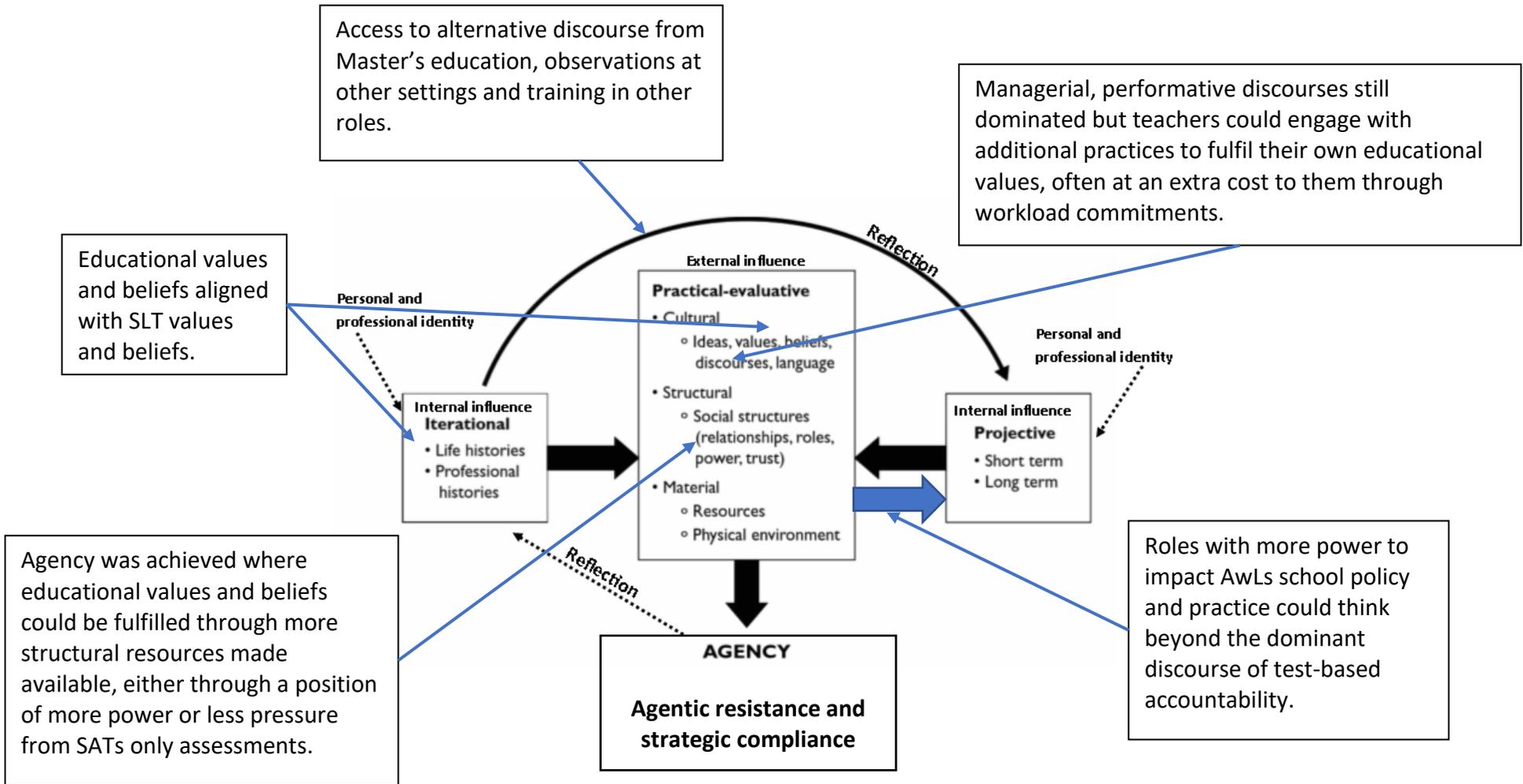


(annotated from Priestley et al., 2013: 152)

### 9.3.3 Agency

Achievements of agency through resistance and strategic compliance showed a complex interaction between the three domains of agency and the internalisation of neoliberal discourses in education. Where teacher values and beliefs aligned with the school's, agency was possible, but only where teachers had the resources, power and trust to shape their own teaching and assessment practices. Part of the agency achieved as strategic compliance with in-school assessment policy came from what Hoyle and Wallace (2007) referred to as 'principled infidelity'. Teachers D and G understood and had accepted (or internalised) the measurement of their own and the school's performance using statutory testing as part of high-stakes accountability for schools. The neoliberal emphasis accepted is one of '...a mode of 'governing through freedom' that requires people to be free and self managing' (Ong, 2007: 4). It is within this neoliberal lens that the schools and teachers operate, thus their acceptance and knowledge of how this management works, enabled strategic compliance by Teachers D and G. This would not have been possible without the enablements of their respective roles, and the absence of direct KS2 SATs pressures on their performance as teachers. Thus, teachers bearing the brunt of SATs based accountability were limited by the dominance of neoliberal discourse which narrows what can be envisaged as possible to the self-management of teacher performance measured through SATs testing. This in turn is influenced by similar pressures put on schools, governed by their freedom to self-manage, but only through the lens of SATs based accountability. The agency seen by Teacher H was represented by his change in role, he was agentic by searching out a non-class-based role where he could shape in-school assessment. Though this was motivated by the tensions caused between AwLs and his principled educational beliefs, he could not demonstrate this agency while still working under the dominance of test-based accountability, whereby his success in SATs contributed hugely to the school's success, through the value of SATs results in the education marketplace.

Figure 9-8 Summary of Neoliberal impact of agency evident in teacher interview data

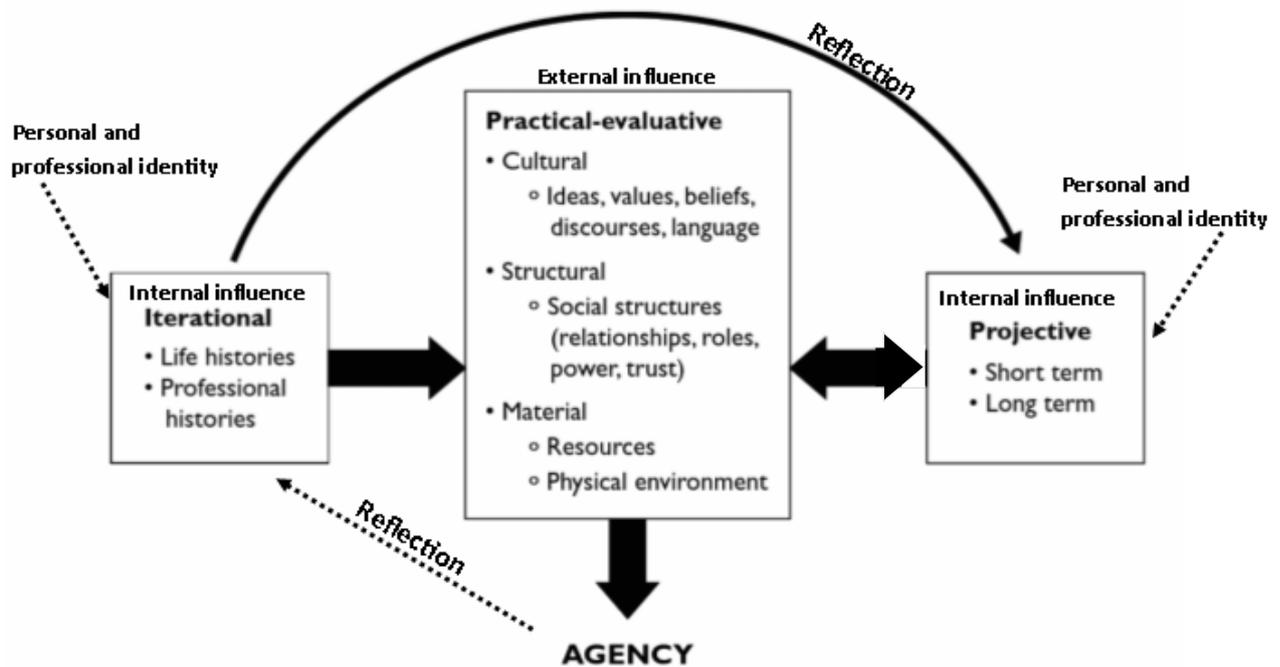


(Annotated from Priestley et al., 2013: 152)

### 9.3.4 Conclusion

Throughout this analysis of teacher agency, neoliberal discourses contributed to limitations in agency achievement wherein they dominated assessment discourse in schools, and impacted on teaching strategies. These discourses are being referred to as *dominant* as they prevented teachers from being able to envisage alternate discourses in either the short term or long term. What was particularly evident was that when agency was achieved for Teachers H and D, this was still limited to the neoliberal discourses at large. However, the resources available to these teachers, through subject or key stage lead positions, enabled them to be strategic in their compliance, and to be able to draw on additional assessment discourses available to them from external experiences, such as Teacher H’s Master’s degree and Teacher D’s training as a KS1 SATs moderator.

Figure 9-9 Adapted theory of ecological agency from neoliberal discourse impact.



(annotated from Priestley et al., 2013: 152)

This analysis of teacher agency through a neoliberal lens additionally adds to the ecological theory of agency by highlighting how impactful the practical-evaluative domain was on the projective domain of the teachers, as shown in Figure 9-9. Enablements and constraints within this domain, particularly concerning marketisation, performativity, and managerialism restricted agency. The ecological theory of agency accommodates this

influence where it is mediated through the iterational domain's life history component. However, what this research shows is that where compliance was conflicted, teachers maintained a highly impactful influence from strongly held educational values and beliefs. Where their projective domain was restricted, this was related to the resources available to them, and how compatible their values were with those that were dominant in national and school policy and practice. This indicates that the practical-evaluative domain impacts on the projective domain of the teachers, in a similar fashion as the projective domain does on the practical-evaluative domain.

## 10 FURTHER ANALYSIS AND DISCUSSION

---

Findings from this thesis draw attention to the function assessment has in Primary schools as a cog in the larger wheel of school accountability (Ball et al., 2012; Green, 2011; Pratt, 2016). This research exemplifies the impact of assessment policy change on a previously well understood, if flawed (Bew, 2011), assessment framework with NCLs. As such, findings are discussed in relation to literature presented in the first half of the thesis, and theory adopting a neoliberal lens to view assessment's place in the wider frame of policy and accountability, and considering the validity of primary assessment according to assessment theory and developments in learning theory set out in Chapter 2.

This chapter will also draw attention to the differences and similarities which emerged from the survey and interview data. It will thus be structured in headings relating to the main findings identified followed by a final section drawing together the impact of neoliberal discourses.

### 10.1 SATS PRESSURES AND THE IMPACT OF HIGH-STAKES TESTING WITHOUT LEVELS.

It was clear both interview settings felt pressure from a key lever of neoliberal ideology, market regulation of primary schools through statutory assessment, shown by efforts to maximise the outcome of the KS2 SATs. This was evident through SLT views, emphasising the need for SATs results to improve each year, and through teacher comments referring directly to SATs pressure felt by Yr6 teachers and students. Often referred to as assessment wash-back, consequences of this accountability pressure were evident in both settings (Koretz, 2008; Gipps, 2011). The increased floor standards in AwLs prompted efforts from both settings to attune their teaching to curriculum areas which could be measured in the KS2 SATs but each setting differed in their approach. In S2 the emphasis was on formative assessment practices to prevent students from falling behind and embedding a mind-set of learning which accepted failure as part of the process. Illeris (2017) conceptualised motivation as an essential dimension of learning; in accepting an emotional impact of failure, S2 showed sensitivity towards students' emotional dimension of learning, understanding that this can impact on student engagement in learning. Learning could be

more effective if students were motivated to work through mistakes, if those mistakes could be caught quickly, and acted on with 'responsive' teaching. This 'responsive' teaching in S2 was exemplified as an improvement in FA practices without levels, ensuring students were prepared for the next lesson. These practices echoed recommendations by the NCR (2011) who referenced the importance of mind-set in learning, and opportunities for students to catch up and learn as a class through small group teaching interventions. However, the SATs and ITAF content had a significant impact on what was targeted by this FA improvement. From then on, S2 engaged in a newly expanded market for in-school assessment resources transitioning towards 100 externally purchased tests for teacher SA which teacher S2 felt to be aligned better with end of KS2 SATs expectations. This is not to discount the authentic learning impact these strategies had, only that they were predominantly intertwined with end of KS2 ITAF and SATs expectations rather than the curriculum itself (Whetton, 2009); and it risks neglecting learning which cannot be represented in these assessments as well as invalidating interpretations that SATs and the narrow ITAFs represent learning across the whole curriculum.

SATs tests and ITAF wash-back was more evident in S1. To work more effectively and compete better in the education marketplace, the SLT focused on encouraging independence in lessons, seemingly to allow students to become accustomed to the absence of teacher support, and therefore be better prepared for SATs. This was discussed by all teachers in S1 who found this aspect of school policy alongside teaching a new and elevated curriculum content challenging. With SATs results referred to in the SLT interview as representing their schools' educational performance, the teachers were required to display their 'functional competency' (Wilkins, 2015) through them, with little trust put in teacher SA (discussed below). Adopting this audit style accountability '...performative regulatory instruments may have [had] a profound, and constraining, impact on teachers' sense of values-led professionalism.' (Wilkins, 2015: 1153). Teachers in S1 felt pressure to engage in TTTT strategies as a means of working harder (Koretz, 2017) since an effective teacher in S1 was one who could demonstrate performance through SATs results. This is evidence of the prevalence of neoliberal discourse on an individual level where Yr 6 teachers were required to self-govern and manage their own success through being held to account by the use of KS2 SATs results.

In S2, their pedagogy became more discursive in response to AwLs, in S1 teachers felt they needed to be less discursive in lessons. It is particularly concerning how deep the belief was, for S1SLT, that SATs results were a valid representation of learning, or, that this was even something which could be questioned when external pressures emphasised this so explicitly. Similar findings are seen by Ward and Quennerstedt (2019) of privileging test-measurable skills without levels through lesson content mirroring test structure and content. Universally, teachers in S1 acknowledged that these strategies did not fulfil in their own educational aims, for example, teaching a broad and balanced maths curriculum when number competency was represented proportionally more the SATs restricting the achievement of teacher agency.

S1 teachers stated that they felt non-core subjects suffered from limited time and they felt pressure to prioritise core subjects and SATs content. While teachers in S2 admitted they gave more time to core subjects with less attention given to non-core subjects. Survey findings supported this view, which was represented more among teachers in SATs years; that the balance between core and non-core assessment had not improved through AwLs.

Regardless of the setting, workload concerns were shared by the teachers. An increasing need to evidence teacher SA judgments showed the presence of managerial discourse (Green, 2011), which as well as encouraging performative behaviours of evidencing against ITAF criteria, also manifested itself in the need to transparently evidence day to day lesson activities. Although this was evident in S2 regarding evidence for moderated assessment, in S1 this resembled the dominance of teacher accountability as viewed through the lens of 'New Public Management', requiring audit trails such as evidencing teacher SA, as well as proof of learning in all lesson activities. Under these performative practices, Biesta (2017: 316) questions whether '...we are measuring what we value, or whether we are valuing what is being measured'.

## 10.2 TIMING AND EXTENT OF GOVERNMENT GUIDANCE ON ASSESSMENT FOR ALL YEARS.

Greater autonomy for schools to develop non statutory assessment frameworks was challenging for both settings, as they had reinvented their approaches several times since levels were removed and this was still seen as a work in progress. For the teachers, feelings

of instability and impending change in addition to curriculum reform made adjusting to AwLs difficult. This was echoed by the survey data which represented teachers of all year groups, showing the majority of participants found the transition difficult and felt more guidance was required. Neither setting had created, then stayed with, their own AwLs tests, trackers and frameworks. Instead, settings purchased assessment resources from private companies, taking advantage of an additional market opened, consisting of advisors for school assessment and resource producers. This was present before levels were removed but not as extensive (Ball, 2018), a similar trend noted by Poet et al. (2018) in Primary schools relying on external resources to develop their assessment policy and Alderton and Pratt (2021) for the use of curriculum tracking software. Of the multiple assessment trackers and core subject tests purchased for non-statutory assessment, SLT in both settings emphasised the importance of those which aligned with and prepared for the SATs. As noted by Pratt and Alderton (2019), resources which support SATs preparation are preferred. The tests specifically provided opportunities for both settings to compare attainment with other schools in non-SATs years, in a similar way SATs without levels now did for Yr6 and that QCA tests had done with levels.

Where guidance was available for end of KS1 and KS2 assessment, it was highly influential in both constructing non-statutory assessment, and also in shaping teaching (discussed in Validity and Teacher trust). As the only official assessment frameworks, they had been used to exemplified non-statutory assessment in non-SATs years. For example, SLT interviews from S1 showed they based their construction of assessment frameworks for all years on the ITAFs. Text responses from the survey also expressed the desire to have ITAF-like assessment frameworks for non-SATs years. Data from questions 7 and 9 showed that assessment guidance was useful and helped with confidence in AwLs; with those who had not assessed with levels more inclined to find this guidance useful. Similarly, teacher interviews showed how important the ITAFs, moderation experience, and the SATs of the previous year, were to their confidence in AwLs. The stripped back assessment guidance aligns with a neoliberal government agenda where central control is small (Fautley, 2017), and market choice encourages what is felt to be best practice (Whitty and Wisby, 2016).

Both survey data and Teacher G's interview data showed even for those who had not taught with levels experienced a difficult transition towards AwLs, suggesting that familiarity with

levels was not the only factor contributing towards a difficult experience. Although there were aspects of NCLs some teachers interviewed preferred, a desire to go back to NCLs was not the only contributing factor for experiencing a difficult transition. Interview data supported this, with reasons given for the difficulties of; late guidance, changes in guidance, speed of the changes, and reliance on level-like systems.

Reliance on resources aligned to SATs expectations and ITAF guidance shows unfortunately that autonomy granted to schools to design their own non-statutory assessment, and mitigate TTTT practices, had been used instead to select assessment resources which aligned to SATs criteria and ITAF expectations. As argued by Schmidt (2017), autonomy is undermined by high accountability especially when it is unclear how to succeed.

### 10.3 WRITING MODERATION AND RELIANCE ON ITAF

Teachers in both settings experienced writing moderation as an ITAF based tick list, and neither setting felt this encouraged good writing. In both setting teachers relied heavily on Governmental guidance, ITAF, and exemplars to embed the ITAF criteria throughout the school year. The use of the ITAFs as a secure fit assessment of writing was strongly disagreed with by teachers in both settings. Although, the clarity of expectation was a positive for S2 teachers as they had struggled to know what was expected in the first year under AwLs. The use of teacher SA for high-stakes accountability purposes has been cautioned against, as the pressure put on the teacher is highly likely to impact on teaching (Baird and Black, 2013; Black, 1998; Harlen 2005; Stobart, 2009; Wiliam, 2001). The findings from this thesis provide evidence of this wash-back from the use (or lack thereof) of teacher SA in KS2 SATs. Teacher SA was often critiqued for being biased ([see section below](#)), however, the issue here was with the moderation expectation of KS2 writing and how it narrowed the teaching of writing to end of KS2 expectations. The ITAFs, though criticised by the teachers, were also praised for making expectation clear for writing assessment, a tension was evident here between the teachers *needing* to know how to assess without levels and teaching according to their beliefs and values. This clarity enabled the evidencing and prioritising of the aspects of writing which were to be measured in moderation, thus demonstrating the impact of the neoliberal managerial discourse and the performative impact on the subsequent teaching of writing (Ball et al., 2012; Green, 2011). The pressure

to meet assessment targets, whether agreed with or not, ultimately led to compliance in the use of TTTT practices in writing, varying in extent for all the teachers. Reducing TTTT, as referred to was a key aim of removing levels and non-statutory assessment guidance in schools. Compared to APP criteria, which was viewed as more comprehensive, ITAF guidance for end of KS1 and KS2 is narrower. It is important to note that the ITAFs instructs that they are not to be used as an assessment framework, and do not represent a programme of study, however the lived moderation experiences of the teachers in this research contradicted this message. When viewed through a Neoliberal lens it is the teacher's duty, not for just themselves, but also for their school to succeed (Davies and Bansel 2007), to succeed with success measured by SATs results. Consequently, there is pressure, or even belief, that a successful teacher is one who *must* do so through the lens of assessment data (Pratt, 2016), as seen through the limited achievement of teacher agency in complying with test preparation practices participants disagreed with. Accountability demands placed on KS2 SATs were felt, and compliance with these pressures resulted in strategic practices to increase the likelihood of success in the SATs. Interview data analysis strongly suggested that teachers were unable to achieve much agency in avoiding this aspect of teaching. Good performances on SATs had become part of the teacher identity for some teachers (Buchanan, 2015; Moore and Clarke, 2016), who even started to doubt views informed by previous teaching experiences which encouraged them to put student learning first above SATs preparation.

A desire for a consistent measure of writing assessment was also a motivation for the reliance on the ITAFs. Interview data showed concerns of consistency in writing AwLs, based on experience of KS2 writing moderation, highlighting inconsistencies compared to other teachers they knew. As Moss (2017: 62) argued under the pressure of accountability 'the assessment tools themselves become the curriculum.'

A survey text response from Q7 referred to the assessment of writing as a farce without levels suggesting instead that assessment of writing should use comparative judgment (CJ). This was also suggested by S1 and S2, as a means of producing consistency of writing assessments. Notably, CJ was discussed in terms of replacing teacher SA at the end of KS2 rather than using it to supplement it as recommended by Heldsinger and Humphry (2010). Although viewed as preferable, CJ has drawbacks especially relating to workload and the

time it takes to judge longer pieces of writing, reflective of Yr6 expectations (Humphry, 2013; Wheadon et al., 2019). Teachers interviewed, who were in SATs years, did not feel AWLs ITAFs improved student writing, rather a higher focus on grammar hindered the production of creative, fluent writing.

#### 10.4 VALIDITY OF SATS

A prominent finding of this thesis is that participants did not feel that SATs or tests in general should be the only form of assessment. The word 'validity' was not used but instead teachers were highly critical of the notion that SATs assessment represented curriculum attainment fully and accurately for all children. As the survey data suggested, which teacher interviewed supported, a hybrid format of SA combining teacher and test assessment was preferable. For KS1 this was the format, with the KS1 teacher interviewed stating that this removed pressure and stress from the KS1 SATs as the test was not relied upon alone; though this did not prevent wash-back from the KS1 ITAFs in this case it did allow assessment of areas Yr6 teachers felt were more valid if teacher assessed for example, comprehension of reading, and using and applying in maths. Teachers in S1 showed more concern with the format of SATs tests disadvantaging their students than in S2. The reading SATs were referred to specifically as under-representing the reading curriculum just as Tennent (2020) observed, SATs reading assessments are affected by construct under-representation, as they are heavily skewed towards comprehension. The SATs reading test was also felt to be affected by construct-irrelevant content, such as writing ability and wider life experiences (Messick 1989a). This was not raised as a concern for S2 which was situated in a more affluent area, but teachers shared concerns regarding the impact stress and anxiety had on students test results. These concerns around how statutory tests affects students' performance were downplayed in SLT interviews, where a 'chance' of a few students not being able to cope well in SATs conditions was discussed; in general, it was felt tests could assess learning reliably, though perhaps not always validly. Reliability of SA was a higher priority for the SLT whereas, validity was a main concern for teachers interviewed.

This research is calling into question validity of SATs results without levels, more so than with NCLs, based on the enhancement of TTTT and test preparation practices evident in the interviews. Such practices were referred to as present with NCLs but there was an emphasis

that test preparation mattered more without levels as there was no other measure to gauge their success against as a school. When considering the consequences of test-based accountability from Koretz (2008) of working more effectively, teaching more, working harder, reallocation, coaching and cheating. S1 and S2 described all of these consequences apart from cheating. Test preparation strategies to familiarise their students with the format and structure of testing, were embedded throughout the school year. Students needed to be aware of the 'tricks' to pass the SATs tests.

A bigger influence of TTTT was evident in S1, where pedagogy was shaped to encourage the level of independence required in tests in lesson activities. Emphasising independence during lessons and minimising teaching fails to align with the developments made in learning theory. Limiting interaction with students, risks losing the student focused social constructivist approach to learning and the formative assessment benefits it yields. As Illeris (2017) posits students need to interact with the teacher, ineffective communication with the teacher inhibits learning. S2 had a different approach where the focus was on more engagement with the teacher or teacher assistant, with more attention given to the mind-set of learning to enable students to learn without a fear of failure. These practices by S2 provided opportunity for all three of Illeris' (2017) dimensions of learning: student motivation and concentration (incentive), previous knowledge (content), and effective teacher communication (environment) (Illeris 2017: 13). Although both settings reference Alison Peacock's work on Growth Mind-set (Swann et al., 2012) this seemed more embedded in pedagogy in S2.

Reallocation was evident with all teachers in relation to their adaptation to AwLs, by targeting areas of the curriculum weighted more heavily in the SATs tests or ITAF for writing and sacrificing non-core subjects to provide extra support for maths, reading and writing. Within core subjects, reallocation of teaching was evident by the focus on ITAF content for writing, comprehension for reading and a focus on number for maths. The major issue with this reallocation, which could also be considered as exemplifying TTTT, was that it invalidates interpretations of SATs outcomes as representative of a wider base of learning, which the test merely samples, as Wiliam (2001: 21) states '...people in the darkness can get up to all kinds of things' Teacher F's experience reflects this when recounting a specific reallocation of teaching time to specific maths objectives, resulting in the lowest ability

group outperforming their own middle group, despite learning less of the curriculum content.

Though some argue that schools and teachers do not have to teach to the test (Bew, 2011; Goldstein, 2017), findings of this research demonstrate that for at least these schools the reality was that to be successful in SATs TTTT, to a greater or lesser extent, has become part of teaching in Yr2 and 6 as a result of the heightened use of test-based accountability in the absence of comprehensive teacher assessment guidance. Removing levels seems to have done little to alleviate TTTT in these schools, as Green (2011: 45) argued, autonomy in the presence of ‘contracts, targets, performance indicators, and monitoring and evaluation systems act as new forms of control.’ Reallocation in S1 impacted more on pedagogy, where lesson content was adapted to better suit test independence. The concern here is one raised by Gipps (2011: 5) that ‘only some material and certain tasks are amenable to...testing’ and thus may be privileged above other learning in the curriculum which cannot be represented in testing.

### 10.5 TEACHER ASSESSMENT: TRUST AND RELIABILITY

Given the emphasis in assessment literature (*inter alia* Broadfoot and Black, 2004; Harlen, 2005; Stobart, 2009; Ward and Quennerstedt, 2019; Wiliam 2001) and government commissioned reports of assessment (Bew, 2011; McIntosh, 2015), that teacher SA needed a more prominent role in KS2 assessment, it is concerning that this thesis findings show examples of two schools both choosing to move further away from teacher SA. This caused considerable tensions with teachers in S1 and to a lesser extend S2. While feeling it was a more accurate representation of student learning, teachers from both settings simultaneously questioned the reliability of teacher SA, something also noted by Poet et al. (2018). Survey respondents concurred that teacher SA was needed for some areas of the curriculum with less support that testing was best suited to some areas of the curriculum.

This data regarding teacher SA exemplifies the impact of managerialism as governance within schools, as Green (2011) argued, not only is the priority given to what can be measured, to measure this performance evidencing was essential, instilling a top-down audit culture (Ball et al., 2012). This was represented more so in S1, where pressure was felt

to provide evidence for all learning and not just for teacher SA. The HT and DHT in S1 related this requirement for teacher SA to be evidenced to monitoring student progress. The effect this had on the teachers in S1 was a feeling of distrust from the SLT. Similarly, Teacher H in S2 felt that their teacher SA was not trusted, accepting that teacher SA had not been valued by the Government even before NCLs were removed. Both settings used online trackers to monitor teacher assessment which further demonstrates the necessity to rely on managerial modes formative assessment aligned to the curriculum resulting in summatively driven formative assessment. The impact of school dependence on trackable assessment platforms as Alderton and Pratt (2021: 12) explain it is that 'the data produces the learner and the teacher as much as they both produce the data.' The emphasis is on a narrow view of learning, restricted to what can be evidenced and tracked rather than trusting the teacher to make a professional judgment based on their experience.

AWLs had resulted in both settings increasing their use of tests for non-statutory SA. SLT in both settings felt tests were more reliable as a predictor of how students would perform in the SATs. Teacher SA was felt to be biased, inconstant, and unreliable, criticisms attributed in the literature to the dominance of testing for SA (Allal, 2013; Collins et al., 2010; Marlow et al., 2014). In S1 teachers felt pressure for their SA to match the SATs outcome, pressure which can be accounted for through the responsibility placed on the individual to succeed through their own efforts within the neoliberal discourse. The SLT in S1 reinforced this, with the HT and DHT indicting disparities between teacher SA and SATs tests were due to too much teacher support in lessons. Formative diagnostic assessments were also moving towards independent mini-tests or cold tasks instead of group and talk based activities, assessments which can be planned, tracked and managed more explicitly. A similar trend of primary schools moving towards testing for TA is evident in recent literature investigating the impact of the assessment changes in primary schools (Ward and Quennerstedt, 2019); with Pratt and Alderton (2019) linking this reliance on tests with attempts to gain control over assessment targets schools otherwise would be unsure how to meet.

The use of external testing in both settings reflected findings from Poet et al. (2018), that they allowed for comparability with other schools, and provided a function to be able to predict how students were progressing towards end of KS2 standards. However, what this demonstrated was a continuation of the influence from end of KS2 test based accountability

as wash-back throughout the primary school, the very thing McIntosh (2015) presented in their review of AwLs. McIntosh (2015) argued this type of wash-back was the fault of NCLs being used outside its original bounds of end of KS1 and KS2 summative assessment. The language of levels may have been what was used to enact these assessment driven practices, but it seems test-based accountability of schools within dominant neoliberal discourses of marketisation, competition and managerialism, is what instigates the practices which continue without levels. As Pratt and Alderton (2017) posited, assessment practices may have changed, but they operate within the same discourses. The resulting impact of these discourses is the reforming of the teacher as a professional. Dominant discourses of test-based accountability and performativity are defining what it means to be a teacher and are able to restrict the agency of teachers. As the interview data showed, only teachers with some element of seniority were able to display agency in how they complied with test preparation and coaching practices; though there were clear tensions where teachers complied. Rather than the teachers contradicting themselves, when complying although conflicted, it is possible that this was instead what Buchannan (2015) identified as an internalisation of test-based accountability and how it is used as a measure of a successful teacher.

## 10.6 NEOLIBERAL IMPLICATIONS OF AWL POLICY REFORM

The analysis of education policy and research data using the framework of dominant neoliberal discourses, set out in [Chapter 4](#), has been central to understanding the impact of AwLs on teaching practice. This will be explored below summarising implications of the neoliberal market agenda for primary assessment.

### 10.6.1 Marketisation of education, education as a commodity

Marketisation in education involves:

- Competition for resources via league tables
- Managerialism encouraging performativity
- Accountability- school success- test based

Therefore, when viewed in this fashion AwLs exists within the regulation of schools using market theory. This is a market led philosophy where school autonomy and punitive accountability measures will drive up education standards through the threat of forced

academisation or a reduction in pupil numbers (see Competition). Embedded in the AwLs reform is the assumption that education functions as a means to economic prosperity in a globalised marketplace. The viewpoint of education as a moral imperative has been lost in policy and discourse (James, 2015). Moreover, using the economy as a justification for marketised regulation of the school system puts pressure on school output, exam results, and not on the process of learning, which was where the participants' values and beliefs lay. This encouraged an objective link between teaching input and exam output, encouraging teaching input which would better fulfil the testing output, so the school had more in the way of capital in the education market. Though this has been an increasing feature in education policy since New Labour reforms in the late '90s, the measure of success is now very narrow as a result of the stripped back non-statutory assessment guidance alongside the prevalence of KS2 SATs for holding schools to account.

#### 10.6.1.1 Competition

School league tables function with the market to inform parental 'choice' of school for their child. Pupil places are attached directly to school finance, therefore, according to this market theory better SATs performing schools attract fuller class sizes thus optimising school funding for that year. However, this assumes that parents live in an area where they have 'choices' of schools with available places in order to be able to select a preference. Additionally, this theory assumes that SATs results are a valid measure of school and teaching quality, a stance not held by participants in this research. According to this practice, schools who perform lower on the league tables with have low pupil sizes thus less money. With a reduced school budget potentially meaning less finance for staff and educational resources how can a school be best place to improve their competitive position in the market? In a neoliberal market there will always be inequalities, such as schools who will never attract enough pupil places when schools are always ranked (Exley and Ball, 2014). In this research pressure was very clearly felt by Yr 6 teachers in both settings to maximise SATs results.

Global competition in the economic job market is presented in policy as the imperative behind educational improvements, and motivates the use of international comparisons with countries who out-perform England in large scale international assessments. This perpetuates the view of education as intellectual capital, and as capital in a global market

but ignores that market theory just might not be improving standards educationally, rather that it is undermining the very improvements policy makers want (Hoyle and Wallace, 2007). For example, while S2 shifted focus towards improving the pupil learning process formatively, S1 encouraged teaching practices which reduced teacher intervention, thus potentially missing FA opportunities for improving learning.

### 10.6.2 Accountability and autonomy within the neoliberal message

This research does not argue that schools should not be held to account, rather that they can be accountable, but in a way that does not undermine their purpose to educate. They should also only be accountable for what they can control. The measure of accountability through the neoliberal lens emphasises and prioritises what can be assessed using SATs. This narrows down the options available to schools to exercise autonomy over in-school assessment and SATs preparation practices, as seen in the shift towards externally purchased tests to represent teacher SA in both S1 and S2. Teacher autonomy over test preparation practices was further restricted when teacher pay was linked to performance which was also linked to pupil assessment performance.

### 10.6.3 Managerialism

The emphasis on managing SATs outcomes dominated teacher practice, and encouraged the use of audit style performance monitoring of teachers. Emphasising what is visible explicitly prioritises this type of learning and assessment. It set teachers up to have to prove themselves, creating a severe lack of trust between SLT and teachers in S1, and also resulted in performative teaching and assessment practices in both interview settings.

#### 10.6.3.1 Performativity

Managerial pressures of proving teaching and learning have occurred increased workload and thus took time away from fulfilling teachers' education values and goals when they were not managerial in nature. To manage these tensions, teachers acted performatively to ensure teaching and learning were evidenced and made visible in the manner required by SLT or government. From this perspective teacher motivation was to ensure it looked like what needs to be done has been done, even though the task itself may not be viewed as worthwhile educationally by the teachers, as seen in [section 9.2.3.2](#) concerning Strategic compliance. As a result, there was conflict for teachers between engaging with

performativity practices, and spending time and energy on teaching which was aligned with their educational values and beliefs.

The next chapter concludes the thesis by directly addressing how the data and findings address the research questions and provide new knowledge, before making recommendations for future assessment policy reform and practise.

# 11 CONCLUSION

---

## 11.1 INTRODUCTION

At the beginning of this thesis a set of research questions were introduced:

- What has been the impact of the assessment reform 'Assessment without Levels' on teaching and assessment practice?
- What are the Policy into Practice implications of the assessment reform 'Assessment without Levels'?

How this research has addressed these questions will be dealt with by reviewing each question and in turn, drawing on research findings and discussion. These conclusions apportion no blame or criticism of schools or teachers who participated, all those interviewed displayed a deep belief in education and were passionate about their roles. Where discussion is made which is critical of approaches adopted by the schools the important part of the discussion is 'why', this is not to try to find 'what works', but to explore pressures, enablements, and constraints on schools and teachers experiencing assessment reform.

## 11.2 WHAT HAS BEEN THE IMPACT OF THE ASSESSMENT REFORM 'ASSESSMENT WITHOUT LEVELS' ON TEACHING AND ASSESSMENT PRACTICE?

There were two main challenges teachers experienced with AwLs; transition, and adapting to a new assessment framework (or lack thereof). Teachers felt the transition was not handled well, with guidance arriving late for those in SATs years, and absent for those not in SATs years. Schools had been placed in a difficult position where previous guidance and support for assessment had been removed, which was previously relied on to anticipate cohort achievement in SATs (Moss 2017). The result of this for both settings was to initially revert to a levels-based assessment before moving on to several externally purchased assessment resources. Teachers lacked both consensus and stability to enable them to confidently assess without levels. The situation was still precarious even after four years of AwLs transition, the schools were still adapting non-statutory SA practices. The instability and uncertainty contributed towards a lack of agentic behaviour from teachers interviewed.

It was difficult for them to envision assessment differently when teachers were still adapting to the most recent incarnation of in-school assessment and unchanging demands of statutory testing. Critique from the participants regarding the handling of AwLs Policy transition was damning. Government policy failed to provide timely guidance and support. End of key stage expectations still needed to be met but how pupils were to meet them was significantly less clear than with NCLs.

The desire to have a more structured assessment framework for non-statutory assessment was apparent. In SATs years the ITAFs quickly became the focus of assessment in Yr2 and 6. This could have been avoided if a comprehensive assessment framework for the end of KS1 and KS2 had been developed alongside training for teachers to implement it (See [Recommendations](#) below). The desire to have a structured assessment framework did not come from an affinity for NCLs but was desired as a means of security - teachers and SLT wanted to be able to track and predict SA so students were prepared for SATs. This meant relying on externally purchased test resources for all years aligned to the format and cohort-referencing of KS2 SATs without levels, in addition to modelling non-statutory teacher SA on ITAFs structure and content. Teachers interviewed displayed little agency concerning TTTT, even with those who were subject or Key Stage leads complied with this practice. Cases where agency was seen were from teachers with leadership responsibility, however this agency was in terms of fulfilling their own educational aims which were hampered by their compliance to test preparation practices, such as curriculum narrowing and reallocation of time to tested core subjects. Where Teacher H was able to do this, they added significantly to their workload, as Ball et al. (2011: 366) state, teachers ‘...are creative and sophisticated and they manage, but they are also tired and overloaded much of the time’.

Neither schools nor teachers demonstrated much autonomy in constructing AwLs using school assessment because of the dominating pressure to perform well in SATs. Setting 2 displayed an encouraging attitude to learning, increasing the amount of small group catch-up sessions, however these were limited to core subjects of Maths and English, perpetuating the divide between core and non-core subjects. Setting 1’s approach demonstrated a more drastic alignment of learning to the conditions of statutory testing. The impact differed between settings studied but improvements in learning, for both settings, centred on improving statutory SATs results, risking the production of inauthentic

teaching practices to meet targets (Wilkins 2015). These practices were documented from both study settings through coaching, reallocation of time to core tested subjects and targeting areas of the curriculum which were represented more on statutory assessments.

The validity of testing was questioned extensively by those interviewed and worsened by a move from both schools to test based SA in school. Teacher assessment, though argued by the teachers as being a better judge of aspects of the curriculum compared to tests, echoed in survey findings, was not trusted, with expectations in S1 that teacher SA should function as a predictor of how students performed in the SATs. Combined with a requirement to evidence their SA, the requirement for teacher SA was declining.

Achievement of agency was additionally hampered by an acceptance and normalisation of testing as an accurate measure of school and teaching quality (Moore and Clarke, 2016). Conflict in these views was evident between teachers, especially between tightly held educational beliefs practices which compromised these. Yet even where these conflicts did appear, how success was measured as a Yr2 or Yr6 teacher was through SATs results, what it meant to identify as a successful teacher included conforming with test-based accountability practices (Buchanan, 2015). In conforming to the accountably systems some teaching practices were valued over others, making educational priorities for schools unclear (Biesta et al., 2015). This aspect of assessment was not reduced under AwLs and is known to hamper student learning yet policy argued that in-school NCLs assessment was the reason for these practices.

Historically the introduction of the standards for teachers has been criticised for de-professionalising teachers by defining measurable instrumentalist quantities of their performance (Buchanan, 2015; Priestley, 2011b). In de-professionalising teachers through previous reforms, teachers have been trained to evidence the standards when qualifying to be a teacher. In doing so teachers are trained with the view that possessing these functional competencies (Wilkins, 2015) defines a professional teacher. The impact of this could be seen in the agency analysis of teachers interviewed, who normalised test-based accountability as part of teaching. Assessment policy without levels amplifies the need for teachers to performatively prove themselves as teachers through ambiguity of assessment expectations, particularly in relation to Ofsted inspections where students' books are the source of evidence for learning common to all schools.

### 11.3 WHAT ARE THE POLICY INTO PRACTICE IMPLICATIONS OF THE ASSESSMENT REFORM

#### ‘ASSESSMENT WITHOUT LEVELS’?

Although the motivation for the Assessment Reform was supported by literature and SLT interviewed, the reality for teachers showed that little of those improvements had been realised. Consequences of NCLs in terms of TTTT, end of key stage assessment criteria influencing all primary years, and levels driving learning up to the next criteria at the expense of deepening knowledge (McIntosh, 2015), were still evident when assessing without levels, and, in some cases, were heightened. In perpetuating the discourses on marketisation and test-based accountability, any alternative discourses are afforded little space (Stevenson, 2010). This lack of alternative discourse hampers attempts made by Policy makers to increase the autonomy of schools and subsequently allow teachers to improve their teaching practices. The ability of teachers to envision different futures is constrained by limited access to educational discourses which complement current assessment priorities.

What Policy has ignored is that assessment discourses present in NCLs were not confined to assessment using levels. Levels, as a criteria were a delivery tool, and a language used to enact assessment under a wider neoliberal education discourse. In AwLs, these discourses still function, but through a different mechanism and language (Pratt and Alderton, 2019). Pragmatically, teacher views demonstrated that assessment had become more constrained by performative practices in AwLs. Through additional expectations to evidence teacher SA, writing with AwLs had been impacted hugely through moderation experiences of a technical judgment of competency. In S1, evidencing extended a performative gaze from the SLT to all lesson work.

This research demonstrates that improvements in education beyond test preparation are unrealistic, with the continued presence of Neo-liberalism with a ‘N’ (Ong, 2007), marketisation and education viewed as an economic commodity, and neo-liberal impact with ‘n’ (Ong, 2007) encouraging performative, responsiblised teachers. Teachers referred to their pay as also related to student attainment, adding financial stakes to their performance. This directly linked students’ exam scores to teachers’ financial situation. If educational success continues to be measured by external standardised testing, the only learning that will still be valued is that which can be tested. Autonomy has not been

afforded to school AwLs, when school regulation measures privilege those who can maximise test output (Priestley et al., 2015b, Wilkins 2015). This compromises the validity of these assessments in demonstrating attainment of the curriculum it intends to sample, as it is in reality only representing a small part of the curriculum.

For autonomy to be achieved teachers need to be able to achieve agency in their classrooms. Allowing schools autonomy for in school assessment while ratcheting accountability measures denies the development of this (Priestley et al., 2015b). The institutionalised use of standardised testing to regulate schools has robbed teachers of trust in their assessment by both other people and themselves. Literature and DfE commissioned reports overwhelmingly favour the inclusion of teacher assessment to increase the validity of educational assessments, and to mitigate the need to teach to the test or rely on coaching strategies to enable students to perform adequately. AwLs has not enabled these changes, instead it has increased pressure on teachers, schools, and students, resulting in an increase of TTTT practices, further invalidating the ability of the SATs to represent the learning they purport to. Instead, there is likely to be a repeat of the 'improvement then plateau' (Koretz, 2008) seen following the National Strategies under the Labour government, where teachers adapt to the new demands, and find creative ways to increase class performance, whilst putting the extra work into teaching they feel to be of value but which is currently hampered by the very assessments which were intended to make it better.

#### **11.4 CONTRIBUTION TO KNOWLEDGE**

There are two areas where this research has contributed new knowledge to the field of educational assessment and teaching practice. The first is adding to, and moving forward discussions in the literature addressing primary assessment practices, and the impact of policy reform on teaching and assessment in the primary school. The second area to which I have contributed new knowledge is in the adaptation and application of the ecological theory of agency developed by Biesta and Tedder (2006). Each of these two distinct contributions will be discussed in turn in this section.

#### 11.4.1 Adding to and moving forward debates around assessment practices and policy in primary education.

The impact of assessment without levels lacks prior research due to the recent nature of the changes, thus my research findings provide new knowledge on how participating schools adapted to this most recent policy change. These findings add to the most recent research into AwLs (Poet et al., 2018; Pratt and Alderton, 2019; Ward and Quennerstedt, 2019) that in the absence of structured assessment guidance, and continued presence of test-based accountability, in-school assessment has narrowed learning more to becoming mainly that which is testable. By viewing education policy through a neoliberal lens, this research has identified the impact of opening further a market for externally purchased school assessment resources, which, due to the pressure on SATs, resulted in in-school assessment relying on those which most closely mimicked SATs testing. An in-depth picture of the implementation of AwLs in the two settings provides a novel view into the tensions teachers face in SATs years, and the importance of school values and ethos evidenced by the resulting differences in pedagogy between settings. This emphasised the impact of school mediation of policy on the AwLs reform.

There is a major concern from this research that SATs without levels are a non-valid form of summative assessment if their intention is to assess student attainment on a curriculum that goes into the depth of a subject as argued in DfE (2010b). The aim for higher autonomy with in-school assessment to improve educational attainment, had been prevented through the continued presence of a pressurised and punitive education market. This next section addresses my second contribution of knowledge of my contribution to the ecological theory of agency.

The ecological theory of agency had been previously applied by others (refs) to curriculum reform, applying it to AwLs reform was a novel approach taken in this research, and adapted further by developing the framework as discussed below. This produced novel insights into teaching practices, and how agentic teachers were when facing pressure to TTTT and narrow the curriculum as identified in the thematic interview data analysis. The fate of teacher SA in any form other than a test is presented as bleak, with teacher views on their own assessment conflicted. This picture of SA after AwLs policy reform moves forward the discussion on the place for teacher SA in statutory assessment, or even if there is one under

the current climate where we value what we can measure, rather than working to develop a way to better measure what we value.

#### 11.4.2 Agency theory and the Neoliberal impact

The ecological theory of agency had been previously applied by others (Priestley et al., 2013) to curriculum reform, applying it to AwLs reform was a novel approach taken in this research, and adapted further by developing the framework as discussed below. This produced novel insights into teaching practices, and how agentic teachers were when facing pressure to TTTT and narrow the curriculum as identified in the thematic interview data analysis. The fate of teacher SA in any form other than a test is presented as bleak, with teacher views on their own assessment conflicted. This picture of SA after AwLs policy reform moves forward the discussion on the place for teacher SA in statutory assessment, or even if there is one under the current climate where we value what we can measure, rather than working to develop a way to better measure what we value.

In Chapter 5 the ecological theory of agency from Biesta and Tedder (2006) and applied to Scottish curriculum policy reform by Priestley et al. (2013), is presented as a theoretical framework relevant for analysing teacher agency during assessment policy reform. Teacher identity was considered and argued as integrated with agency (Beauchamp and Thomas, 2009; Buchanan, 2015; Day et al., 2006). Thus, the ecological theory of agency was built on by linking in theory of teacher identity and agency into the model shown in Figure 5-1 in Chapter 5. Essentially, links were made between what is referred to as *external* and *internal* influences on teacher identity, how identity can be in conflict with expectations in practice, and how resolving these conflicts provide opportunity for agency. Figure 5-2 shows Figure 5.1 annotated to include these elements, which allowed for analysis of teacher interview data, in terms of teacher agency, to focus on teachers' differing roles and levels of experience in the face of assessment policy reform. The link between agency outcome and a teachers iterational domain was indicated as reflection, making clearer the role experience plays in the reforming of teacher identity, and the impact this has on future practice thus future agency achievement.

The annotated model of ecological agency was then applied to teacher interview data, analysing teacher agency in response to AwLs policy reform. This topic is both novel and

under researched, and within what literature there is (as discussed above), lacks a focus on teacher agency in assessment practices in England. In applying this theory, I have contributed new knowledge to this field and provided vital insights into how national policy, when mediated through neoliberal discourses, restricts teacher agency in their efforts to mitigate the unintended consequences of statutory testing.

Through the analysis of teacher agency, ecological theory has further been adapted to account for the impact of neoliberal discourse on the practical evaluative domain. Figure 9-9 shows a key adaptation which emphasises the impact the practical-evaluative domain had on the teachers' projective domain. It was not just limited through experience and temporal interactions, neoliberal discourses were so dominant their impact restricted short-term and long-term considerations of the future based on what was available to them in terms of resources, positions of power and trust. The dominance of neoliberal discourses through AwLs reform via the removal of in-school assessment guidance, and a diminishing role of teacher SA, demonstrated strong impact of the practical evaluative domain on the projective domain. This was in relative rather than reflective terms, which was not previously presented in the ecological theory of agency.

## 11.5 RESEARCH DESIGN REFLECTIONS

Here I will review the study design, methods and theoretical underpinnings of the research. This research is limited in its scope from the small sample size of two school settings and survey responses. The responses from the participants cannot be interpreted as being representative of teachers at large. However, in adopting an interpretivist paradigm I prioritised the context rich nature qualitative interviews, alongside survey data which although smaller in response than hoped, did represent a balance of all primary year groups. In addition, the research aimed for credibility and trustworthiness above generalisability of findings (Guba and Lincoln, 1986). The research was limited by the resources available, as a PhD project I was the soul researcher and had a tight timescale to recruit primary schools for interviews and participants for the survey. In hindsight and given more time I would have attempted to recruit additional schools increasing the strength of triangulation between many sources of data. I am disappointed that more people did not complete the online survey from the social media promotion. Due to the nature of this recruitment process, I will never know for sure why so many potential participants did not progress past the first few pages. Perhaps teachers did not have time for online surveys. From my own experience it would not have been high on my priorities. Knowing this now I would have a different approach of asking schools if I could have 5 min of a staff meeting to discuss AwLs and provide them with the option of filling out the survey on a paper format. This was the approach utilised for the interview study settings.

With this considered I still feel the mixed method pragmatic research design was appropriate and worked well for the research focus. There are areas of the research which on reflection would have been useful to cover for example, including writing assessment in the survey from the start, including questions relating to assessment theory literacy and exploring teachers understanding of the interaction between learning and assessment.

The three main theoretical lenses adopted to view the topic through; assessment validity, neoliberal ideology and agency, all emerged out of the pilot study interview data analysis. Had I been cognisant of these theories at the start of the research as I am now, they may have influenced how I approached the interview questions and the survey. However, there

may have been a benefit from me still identifying more as a teacher than a researcher by not being far removed from the job participants occupied.

## 11.6 RECOMMENDATIONS

Considering the impact of neoliberal discourse and the conclusion reached, as outlined above, recommendations will now be made for further research and for schools, teachers and policy makers.

### 11.6.1 Further research

Since data collection a number of changes have already occurred. Student writing assessment is no longer a secure-fit assessment and instead a best-fit, in response to teacher concerns. KS1 ITAFs have stayed the same as teacher SA is still moderated as part of SATs. However, in KS2 there is only an assessment framework for writing, removing any guidance for teacher SA in reading and maths.

The covid-19 pandemic has shifted the landscape of primary assessment which seemed to be moving further away from increasing the inclusion of teacher SA. Teacher assessment has had to be utilised to assess students at the end of KS2 ahead of their transition to secondary education. Research into the development of AwLs in primary schools in light of a recent reliance on teacher SA is crucial. This is to address the following points this research raises going forward:

- How has the profile of teacher assessment changed as a result of its use in the covid-19 pandemic?
- In what ways has AwLs developed since this research took place?
- What position are teachers in to be agentic in their assessment practices without levels?

### 11.6.2 Schools, teachers, and policy makers

As with many others who have written on the topic, as well as a recommendation since TGAT (1988), I recommend that teacher SA should form a larger part of end of KS2 assessment (Bew, 2010; Stobart, 2005; Tennent, 2020). This would require policy makers to reform current statutory end of KS2 assessment to become a blend of teacher SA and test. The form this takes, based on my research could see areas such as reading comprehension

still assessed via a test but other areas of the reading curriculum based on teacher SA. The same can be applied to maths where teachers supported the use of tests for calculation but preferred teacher SA for areas such as using and applying. Decisions, by academics, researchers and teachers in primary assessment, would need to be made on what the most valid assessment method is for different areas of maths, reading and writing. Currently, the test dominated system in place favours reliability, and the comparability it provides, over how valid the assessments are. If this recommendation is adopted, then teacher assessment frameworks would need to be developed to be comprehensive alongside CPD for teachers. Moderation will also need to improve. This is a challenge when in-school assessment guidance is extremely limited preventing the development of a shared knowledge and understanding of how the curriculum is represented in student's work in different schools. There lacks a foundation for moderators to base their judgments on.

Years of SA dominated by statutory testing may have deskilled teachers in their own assessments, therefore teachers need training and trust to develop rigorous and dependable but most importantly valid assessments (James 2017; Marlow et al., 2014; McIntosh 2019; William, 2015). Data from this doctoral research has shown that teachers' literacy regarding assessment theory was minimal. Considerations of teacher bias and consistency of assessment between schools were made but teachers did not discuss assessment in terms of assessment theory for example, considering the validity and reliability of assessment. Therefore, developing primary teacher assessment literacy, may provide a deeper understanding of assessment and its relationship to learning. To this end, I recommend that assessment theory form a larger part of Initial Teacher Education so that early career teachers go into the profession with a more developed perspective on assessment and how it links to learning theory. For those already in the profession, CPD should be offered to develop knowledge of assessment theory and provide an additional discourse for the use and purpose, but also the realistic bounds of assessment.

I acknowledge that without these changes at policy level schools and teachers taking on board recommendations to invest in their teachers and trust their dedication to their career and reap the potential benefits of students who receive an education where learning, rather than assessment, comes first, is not straightforward as KS2 SATs are embedded in a marketised, accountability laden regulation of education and believed to represent the very

learning they prevent because of their dominance. The current performative driven view of professionalism and what it means to be a good teacher displays the cruel optimism described by Moore and Clarke (2016), because justifications for AwLs provided teachers and SLT in this study with hope, hope to improve upon the down sides of NCLs, TTTT and pushing on through levels rather than deepening understanding. Yet, by engaging in and adopting AwLs policy, these schools have replicated or reinforced the very practices they hoped to remedy. It is well beyond the scope of this research to recommend how Primary school accountability can be reformed as it is as much of a victim of Neo-Liberal market agendas as other areas of our life. However, it would be negligent of me not to provide support and data from this research to the cause. How AwLs was adopted in each setting did matter. SLT in S2 put authentic pedagogical practices in place, though these privileged core SATs subjects, student learning was at their heart. Therefore, how schools and SLT mediate policy is critical to the impact of AwLs reform on student learning. SATs pressure cannot be ignored, yet some teachers displayed agency in ensuring their compliance to AwLs policy impacted less on their students' learning than if the teacher only engaged on a performative level. However, this comes at a cost for the teacher (Wilkins et al, 2012). I recommend that school SLT put student learning and teacher wellbeing at the heart of their practice. Workload demands under AwLs policy had increased according to participants leaving less time to teach pupils to learn rather and instead worked to ensure students performed well in SATs. To this end I encourage SLT to research elements of their practice in a critical manner which may provide wider discourses than those available in practice and policy.

My last recommendation is to policy makers, apart from abandoning policy dominated by neoliberal ideology, they should expand their view of how education can be successful. The current neoliberal motivated pressure on statutory assessment is not working to authentically increase attainment. Perhaps the minister for education should also engage in education research and invest their time in studying assessment theory and become more literate from engagement in the wider literature, expanding their discourses regarding education and assessment beyond which they are currently confining themselves.

## REFERENCES

---

- Allal, L. (2013) teachers' professional judgement in assessment: a cognitive act and a socially situated practice, *Assessment in Education: Principles, Policy & Practice*, 20(1), pp. 20-34
- Alderton, J. and Pratt, N. (2021) Filling gaps: assessment software and the production of mathematics and its teaching and learning in primary schools. *Critical Studies in Education*, pp.1-15.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1985) *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- American Educational Research Association, American Psychological Association, & National Council, on Measurement in Education. (1999) *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- American Educational Research Association, American Psychological Association, & National Council, on Measurement in Education. (2014) *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Angus, L. (2015) school choice: Neoliberal education policy and imagined futures. *British Journal of Sociology of Education*, 36(3), pp. 395-413.
- Archer, M.S. (1995) *Realist social theory: the morphogenetic approach*, Cambridge: Cambridge University Press.
- Argyris, C. and Conant, J. B. (1985) *Theories of Action*. In Argyris, C. Putman, R. and McLain Smith D. (1985) *Action Science*, 13, pp. 80-102. Wiley.
- Assessment Reform Group (1999) *Assessment for learning: beyond the black box*. Cambridge: University school of Education.
- Assessment Reform Group (2002a) *Testing, motivation and learning*. Cambridge: University school of Education.
- Assessment Reform Group (2002b) *Assessment for learning: 10 principles*. Cambridge: University school of Education.
- Assessment Reform Group (2006) *The role of teachers in the assessment of learning*. London: University Institute of Education.
- Bailey, G. (2014) Accountability and the rise of "play safe" pedagogical practices, *Education + Training*, 56(7), pp. 663-674,
- Bailey: L., & 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 Bochell,

H. and Powell, M. (eds) *The Coalition Government and Social Policy: Restructuring the Welfare State*, pp. 127-152. Bristol: Policy Press.

- Baird, J. (2010) Beliefs and practices in teacher assessment. Editorial. *Assessment in Education: Principles, Policy & Practice*, 17(1), pp. 1-5.
- Baird, J.A., Andrich, D., Hopfenbeck, T.N. and Stobart, G. (2017) Assessment and learning: Fields apart? *Assessment in Education: Principles, Policy & Practice*, 24(3), pp. 317-350.
- Baird, J. A. and Black, P. (2013) Test theories, educational priorities and reliability of public examinations in England. *Research Papers in Education*, 28(1), pp. 5-21.
- Ball, S. J. (1998) Big policies/small world: An introduction to international perspectives in education policy. *Comparative education*, 34(2), pp. 119-130.
- Ball, S. J. (2003) The teacher's soul and the terrors of performativity. *Journal of education policy*, 18(2), pp. 215-228.
- Ball, S. J. (2012) *Global Education Inc.: New policy networks and the neoliberal imaginary*. Routledge.
- Ball, S.J. (2015) 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.
- Ball, S. J. (2015) Education, governance and the tyranny of numbers. *Journal of Education Policy*, 30(3), pp. 299-301
- Ball, S. J. (2016) Subjectivity as a site of struggle: refusing neoliberalism?. *British Journal of Sociology of Education*, 37(8), pp. 1129-1146.
- Ball, S. J. (2018) Commercialising education: profiting from reform! *Journal of Education Policy*, 33(5), pp. 587-589.
- 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.
- Ball, S.J., Maguire, M., Braun, A. and Hoskins, K. (2011) Policy actors: Doing policy work in schools. *Discourse: Studies in the cultural politics of education*, 32(4), pp. 625-639.
- Ball, S., Maguire, M., Braun, A., Perryman, J. and Hoskins, K. (2012) Assessment technologies in schools: 'Deliverology' and the 'play of dominations'. *Research Papers in Education*, 27(5), pp. 513-533.
- Baltodano, M. (2012) Neoliberalism and the demise of public education: The corporatization of schools of education. *International Journal of Qualitative Studies in Education*, 25(4), pp. 487-507.
- Beauchamp, C. and Thomas, L. (2009) Understanding teacher identity: An overview of issues in the literature and implications for teacher education. *Cambridge journal of education*, 39(2), pp. 175-189.

- Benton, T. and Elliott, G. (2016) The reliability of setting grade boundaries using comparative judgement, *Research Papers in Education*, 31(3), pp. 352-376,
- Bennett, R. E. (2009). *A critical look at the meaning and basis of formative assessment*. Princeton, NJ: Educational Testing Service.
- Bennet, R. E. (2011) Formative assessment: a critical review, *Assessment in Education: Principles, Policy & Practice*, 18(1), pp. 5-25,
- Bew, P. (2011) *Review of Key Stage 2 testing, assessment and accountability: Progress Report*. London: Department for Education.
- Biesta, G. (2008) Good education in an age of measurement: on the need to reconnect with the question of purpose in education, *Educ Asse Eval Acc*, 21, pp. 33-46
- Biesta, G. (2017) Education, Measurement and the Professions: Reclaiming a space for democratic professionalism in education. *Educational Philosophy and Theory*, 49(4), pp. 315-330.
- Biesta, G., Priestley, M. and Robinson, S. (2015) The role of beliefs in teacher agency. *Teachers and Teaching*, 21(6), pp. 624-640.
- Biesta, G., Priestley, M. and Robinson, S. (2017) Talking about education: Exploring the significance of teachers' talk for teacher agency. *Journal of Curriculum Studies*, 49(1), pp. 38-54.
- Biesta, G. J. and Burbules, N. C. (2003) *Pragmatism and educational research*. Rowman & Littlefield.
- Biesta, G. and Tedder, M. (2006) *How is agency possible? Towards an ecological understanding of agency-as-achievement*. The Learning lives Project.
- Biesta, G. and Tedder, M. (2007) Agency and learning in the lifecourse: Towards an ecological perspective. *Studies in the Education of Adults*, 39(2), pp. 132-149.
- Black, P. (1998) Learning, League Tables and National Assessment: opportunity lost or hope deferred? *Oxford Review of Education*, 24(1), pp.57-68.
- Black, P. (2000) Research and the Development of Educational Assessment, *Oxford Review of Education*, 26(3-4), pp. 407-419.
- Black, P. (2015) Formative assessment – an optimistic but incomplete vision, *Assessment in Education: Principles, Policy & Practice*, 22(1), pp.161-177,
- Black, P., Harrison, C., Hodgen, J., Marshall, B. and Serret, N. (2010) Validity in teachers' summative assessments. *Assessment in Education: Principles, Policy & Practice*, 17(2), pp. 215-232.
- Black, P., Harrison, C., Hodgen, J., Marshall, B. and Serret, N. (2011) Can teachers' summative assessments produce dependable results and also enhance classroom learning? *Assessment in Education: Principles, Policy & Practice*, 18(4), pp.451-469.
- Black, P., McCormick, R., James, M. and Pedder, D. (2006) Learning How to Learn and Assessment for Learning: a theoretical inquiry, *Research Papers in Education*, 21(02), pp. 119-132.

- Black, P. and Wiliam, D. (1998a) Assessment and classroom learning. *Assessment in Education: principles, policy & practice*, 5(1), pp. 7-74.
- Black, P. and Wiliam, D. (1998b) *Inside the black box: Raising standards through classroom assessment*. London: GL Assessment.
- Black, P. & Wiliam, D. (2003) 'In praise of educational research': formative assessment, *British Educational Research Journal*, 29(5), pp. 623-637
- Black, P. and Wiliam, D. (2006) The reliability of assessments. *Assessment and learning*, pp. 119-131.
- Black, P. and Wiliam, D. (2007) Large-scale assessment systems: Design principles drawn from international comparisons. *Measurement*, 5(1), pp. 1-53.
- Black, P. & Wiliam, D. (2009) Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), pp. 5-31.
- Black, P. and Wiliam, D. (2018) Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, 25(6), pp. 551-575.
- Black, P., Swann, J and Wiliam, D. (2006) school pupils' beliefs about learning, *Research Papers in Education*, 21(02), pp. 151-170,
- Bloom, B. S. (1956) Taxonomy of educational objectives. *Cognitive domain*. 1. New York: McKay.
- Bloom, B. S., Hastings, J.T. and Madaus, G.F. (1971) *Handbook on Formative and Summative Evaluation of Student Learning*, New York: McGraw-Hill Book Co.
- Borsboom, D., Cramer, A. O., Kievit, R. A., Scholten, A. Z. and Franić, S. (2009) *The end of construct validity*. In *The Concept of Validity: Revisions, New Directions and Applications*, Oct, 2008. Information Age Publishing.
- Boud, D. (2000) Sustainable assessment: rethinking assessment for the learning society. *Studies in continuing education*, 22(2), pp. 151-167.
- Boyne, R. (2000) "Post-Panopticism", *Economy and Society*, 29(2), pp: 285-307.
- Bramley, T. and Vitello, S. (2019) The effect of adaptivity on the reliability coefficient in adaptive comparative judgement. *Assessment in Education: Principles, Policy & Practice*, 26(1), pp. 43-58.
- Braun, V. and Clarke, V. (2006) Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2), pp. 77-101.
- Braun, V. and Clarke, V. (2014) What can "thematic analysis" offer health and wellbeing researchers? *International journal of qualitative studies on health and well-being*, 9.
- Braun, V. and Clarke, V. (2021) Can I use TA? Should I use TA? Should I not use TA? Comparing reflexive thematic analysis and other pattern-based qualitative analytic approaches. *Counselling and Psychotherapy Research*, 21(1), pp. 37-47.

- Broadfoot, P and Black, P. (2004) Redefining assessment? The first ten years of Assessment in Education, *Assessment in Education*, 11(1).
- Broadfoot, P. M., Daugherty, R., Gardner, J., Gipps, C. V., Harlen, W. and James, M. (1999) *Assessment for learning: Beyond the black box*. Cambridge, UK: University of Cambridge school of Education.
- Broadfoot, P. M., Daugherty, R., Gardner, J., Harlen, W., James, M. and Stobart, G. (2002) *Assessment for learning: 10 principles*. Cambridge, UK: University of Cambridge school of Education.
- Bronfenbrenner, U. (1977) Toward an experimental ecology of human development, *American Psychologist*, 32, pp. 513– 31.
- Bronfenbrenner, U. (1979) *The Ecology of Human Development*, Cambridge, MA: Harvard University Press.
- Bruner, J. S. (2009) *Actual minds, possible worlds*. Harvard University Press.
- Bryant, A, & Charmaz, K. (eds) (2019) *The SAGE Handbook of Current Developments in Grounded Theory*, London: SAGE Publications.
- Buchanan, R. (2015) teacher identity and agency in an era of accountability. *Teachers and Teaching*, 21(6), pp. 700-719.
- Buchanan, K. and Bardi, A. (2015) The roles of values, behavior, and value-behavior fit in the relation of agency and communion to well-being. *Journal of personality*, 83(3), pp. 320-333.
- Burgess, S., Briggs, A., McConnell, B. and Slater, H. (2006) *school choice in England: Background facts*. Bristol: The Centre for Market and Public Organisation.
- Burgess, S., Greaves, E., Vignoles, A. and Wilson, D. (2009) *Parental Choice of Primary school in England: What "type" of school Do Parents Choose?* Bristol: The Centre for Market and Public Organisation
- Burgess, S., Greaves, E. and Vignoles, A. (2019) school choice in England: evidence from national administrative data. *Oxford Review of Education*, 45(5), pp. 690-710.
- Campbell, D.T. (1976) Assessing the impact of planned social change. *Occasional Paper Series*, 8.
- Charmaz, K. (2020) "With constructivist grounded theory you can't hide": Social justice research and critical inquiry in the public sphere. *Qualitative Inquiry*, 26(2), 165-176.
- Charmaz, K. (2019) With Constructivist Grounded Theory You Can't Hide, *Social Justice Research and Critical Inquiry in the Public Sphere*, Qualitative Inquiry.
- Cizek, G. J. (2016) Validating test score meaning and defending test score use: Different aims, different methods. *Assessment in Education: Principles, Policy & Practice*, 23(2), pp. 212-225.

- Christodoulou, D. (2017) *Making good progress?: The future of Assessment for Learning*. Oxford University Press-Children.
- Cohen, L., Manion, L. and Morrison, K. (2007) *Research Methods in Education*. 6<sup>th</sup> edn. Abingdon: Routledge.
- Cohen, L., Manion, L. and Morrison, K. (2018) *Research Methods in Education*. 8<sup>th</sup> edn. Abingdon: Oxon.
- Coles, A. and McGrath, J. (2010) *Your Education Research Project Handbook*. London: Pearson.
- Collins, S., Reiss, M & Stobart, G. (2010) What happens when high-stakes testing stops? teachers' perceptions of the impact of compulsory national testing in science of 11-year-olds in England and its abolition in Wales, *Assessment in Education: Principles, Policy & Practice*, 17(3), pp. 273-286,
- Conservative Party (2007) *Raising the Bar, Closing the Gap: A policy green paper*. London: Conservative Party
- Corbin, J., & Strauss, A. (2008). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. 3rd edn. Sage Publications.
- Courtney, S. J. (2016) Post-panopticism and school inspection in England. *British Journal of Sociology of Education*, 37(4), pp.623-642.
- Cox, H. (2008) The development process for the optional teacher assessment materials for English for use in Wales at the end of Key Stage 2, *Education 3-13*, 36(4), pp. 313-323
- Creswell, J. W. (2009) *Research design: Qualitative and mixed methods approaches*. London: Sage Publications.
- Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.
- Creswell, J. W. and Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory into practice*, 39(3), pp. 124-130.
- Cronbach, L. J., & Meehl, P. E. (1955) Construct validity in psychological tests. *Psychological bulletin*, 52(4), p.281.
- Cronbach, L.J. (1971) Validity. *Educational measurement*, 2, pp. 443-597.
- van Daal, T., Lesterhuis, M., Coertjens, L., Donche, V. and De Maeyer, S. (2016) Validity of comparative judgement to assess academic writing: examining implications of its holistic character and building on a shared consensus. *Assessment in Education: Principles, Policy & Practice*, pp. 1-16.
- van Daal, T., Lesterhuis, M., Coertjens, L., Donche, V. and De Maeyer, S. (2019) Validity of comparative judgement to assess academic writing: Examining implications of its holistic

character and building on a shared consensus. *Assessment in Education: Principles, Policy & Practice*, 26(1), pp. 59-74.

Daniels, H. (2016) *Vygotsky and pedagogy*. Routledge.

Daugherty, R. (2007) Mediating academic research: the Assessment Reform Group experience. *Research papers in education*, 22(2), pp. 139-153.

Davies, B. and Bansel, P. (2007). Neoliberalism and education. *International journal of qualitative studies in education*, 20(3), pp. 247-259.

Day, C., Kington, A., Stobart, G. and Sammons, P. (2006) The personal and professional selves of teachers: Stable and unstable identities. *British educational research journal*, 32(4), pp. 601-616.

Dearing, R. (1994) *The Dearing Review: The National Curriculum and its Assessment: Final Report*.

Denscombe, M. (2007) *The good research guide: for small-scale social research projects*. 3<sup>rd</sup> edn. Maidenhead: Open University Press.

Denscombe, M. (2009) *Ground rules for social research: Guidelines for good practice*. UK: McGraw-Hill Education.

Denscombe, M. (2010). *Ground rules for social research: Guidelines for good practice*. 2<sup>nd</sup> edn. UK: McGraw-Hill Education.

Denscombe, M. (2014). *The Good Research Guide: For Small-scale Social Research Projects*. 5<sup>th</sup> edn. UK: McGraw-Hill Education.

Denzin, N. K. (1978) *The research act: A theoretical introduction to sociological methods* (2nd ed.). New York: McGraw-Hill.

Denzin, N. K. and Lincoln, Y. S. (1998) *Collecting and interpreting qualitative materials*. London: Sage.

Denzin, N.K. and Lincoln, Y.S. (2005) Introduction: The discipline and practice of qualitative research. In N.K. Denzin and Y.S. Lincoln, eds. *The Sage handbook of qualitative research* (3rd edn., pp.1-32). Thousand Oaks, CA: Sage.

DES (1988a) *Task Group on Assessment and Testing: a report* (London, Department of Education and Science and the Welsh Office).

Department for Children, schools, Families. (2009) *Getting to Grips with Assessing Pupils' Progress*. London: Department for Children, schools and Families.

Department for Children, schools and Families. (2009) *Report of the Expert Group on Assessment*: London: Department for Children, schools and Families.

Department for Education. (2013) *The National Curriculum in England Key Stages 1 and 2 Framework Document*. London: Crown Copyright.

- Department of Education and Science. (1988b) *Task Group of Assessment and Testing Report. A Digest for schools*. London: HMSO.
- Department for Education and Employment. (1998) *National literacy strategy framework*. London: Department for Education and Employment.
- Department for Education (2010a). *The case for change*. London: Her Majesty's Stationery Office.
- Department for Education. (2010b) *The importance of teaching: the schools white paper 2010*. London: Her Majesty's Stationery Office.
- Department for Education. (2011a) *The National Strategies 1997–2011. A Brief Summary of the Impact and Effectiveness of the National Strategies*. London: Department for Education.
- Department for Education. (2013) *The national curriculum in England Key stages 1 and 2 framework document*. London: Crown Copyright
- Department for Education. (2014) *National Curriculum in England: Framework for Key Stages 1 to 4*. London: Crown Copyright.
- Department for Education. (2016) *How Progress 8 and Attainment 8 Measures are Calculated*. London: Department for Education.
- Department for Education. (2016) *Primary school accountability in 2016 A technical guide for primary maintained schools, academies and free schools*. London: Crown Copyright.
- Department for Education. (2017) *Analysis of teacher supply, retention and mobility*. . Available at: <https://www.gov.uk/government/statistics/teachers-analysis-compendium-2017> [Accessed 20 June 2020]
- Department for Education. (2017) *Primary Assessment in England: Government Consultation Response*. London: Department for Education
- Department for Education. (2018) *Analysis of teacher supply, retention and mobility*. Available at: <https://www.gov.uk/government/statistics/teachers-analysis-compendium-4> [Accessed 20 June 2020]
- Department for Education and Employment. (1997) *Excellence in schools*. London: Department for Education and Employment.
- Department for Education and Employment. (1997) *Implementation of the National Literacy Strategy*. London: Department for Education and Employment.
- Department for Education and Skills (2001) *schools: achieving success*. Department for Education and Skills publications.
- Department for Education and Skills (2005) *Higher standards, better schools for all: More choice for parents and pupils*. London: Department for Education and Skills publications.
- Department for Education and Skills (2006) *Primary framework for literacy and mathematics*. London: Department for Education and Skills.

- Diamond, J. B. (2007) Where the rubber meets the road: Rethinking the connection between high-stakes testing policy and classroom instruction. *Sociology of Education*, 80(4), pp. 285-313.
- Dweck, C. (2015) Carol Dweck revisits the growth mindset. *Education Week*, 35(5), pp. 20-24.
- Dweck, C. (2017) *Mindset-updated edition: Changing the way you think to fulfil your potential*. UK: Hachette
- Education Policy Institute (2019) *Analysis: Could phased bursaries solve the teacher retention crisis?* Available at: <https://epi.org.uk/publications-and-research/phased-bursaries-teacher-retention/> [Accessed 20 June 2020]
- Emirbayer, M. and Mische, A. (1998) What is agency? *American journal of sociology*, 103(4), pp. 962-1023.
- Engeström, Y. (1987) *Learning by Expanding: an Activity Theoretical Approach to Developmental Research*. Orienta-Konsultit, Helsinki.
- Exley, S., & Ball, S. J. (2014) Neo-liberalism and English education. *Neo-liberal educational reforms: A critical analysis*, pp. 13-31.
- Fautley, M. (2017). Policy and the Question of Assessment. In Schmidt, P. K. and Colwell, R. (2017) *Policy and the political life of music education*. Oxford University Press.
- Foucault, M. (1980) *Power/knowledge: Selected interviews and other writings, 1972-1977*.
- Furlong, J. (2013) Globalisation, neoliberalism, and the reform of teacher education in England. In *The educational forum* (77( 1), pp. 28-50). Taylor & Francis Group.
- Gardner, J. (2010) *Developing teacher assessment*, McGraw Hill/Open University Press: Maidenhead.
- Gardner, J. ed.. (2011). *Assessment and Learning*. SAGE: London
- Gipps, C. (2011) *Beyond Testing (Classic Edition): Towards a Theory of Educational Assessment*, Taylor & Francis Group
- Gipps, C. and Stobart, G. (2009) *Fairness in assessment*. In *Educational assessment in the 21st century* (pp. 105-118). Springer: Dordrecht.
- Glaser, B.G. (1963) Retreading research materials: the use of secondary analysis by the independent researcher, *The American Behavioural Scientist*, 6 (10), pp. 11–14.
- Glaser B.G., Strauss, A. (1967) *Discovery of Grounded Theory. Strategies for Qualitative Research*. Sociology Press.
- Glass, G.V. (1976) Primary, secondary and meta-analysis of research, *Educational Researcher*, 5 (10), pp. 3–8.

- von Glasersfeld, E. (1984) *An introduction to radical constructivism*. In Watzlawick, P. ed. (1984) *The invented reality*. New York: Norton.
- von Glasersfeld, E., (1987) Learning as a constructive activity. Problems of representation in the teaching and learning of mathematics, pp.3-17.
- von Glasersfeld, E. (1997). Homage to Jean Piaget (1896–1982). *The Irish Journal of Psychology*, 18(3), 293-306.
- von Glasersfeld, E. (1998) Cognition, construction of knowledge, and teaching. In *Constructivism in science education* (pp. 11-30). Springer, Dordrecht.
- von Glasersfeld, E. (2001) The radical constructivist view of science. *Foundations of science*, 6(1-3), pp. 31-43.
- Goldstein, H. (2017) A response to ‘assessment and learning: fields apart?’ *Assessment in Education: Principles, Policy & Practice*, 24(3), pp. 388-393.
- Gorard, S. (2012) The Increasing Availability of Official Datasets: Methods, Limitations and Opportunities for Studies of Education, *British Journal of Educational Studies*, 60(1), pp.77-92.
- Gorard, S.(2002) The Role of Secondary Data in Combining Methodological Approaches, *Educational Review*, 54(3), pp. 231-237,
- Gorard, S. and Taylor, C. (2004) *Combining methods in educational and social research*. McGraw-Hill Education (UK).
- Green, J. (2009) The deformation of professional formation: Managerial targets and the undermining of professional judgement. *Ethics and Social Welfare*, 3(2), pp. 115-130.
- Green, J. (2011) *Education, professionalism, and the quest for accountability: Hitting the target but missing the point*. Routledge.
- Greener, I. and Powell, M. (2009) The evolution of choice policies in UK housing, education and health policy. *Journal of Social Policy*, 38(1), pp. 63-81.
- Grenfell, M., & James, D. (1998) *Bourdieu and education: Acts of practical theory*. New York
- Guba, E. G. and Lincoln, Y. S. (1989) *Fourth generation evaluation*. Sage.
- Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. *Handbook of qualitative research*. pp.105-117.
- Guba, E. G. and Lincoln, Y. S. (2005) Paradigmatic Controversies, Contradictions, and Emerging Confluences. In N. K. Denzin & Y. S. Lincoln, eds, *The Sage handbook of qualitative research* (pp. 191–215). Sage Publications Ltd.
- Hadar, L.L. and Benish-Weisman, M. (2019) teachers’ agency: Do their values make a difference?. *British Educational Research Journal*, 45(1), pp. 137-160.

- Hargreaves, A. (2007) Sustainable professional learning communities. *Professional learning communities: Divergence, depth and dilemmas*, pp.181-195.
- Harlen, W. (2004a) *A Systematic Review of the Evidence of the Impact on Students, teachers and the Curriculum of the Process of Using Assessment by teachers for Summative Purposes*. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.
- Harlen, W. (2004b) *A Systematic Review of the Evidence of the Reliability and Validity of Assessment by teachers for Summative Purposes*. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.
- Harlen, W. (2005) 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.
- Harlen, W. (2005) teachers' summative practices and assessment for learning—tensions and synergies. *Curriculum Journal*, 16(2), pp. 207-223.
- Harlen, W. and Deakin Crick, R. (2003) Testing and motivation for learning. *Assessment in Education: principles, policy & practice*, 10(2), pp. 169-207.
- Harlen, W. and Gardner, J. (2010) Assessment to support learning. *Developing teacher assessment*, pp. 15-28.
- 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.
- Heldsinger, S. and Humphry, S. (2010) Using the method of pairwise comparison to obtain reliable teacher assessments. *The Australian Educational Researcher*, 37(2), pp. 1-19.
- Heldsinger, S.A & Humphry, S.M. (2013) Using calibrated exemplars in the teacher-assessment of writing: an empirical study, *Educational Research*, 55(3), pp. 219-235.
- Hill, D., & Kumar, R., eds. (2012) *Global neoliberalism and education and its consequences*. Routledge.
- Holloway, I. and Wheeler, S. (2002) *Qualitative Research in Nursing*. 2nd edn. Oxford: Blackwell Publishing.
- Hopfenbeck, T.N., Lenkeit, J., El Masri, Y., Cantrell, K., Ryan, J. and Baird, J.A. (2018) Lessons learned from PISA: A systematic review of peer-reviewed articles on the programme for international student assessment. *Scandinavian Journal of Educational Research*, 62(3), pp. 333- 353.
- Hoyle, E. and Wallace, M. (2007) Educational reform: An ironic perspective. *Educational management administration & leadership*, 35(1), pp. 9-25.

- Humphry, S. M. and McGrane, J. A. (2015) Equating a large-scale writing assessment using pairwise comparisons of performances. *The Australian Educational Researcher*, 42(4), pp. 443-460.
- Illeris, K. (2016). *How we learn : Learning and non-learning in school and beyond*. Abingdon: Routledge.
- Illeris, K. (2017) *Learning, Development and Education : From Learning Theory to Education and Practice*. Abingdon: Routledge.
- James, M. (2006). Assessment, teaching and theories of learning. In J. Gardner Ed. (2006) *Assessment and Learning*, 1<sup>st</sup> edn. London: Sage.
- James, M. (2015) Assessing without levels: From measurement to judgement. In *Keynote address given at conferences for school leaders in 2014*.
- James, M. (2017) (Re) viewing assessment: changing lenses to refocus on learning. *Assessment in Education: Principles, Policy & Practice*, 24(3), pp. 404-414.
- James, M. and Lewis, J. (2012) Assessment in Harmony with our Understanding of Learning: Problems and Possibilities. In Gardner, J. (Ed.). (2012). *Assessment and learning*. Sage.
- James, M., Oates, T., Pollard, A. and Wiliam, D. (2011b) *The Framework for the National Curriculum – a report by the Expert Panel for the National Curriculum*. London: Department for Education.
- Jarvis, P., Holford, J. & Griffin, C. (2003) *The theory & practice of learning*, 2nd edn. Kogan Page: London.
- Johnson, S. (2013) On the reliability of high-stakes teacher assessment, *Research Papers in Education*, 28(1), pp. 91-105,
- Kane, M. T. (1992) An argument-based approach to validity. *Psychological bulletin*, 112(3), p. 527.
- Kane, M. T. (2010) Validity and fairness. *Language testing*, 27(2), pp. 177-182.
- Kane, M. T. (2016) Explicating validity. *Assessment in Education: Principles, Policy & Practice*, 23(2), pp. 198-211.
- Koretz, D.M. (2008) *Measuring up*. Harvard University Press.
- Koretz, D. (2016). Making the term ‘validity’ useful. *Assessment in Education: Principles, Policy & Practice*, 23(2), pp. 290-292.
- Koretz, D. (2017) *The testing charade: Pretending to make schools better*. University of Chicago Press.
- Kumar, R., & Hill, D. (2012) Introduction: Neoliberal capitalism and education. In Hill, D. and Kumar, R., eds. (2012) *Global neoliberalism and education and its consequences*. Routledge.

- Kyriacou, C (2001) teacher Stress: Directions for future research, *Educational Review*, 53(1), pp. 27-35.
- Lawy, R and Tedder, M., (2012) Beyond compliance: teacher education practice in a performative framework, *Research Papers in Education*, 27(3), pp. 303-318.
- Lincoln, Y. S., & Guba, E. G. (1986) But is it rigorous? Trustworthiness and authenticity in naturalistic evaluation. *New directions for program evaluation*, 1986(30), pp.73-84.
- Lilly, J., Peacock, A., Shoveller, S. and Struthers, D.R. (2014) *Beyond levels: Alternative assessment approaches developed by teaching schools: research report*. London: National College for Teaching and Leadership.
- Linn, R. L. (2000) Assessments and accountability. *Educational researcher*, 29(2), pp. 4-16.
- Lucey, H. and Reay, D. (2002) A market in waste: psychic and structural dimensions of school-choice policy in the UK and children's narratives on 'demonized' schools. *Discourse: studies in the cultural politics of education*, 23(3), pp. 253-266.
- Marlow, R., Norwich, B., Ukoumunne, O. C., Hansford, L., Sharkey, S. and Ford, T. (2014) A comparison of teacher assessment (APP) with standardised tests in primary literacy and numeracy (WIAT-II), *Assessment in Education: Principles, Policy & Practice*, 21(4), pp. 412-426.
- Merriam-Webster. (n.d.). Merriam-Webster.com dictionary. Available at <https://www.merriam-webster.com/dictionary/assess> [Accessed 1 May 2020]
- Metcalf, S. (2017) *Neoliberalism: The Idea That Swallowed The World: a talk*. Available at: <https://www.theguardian.com/news/2017/aug/18/neoliberalism-the-idea-that-changed-the-world> [Accessed 2 August 2018]
- Messick, S. (1986) The once and future issues of validity: Assessing the meaning and consequences of measurement. *ETS Research Report Series*, 1986(2), pp. i-24.
- Messick, S. (1990) Validity of test interpretation and use.
- Messick, S. (1996) Validity and washback in language testing. *Language testing*, 13(3), pp. 241-256.
- Messick, S. (1998) Test validity A matter of consequence. *Social Indicators Research*, 45(1-3), pp. 35-44.
- McIntosh, J. (2015) *Final report of the Commission on Assessment without Levels*
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Sage.
- Miles, M. B., Huberman, A. M. and Saldana, J. (2014) *Qualitative Data Analysis- Methods Sourcebook*. 3<sup>rd</sup> edn. London. SAGE
- Moore, A. and Clarke, M. (2016) 'Cruel optimism': teacher attachment to professionalism in an era of performativity. *Journal of Education Policy*, 31(5), pp. 666-677.

- Morgan, D. L. (2014) Pragmatism as a paradigm for social research. *Qualitative inquiry*, 20(8), pp. 1045-1053.
- Moss, G. (2009) The politics of literacy in the context of large-scale education reform. *Research Papers in Education*, 24(2), pp. 155-174.
- Moss, A. (2016) Shifting the focus of validity for test use. *Assessment in Education: Principles, Policy & Practice*, 23(2), pp. 236-251.
- Moss, G. (2017) Assessment, accountability and the literacy curriculum: reimagining the future in the light of the past. *Literacy*, 51(2), pp.56-64.
- Murphy, M, ed. (2013). *Social theory and education research: understanding Foucault, Habermas, Bourdieu and Derrida*. Routledge.
- National Statistics (2020) *Search for schools and colleges to compare*. Available at: <https://www.compare-school-performance.service.gov.uk/find-a-school-in-england> [Accessed 20 June 2020]
- Newton, P. E. (2011) A level pass rates and the enduring myth of norm-referencing. *Research Matters*, 2, pp. 20-26.
- Newton, P. E., and Baird, J. (2016) The great validity debate. pp. 173-177.
- Newton, P. and Shaw, S. (2014) *Validity in educational and psychological assessment*. Sage.
- Newton, P. E., & Shaw, S. D. (2016) Disagreement over the best way to use the word 'validity' and options for reaching consensus. *Assessment in Education: Principles, Policy & Practice*, 23(2), pp. 178-197.
- Nisbet, I. and Shaw, S. D. (2019) Fair assessment viewed through the lenses of measurement theory. *Assessment in Education: Principles, Policy & Practice*, 26(5), pp. 612-629.
- No More Marking. (2021) *Comparative Judgement for schools*. Available at: <https://www.nomoremarking.com> [Accessed 17 June 2021]
- Norris, C. (1997) *Against relativism: Philosophy of science, deconstruction, and critical theory*. Blackwell.
- Ofsted. (2015) *school inspection handbook*. London: Ofsted
- Oliver, D. G., Serovich, J. M., & Mason, T. L. (2005). Constraints and opportunities with interview transcription: towards reflection in qualitative research. *Social Forces*, 84(2), pp. 1273+
- Olson, D.R. and Bruner, J.S. (1996) Folk psychology and folk pedagogy. *The handbook of education and human development*, pp.9-27.
- Ong, A. (2007). Neoliberalism as a mobile technology. *Transactions of the Institute of British Geographers*, 32(1), 3-8.

- Opie, C. (2004). *Doing Educational Research*. London. SAGE
- Oxford English Dictionary (2020) *OED Online*. Available at: <https://www-oed-com.ezproxy.bcu.ac.uk/> [Accessed July 01, 2020]
- Perryman, J., Ball, S. J., Braun, A., & Maguire, M. (2017) Translating policy: Governmentality and the reflective teacher. *Journal of Education Policy*, 32(6), 745-756.
- Perryman, J., Maguire, M., Braun, A., & Ball, S. (2018) Surveillance, Governmentality and moving the goalposts: The influence of Ofsted on the work of schools in a post-panoptic era. *British journal of educational studies*, 66(2), 145-163.
- Piaget, J. (1952 [1936]) *The origins of intelligence in children*, 2nd edn.
- Plank, S. B. (2013) Pressures of the Season: An examination of Classroom Quality and High-Stakes Accountability, *American Education Research Journal*, 50(5), pp. 1152-1182.
- Poet, H., Sharp, C., Garry, J., Harland, J., Keightley, G., Kirkup, C. and Martin, K., 2018. Assessment without levels: qualitative research. December 2018.
- Pollitt, A. (2012). The method of adaptive comparative judgement. *Assessment in Education: principles, policy & practice*, 19(3), pp. 281-300.
- Pratt, N. (2016) Neoliberalism and the (internal) marketisation of primary school assessment in England. *British Educational Research Journal*, 42(5), pp. 890-905.
- Pratt, N. (2018) Playing the levelling field: teachers' management of assessment in English primary schools. *Assessment in Education: Principles, Policy & Practice*, pp. 1-15.
- Pratt, N. and Alderton, J. (2017) It's different, it's difficult, it's unknown': Letting go of levels.
- 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.
- Alderton, J., & Pratt, N. (2021) Filling gaps: assessment software and the production of mathematics and its teaching and learning in primary schools. *Critical Studies in Education*, pp. 1-15.
- Priestley, M. (2011a) Whatever happened to curriculum theory? Critical realism and curriculum change. *Pedagogy, Culture & Society*, 19(2), pp.221-237.
- Priestley, M. (2011b) Schools, teachers, and curriculum change: A balancing act?. *Journal of educational change*, 12(1), pp.1-23.
- Priestley, M., Biesta, G. and Robinson, S. (2013) Teachers as agents of change: teacher agency and emerging models of curriculum. *Reinventing the curriculum: New trends in curriculum policy and practice*, pp.187-206.
- Priestley, M., Biesta, G. and Robinson, S. (2015a) *Teacher agency: An ecological approach*. Bloomsbury Publishing.

- Priestley, M., Biesta, G. and Robinson, S. (2015b) Teacher agency: what is it and why does it matter?. In *Flip the system* (pp. 134-148). Routledge.
- Priestley, M., Philippou, S., Robinson, S. & Biesta, G. (2016) *The teacher and the Curriculum: Exploring teacher Agency*.
- Pritchard, A. (2017) *Ways of Learning: Learning Theories for the Classroom*, 4th edn, Routledge, Milton.
- Punch, K.F. (2009). *Introduction to Research Methods in Education*. London, SAGE
- R-Project (2021) (<https://www.r-project.org/> [accessed: 20/12/2021])
- Sayer, A. (1999) *Realism and social science*. Sage.
- Scharaschkin, A. (2017) Commentary on Baird, J., Andrich, D., Hopfenbeck, TN and Stobart, G., 'assessment and learning: fields apart'. *Assessment in Education: Principles, Policy & Practice*, 24(3), pp.454-462.
- Schmidt, P. (2017) Why policy matters: Developing a policy vocabulary within music education. *Policy and the political life of music education*, pp. 11-36.
- Schoenfeld, A.H. (2017) On learning and assessment. *Assessment in Education: Principles, Policy & Practice*, 24(3), pp.369-378.
- Schutt, R.K. (2007) *The Blackwell Encyclopaedia of Sociology*, Volume III (Ed. G. RITZER) (Oxford, Blackwell).
- Sfard, A. (1998) On two metaphors for learning and the dangers of choosing just one. *Educational researcher*, 27(2), pp.4-13.
- Shain, F & Gleeson, D. (1999) Under new management: changing conceptions of teacher professionalism and policy in the further education sector, *Journal of Education Policy*, 14:4, 445-462,
- Shepard, L. A. (2016). Evaluating test validity: Reprise and progress. *Assessment in Education: Principles, Policy & Practice*, 23(2), 268-280. Sireci, S. G. (2016). On the validity of useless tests. *Assessment in Education: Principles, Policy & Practice*, 23(2), 226-235.
- Shute, V.J., 2008. Focus on formative feedback. *Review of educational research*, 78(1), pp.153-189.
- Singh, K. (2007). *Quantitative social research methods*. SAGE Publications.
- Sireci, S. G. (2016) On the validity of useless tests. *Assessment in Education: Principles, Policy & Practice*, 23(2), pp. 226-235.
- Skinner, B., Leavey, G., & Rothi, D. (2018) Managerialism and teacher professional identity: impact on well-being among teachers in the UK. *Educational Review*, pp. 1-16.
- Snow, R. E., & Lohman, D. F. (1993) Cognitive psychology, new test design, and new test theory: An introduction. *Test theory for a new generation of tests*, pp. 1-17.

- Stake, R.E. (1995). *The Art of Case Study Research*. London. SAGE
- Stake, R.E. (2006). *Multiple Case Study Analysis*. London, Gilford Press
- Standards and Testing Agency (2020) *Understanding scaled scores at key stage 2*. Available at: <https://www.gov.uk/guidance/scaled-scores-at-key-stage-2>[Accessed 20 June 2020]
- Statskingdom (2021) [https://www.statskingdom.com/170median\\_mann\\_whitney.html](https://www.statskingdom.com/170median_mann_whitney.html) [accessed: 20/12/2021].
- Stenner, A. J., Smith III, M. and Burdick, D. S. (1983) Toward a theory of construct definition. *Journal of educational measurement*, pp. 305-316.
- Stevenson, N. (2010) Education, neoliberalism and cultural citizenship: Living in 'X Factor' Britain. *European Journal of Cultural Studies*, 13(3), pp. 341-358.
- Stobart, G. (2009) Determining validity in national curriculum assessments, *Educational Research*, 51:2, pp. 161-179,
- Subedi, B. P. (2016) Using Likert type data in social science research: Confusion, issues and challenges. *International journal of contemporary applied sciences*, 3(2), pp. 36-49.
- Tarricone, P. and Newhouse, C. P. (2016) Using comparative judgement and online technologies in the assessment and measurement of creative performance and capability. *International Journal of Educational Technology in Higher Education*, 13(1), pp. 1-11.
- Tennent, W. (2020) The assessment of reading comprehension in English primary schools: investigating the validity of the Key Stage 2 reading standard assessment test (SAT). *Education*, 3(13), pp.1-14.
- Thurstone, L. L. (1927) A Law of Comparative Judgment. *Psychological Review*, 34, pp.273–286.
- Tierney, R.D. (2006) Changing practices: Influences on classroom assessment. *Assessment in Education*, 13(3), pp.239-264.
- Tobin, M., Nugroho, D., and Lietz, P (2016) Large-scale assessments of students' learning and education policy: synthesising evidence across world regions. *Research Papers in Education*, 31(5), pp.578-594
- Torrance, H (2011) Using Assessment to Drive the Reform of schooling: Time to Stop Pursuing the Chimera? *British Journal of Educational Studies*, 59(4), pp. 459-485
- Training and Development Agency (2007) *The Professional Standards For Teachers*. London: TDA
- Vanderstraeten, R. and Biesta, G. (2006) How is education possible? Pragmatism, communication and the social organisation of education. *British Journal of Educational Studies*, 54(2), pp.160-174.

- Verhavert, S., Bouwer, R., Donche, V. and De Maeyer, S. (2019) A meta-analysis on the reliability of comparative judgement. *Assessment in Education: Principles, Policy & Practice*, 26(5), pp. 541-562.
- Vygotsky, L. S. (1978) *Mind in Society: The Development of Higher Psychological Processes*. Cambridge: Harvard University Press.
- Vygotsky, L. S. (1986 [1934]) *Thought and language*. Cambridge, MA: MIT Press.
- Ward, G. and Quennerstedt, M. (2019) Curiosity killed by SATs: An investigation of mathematics lessons within an English primary school. *Education 3-13*, 47(3), pp.261-276.
- Webb, J., Schirato, T. and Danaher, G. (2002). *Understanding Bourdieu*. Sage.
- Webb, R. and Vulliamy, G. (1999) Changing times, changing demands: a comparative analysis of classroom practice in primary schools in England and Finland. *Research papers in education*, 14(3), pp.229-255.
- Webb, R., Vulliamy, G., Sarja, A., Hämäläinen, S. and Poikonen, L. (2009) Professional learning communities and teacher well-being? A comparative analysis of primary schools in England and Finland. *Oxford Review of Education*, 35(3), pp.405-422.
- West, A. and Ylönen, A. (2010) Market-oriented school reform in England and Finland: school choice, finance and governance. *Educational Studies*, 36(1), pp.1-12.
- Wheadon, C., Barmby, P., Christodoulou, D. and Henderson, B. (2020) A comparative judgement approach to the large-scale assessment of primary writing in England. *Assessment in Education: Principles, Policy & Practice*, 27(1), pp. 46-64.
- Whetton, C. (2009) A brief history of a testing time: national curriculum assessment in England 1989–2008, *Educational Research*, 51(2), pp. 137-159.
- Whitty, G., & Wisby, E. (2016). Education in England—a testbed for network governance? *Oxford Review of Education*, 42(3), pp. 316-329.
- William, D. (1996). Standards in examinations: a matter of trust?. *The Curriculum Journal*, 7(3), 293-306.
- William, D. (2001) Reliability, validity, and all that jazz. *Education 3-13*, 29(3), pp.17-21.
- William, D. (2006) Formative assessment: Getting the focus right. *Educational assessment*, 11(3-4), pp.283-289.
- William, D. (2010) An integrative summary of the research literature and implications for a new theory of formative assessment. *Handbook of formative assessment*, pp. 18-40.
- William, D. (2011) What is assessment for learning? *Studies in educational evaluation*, 37(1), pp. 3-14.

- William, D. (2017) Assessment and learning: some reflections. *Assessment in Education: Principles, Policy & Practice*, 24(3), pp.394-403.
- William, D; Black, P. (1996). Meanings and Consequences: a basis for distinguishing formative and summative functions of assessment? *British Educational Research Journal*, 22(5), pp. 537-548.
- William, D., Lee, C., Harrison, C. and Black, P. (2004) teachers developing assessment for learning: Impact on student achievement. *Assessment in education: principles, policy & practice*, 11(1), pp.49-65.
- Williams, J.C. (2015) "Assessing Without Levels": preliminary research on assessment literacy in one primary school. *Educational Studies*, 41(3), pp.341-346.
- Wilkins, C. (2011) Professionalism and the post-performative teacher: new teachers reflect on autonomy and accountability in the English school system. *Professional development in education*, 37(3), pp.389-409.
- Wilkins, C. (2015) Education reform in England: quality and equity in the performative school. *International Journal of Inclusive Education*, 19(11), pp.1143-1160.
- Wilkins, C., Busher, H., Kakos, M., Mohamed, C. and Smith, J. (2012) Crossing borders: new teachers co-constructing professional identity in performative times. *Professional development in education*, 38(1), pp.65-77.
- Winter, C. (2017) Curriculum policy reform in an era of technical accountability: 'fixing' curriculum, teachers and students in English schools. *Journal of Curriculum Studies*, 49(1), pp. 55-74.
- de Winter, J. F., & Dodou, D. (2010) Five-point likert items: t test versus Mann-Whitney-Wilcoxon (Addendum added October 2012). *Practical Assessment, Research, and Evaluation*, 15(11), pp.1-17
- Wolming, S. and Wikström, C. (2010) The concept of validity in theory and practice. *Assessment in Education Principles, Policy & Practice*, 17(2), pp. 117-132.
- Wyatt-Smith, C., Klenowski, V. and Gunn, S. (2010) The centrality of teachers' judgement practice in assessment: A study of standards in moderation. *Assessment in Education: Principles, policy & practice*, 17(1), pp. 59-75.
- Wyse, D & Torrance, H (2009) The development and consequences of national curriculum assessment for primary education in England. *Educational Research*, 51(2), pp. 213-228.
- Yeager, D. S., & Dweck, C. S. (2020) What can be learned from growth mindset controversies?. *American Psychologist*, 75(9), pp.1269-1284.
- Yin, R.K. (2009) *Case Study Research: Design and methods*. 4<sup>th</sup> edn. London: SAGE.

## APPENDICES

---

### APPENDIX 1 RELEASE OF ASSESSMENT GUIDANCE

Assessment without levels testing and guidance	Date
2016 key stage 1 and 2 Sample questions and mark scheme	14 <sup>th</sup> July 2014
National Curriculum in England: English and Maths programmes of study	16 <sup>th</sup> July 2014
National Curriculum and assessment: information for schools	21 <sup>st</sup> August 2014
National Curriculum in England: framework for key stages 1 to 4	2 <sup>nd</sup> Dec 2014
National Curriculum in England: primary curriculum	6 <sup>th</sup> May 2015
Interim teacher assessment frameworks at the end of key stage 1 and 2	17 <sup>th</sup> Sept 2015
2016 teacher assessment exemplification: end of key stage 2 Mathematics Working at the expected standard	Jan 2016
2016 teacher assessment exemplification: end of key stage 1 Mathematics Working at the expected standard	Jan 2016
2016 teacher assessment exemplification: end of key stage 1 English writing Working at greater depth within the expected standard: Ali	Feb 2016
2016 teacher assessment exemplification: end of key stage 2 English writing Working towards the expected standard: Alex	Feb 2016
2016 teacher assessment exemplification: end of key stage 2 English writing Working at the expected standard: Leigh	Revised Mar 2016
2016 teacher assessment exemplification: end of key stage 2 English writing Working at Greater depth Within the expected standard: Frankie	Revised Mar 2016

2016 teacher assessment exemplification: end of key stage 1 English writing Working at the expected standard: Sam	Revised Mar 2016
2016 teacher assessment exemplification: end of key stage 1 English writing Working at greater depth within the expected standard: Ali	Revised Mar 2016
2016 teacher assessment exemplification: end of key stage 1 Reading Working at greater depth within the expected standard	Apr 2016
Interim teacher assessment frameworks at the end of key stage 2	July 2016
Interim teacher assessment frameworks at the end of key stage 1	Dec 2016
Teacher assessment frameworks at the end of key stage 1 and 2 for 2017/18	15 <sup>th</sup> Sept 2017
2018 teacher assessment exemplification: KS2 English writing	19 <sup>th</sup> Oct 2017
2018 teacher assessment exemplification: KS1 English writing	19 <sup>th</sup> Oct 2017
Teacher assessment frameworks at the end of key stage 1 and 2 for 2018/19 onwards	14 <sup>th</sup> Feb 2018
Teacher assessment exemplification: KS1 mathematics	10 <sup>th</sup> Dec 2018
Teacher assessment exemplification: KS1 English reading	11 <sup>th</sup> Dec 2018

## APPENDIX 2 APPROVAL OF ETHICAL APPLICATION RELEASE OF ASSESSMENT GUIDANCE



Faculty of Health, Education and Life Science Research Office  
Faculty of Health, Education and Life Sciences  
Birmingham City University  
Westbourne Road  
Birmingham  
B15 3TN

[HELS\\_Ethics@bcu.ac.uk](mailto:HELS_Ethics@bcu.ac.uk)

15/02/2017  
Miss victoria birmingham

Dear Miss victoria birmingham

**Re:** The government's "assessment without levels policy" impact on assessment practice in primary schools. - birmingham /Feb /2017 /Am /0626

Thank you for your application for approval of amendments regarding the above study. I am happy to take Chair's Action and approve the amendments which means you may continue your research.

The Committee's opinion is based on the information supplied in your application. If you wish to make any substantial changes to the research please contact the Committee and provide details of what you propose to alter. A substantial change is one that is likely to affect the

- safety and well-being of the participants;
- scientific value of the study;
- conduct or management of the study.

The Committee should also be notified of any serious adverse effects arising as a result of this research. The Committee is required to keep a favourable opinion under review in the light of progress reports.

I wish you every success with your study.

Yours sincerely,

Dr Carolyn Blackburn

On behalf of the Faculty Academic Ethics Committee

## Assessment Without Levels Study

---

### Page 1: Page 1

Thank you for your interest in my research. This is a short survey to get an idea of what teachers are thinking about Assessing Without Levels.

This survey is specifically for Class Teachers in KS1 and KS2, who teach using the National Curriculum. It is **not** suitable for EYFS Class Teachers.

During the survey the term Teacher Summative Assessment will refer to any Summative Assessment carried out by the teacher to determine the achievement of a child in National Curriculum expectations.

Assessment Without Levels will refer to the assessment of children after Levels were removed from the assessment of learning.

The term External Tests will be used to describe assessment that is supplied by an organisation outside of the school including SATs and any optional test schemes your school uses at the end of School years.

Core or Foundation subjects relate (here) to only Maths, Writing, Reading.

Non-core subjects or non-foundation subjects relate (here) to all other subjects taught.

### Page 2: Data Protection

Data obtained through this questionnaire may be reproduced and published in a variety of forms and for a variety of audiences related to the broad nature of the research (i.e. conferences, peer reviewed journal, articles etc.), but will all be done anonymously.

All data collected in this survey will be held anonymously and securely. Although I ask for some personal details in this survey, these are for purposes of inclusion and equality only and will not be revealed to any others.

By clicking 'next' and continuing with the survey you are giving consent for your responses to be used anonymously as part of this research.

Once you click 'next' you will be directed to the survey.

### Page 3: Participant information

What Key Stage do you teach?

- KS1                       KS2

What year do you currently teach?

- Year 1                       Year 2                       Year 3  
 Year 4                       Year 5                       Year 6

What years have you previously taught?

- Year 1                       Year 2                       Year 3  
 Year 4                       Year 5                       Year 6

Have you assessed using National Curriculum Levels?

- Yes                       No

Have you assessed without National Curriculum Levels?

- Yes                       No

3 / 15

How many years have you been teaching?

- Less than 1 year                       1-5 years                       6- 10 years  
 11-15 years                       16-20 years                       21+ years

4 / 15

## Page 4: Experience Assessing Summatively.

Please select how much you agree or disagree with each of these statements by clicking only one item per row.

These questions relate to Teachers' Summative assessment of Maths and Reading.

[+ More info](#)

	Strongly Disagree	Disagree	Neither agree or Disagree	Agree	Strongly Agree	N/A
I feel I can confidently assess WITHOUT levels.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
There has been an easy transition from assessing WITH levels to without levels.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Assessing WITH levels was a better way of tracking children's progress.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

5 / 15

There has been enough guidance from the Government to help teachers assess summatively WITHOUT levels.	<input type="checkbox"/>					
The interim framework helps me understand how to summatively assess children's work WITHOUT levels.	<input type="checkbox"/>					
Detailed exemplars for each year and each subject would help improve teacher confidence and accuracy when assessing maths and reading WITHOUT levels.	<input type="checkbox"/>					

Please use this space to leave any comments about Teacher Assessment of Reading and Maths.

6 / 15



## Page 5: Experience Assessing Summatively cont...

Please select how much you agree or disagree with each of these statements by clicking only one item per row.

These questions relate to External Summative Tests of reading and maths.

	Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree
External tests, like the KS1 and KS2 SATs, are a good way to assess CORE subjects.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
External tests, like the KS1 and KS2 SATs, are better now assessing WITHOUT levels.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Some parts of the curriculum are better assessed in external tests.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Some parts of the curriculum are better assessed in teacher summative assessment.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Only external tests should be used to assess CORE subjects at the end of KS1 and KS2.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Only teacher summative assessment should be used to assess CORE subjects.	<input type="checkbox"/>				
---	--------------------------	--------------------------	--------------------------	--------------------------	--------------------------

Please use this space to leave any comments about the external assessment (SATs or anyother test schemes used by the school) of Reading and Maths.

## Page 6: Experience Assessing Summatively cont...

Please select how much you agree or disagree with each of these statements by clicking only one item per row.

These questions will relate to the summative assessment of non-core subjects-for example: history, geography, RE, PE, art...

	Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	N/A
Assessing WITHOUT levels takes focus away from NON-CORE subjects.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Assessing WITHOUT levels has not improved assessment of NON-CORE subjects.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Assessing WITHOUT levels has brought a balance to assessment of CORE and NON-CORE subjects.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Assessing WITHOUT was a balanced way to assess CORE and NON-CORE subjects.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Tests are used to assess NON-CORE subjects in my school.	<input type="checkbox"/>					
Test should be used to assess NON-CORE subjects.	<input type="checkbox"/>					

Please use this space to leave any comments about the assessment of non-core subjects.

## Page 7: Experience Assessing Summatively cont...

Please select how much you agree or disagree with each of these statements by clicking only one item per row.

These questions will relate to the summative assessment of Writing.

Please don't select more than 1 answer(s) per row.

Please select at least 6 answer(s).

	Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree	N/A
The interim framework helps me to assess writing WITHOUT levels.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Moderation courses make it clear how to assessment writing WITHOUT levels.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Assessing writing WITH levels enabled consistent assessment within and between schools.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Assessing writing WITHOUT levels enables consistent assessment within and between schools.	<input type="checkbox"/>					
Teaching and testing grammar improves children's ability to write.	<input type="checkbox"/>					
The increased focus on grammar impleads children's writing.	<input type="checkbox"/>					
A variety of writing exemplars would help in assessing writing WITHOUT levels.	<input type="checkbox"/>					

Please use this space to leave any comments about the assessment of writing WITHOUT levels.

Has this questionnaire covered all the aspects of Assessment Without Levels that you feel are important? \* *Required*

Yes
  No

If you answered no, please leave a comment below:

## Page 8: Thank you!

Thank you for taking the time to answer my questionnaire. Your contribution to my research is greatly appreciated.

If you would like anymore information regarding this survey or claification in how your data will be used please contact [victoria.birmingham@bcu.ac.uk](mailto:victoria.birmingham@bcu.ac.uk)

---

## APPENDIX 4 PILOT STUDY INTERVIEW SCHEDULE

### Pilot study Semi-structured interview Part 1

What have been your experiences assessing maths without levels?

What have been your experiences assessing reading without levels?

What have been your experiences assessing maths with levels?

What have been your experiences assessing reading with levels?

How do you think they compare?

How has assessing without levels affected how you assess noncore subjects?

### Pilot study Semi-structured interview Part 2

How would you describe the relationship between external tests SATs and teacher summative assessment?

Are there areas of the curriculum that you feel are best suited to being assessment via a test or teacher assessment?

What impact has the government interim framework had on your assessment?

What additional information or resource do you feel would positively impact your teacher assessment?

## APPENDIX 5 INFORMATION SHEETS AND CONSENT FORMS

### The Government's "assessment without levels policy" impact on assessment practice in primary schools.

My name is Victoria Birmingham and I am PhD student in the School of Education at Birmingham City University. I have been a primary school teacher since 2012 and it was experiencing the changes in assessment that has driven me to explore the topic of assessment in a PhD. From starting in February 2016 I am now at the stage where I am looking for Primary schools who will be interested in taking part in my research.

#### **Why is the research being done?**

The changes in assessment are very new so there is a very limited amount of research into it. So I would like to build up that research by discovering teacher's and school's experiences of assessing without levels.

#### **What I want to find out?**

These are the research questions I want to find out in the study:

- How has assessment without levels affected teachers' assessment in KS1 and KS2?
- To what extent do teachers' assessments correlate with external summative tests (SATs) at the end of KS2 and KS1 in maths and reading since the removal of levels?
- How are the new curriculum objectives assessed in teachers' summative assessment and external national summative tests (SATs) at the end of KS1 and KS2 in maths and reading since the removal of levels?

I'll be using questionnaires and interviews with teachers to explore these questions.

#### **Your school's role in the research and how will your staff contribute?**

Should you agree to take part in this study, your contribution will be by allowing me to conduct research in your school and have access to your teaching staff and participating in an interview about your experiences of the change in assessment.

All your KS1 and KS2 teaching staff will be invited to fill out a short questionnaire about their experiences assessing. It will be online so they can be filled out easily at a time to suit them. I would also like to introduce my study to your teacher in a Staff meeting, if at all possible where paper copies of the survey will be available.

The next stage will be teacher interviews of those who teach year 6 and/or year 2 this academic year. This will be the main focus of my research as these are the only years assessment is performed in an external test and using teacher assessment.

Second interviews will take place with the same year 2 and/or Year 6 teachers from this academic year. Here anonymised assessment data from the KS2 and KS1 SATs and the teachers summative assessment will be discussed. Please note that in no way will the test data be used to judge teachers own summative assessment. The purpose of comparing the two sets of assessment data is to explore how the new curriculum is represented and assessed in the different assessment forms.

My aim is to make your staff feel as comfortable as possible. They are helping me out by taking part and so I am very grateful for their time. Therefore, they are free to take part in whichever stages of the research they like. If they do not wish to take part I will completely respect this decision.

### **How does contributing benefit your school?**

By taking part in this research you will play a significant role in the development research into the assessment changes in primary school. The originality of this research means that the findings will be presented at conferences and in publications in the field, meaning that you will help shine a light on the changes in assessment and how they are impacting school life.

Furthermore, by reflecting on their practice, it is hoped that the teachers will also benefit in terms of their professional development. I will be more than happy to share my findings with you and run staff meetings to discuss them and my further research. You will also have access to my thesis once completed.

### **What will happen after the data has been collected?**

The paper copies of the questionnaires will be shredded as soon as I have typed up the findings onto a password-protected device. The voice recordings from the interviews will be stored on a password-protected device. I will then transcribe these interviews and delete the recordings after 2 years of the research completion in line with BCU guidelines. I will then analyse my quantitative and qualitative findings and highlight key findings from the research.

### **Confidentiality Agreement**

During the write up of the research, you, your school, your staff and children reserve the right to anonymity. I will use pseudonyms when referring them in my thesis so that nobody will be able to attribute what has been said to a particular person. This is part of a confidentiality agreement I have for the research.

*Thank you for your time reading this information sheet. If you have any further questions about this research please contact me on [victoria.birmingham@bcu.ac.uk](mailto:victoria.birmingham@bcu.ac.uk) or my Director of Studies on [Martin.Fautley@bcu.ac.uk](mailto:Martin.Fautley@bcu.ac.uk).*

### Head Teacher Consent Form

The aim of my research is to find out about teacher' thoughts and experiences of Assessment Without Levels. I seek to understand more about how the changes have affected assessment in primary schools and how the new curriculum is represented in external tests and teachers' summative assessment. This will be discovered both through a questionnaire, speaking to them in interviews and looking at past assessment data. By consenting to this research you will provide provisional agreement that some of your teachers may take part in my study.

However, on an individual level each teacher will need to make their own decisions about their level of involvement. Teachers may decide to just fill in the questionnaire, not take part at all, or complete all levels of the data collection. The interviews will only need to be with the teachers who taught year 2 and year 6 this academic year. This is a decision that I will encourage them to make for themselves. Further to this, teachers who do decide to take part will be able to withdraw from the study at any point throughout and after data collection without prejudice. All responses from teachers in the research will be made confidential and they will be anonymised in the write up of the research.

The purpose of this consent form is to ensure that you understand what my research entails and what your school will be required to do. In addition to this form, it will also be necessary to seek consent from teachers involved in the interviews and questionnaires.

Please contact me with any questions you may have about my research and the information I have provided you with at [victoria.birmingham@bcu.ac.uk](mailto:victoria.birmingham@bcu.ac.uk)

Once you have read the information sheet provided, please tick the box below as appropriate.

	Yes	No
I agree for my school to be participate in this research project.	<input type="checkbox"/>	<input type="checkbox"/>
I agree to the schools assessment data being used anonymously.	<input type="checkbox"/>	<input type="checkbox"/>

School Name and Address: \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

Name: \_\_\_\_\_ Date \_\_\_\_\_

Signature: \_\_\_\_\_

## Teacher information sheet

The Government's "assessment without levels policy" impact on assessment practice in primary schools.

### Aims of my study

My name is Victoria Birmingham and I am currently undertaking a PhD in the School of Education at Birmingham City University. The following is an information sheet about the research I am undertaking, which I have provided to help you make an informed decision about whether you would like to participate in the study.

As you will know and have already experienced the change in assessment following the removal of levels has been a big one. The way core subjects are assessed has changed and these changes need to be explored. Researching this will promote good practice, give a view into the experiences of schools and teachers working with and assessing Curriculum 2014 and explore the balance between external testing and teachers' summative assessment.

### Research questions

These are the research questions I want to find out in the study:

- Has assessment without levels affected teachers' assessment in KS1 and KS2?
- To what extent do teachers' assessments correlate with external summative tests (SATs) at the end of KS2 and KS1 in maths and reading since the removal of levels?
- How are the new curriculum objectives assessed in teachers' summative assessment and external national summative tests (SATs) at the end of KS1 and KS2 in maths and reading since the removal of levels?

### Invitation to participate

I would like to invite you to participate in my PhD research into Assessment Without Levels. You have been chosen to participate because as teachers you are at the forefront of the changes in assessment therefore your opinions and experiences are extremely valuable; they will form the bases of my research. Please note that all participation is on a voluntary bases.

### How you will participate

The study will comprise of a few stages: a questionnaire, semi structured interviews and access to end of year SATs and teacher assessment data.

The questionnaire is short and can be done online through the link I will provide. It can be filled out by all members of teaching staff.

The next stage of the study will be teacher interviews of those who taught Year 6 and Year 2 this academic year. This will be the main focus of my research as these are the only years assessment is performed in an external test and using teacher assessment.

Anonymised assessment data from the KS2 and KS1 SATs and your summative assessment will also be discussed. Please note that in no way will the test data be used to judge your own summative assessment. The purpose of comparing the two sets of assessment data is

to explore how the new curriculum is represented and assessed in the different assessment forms.

My aim is to make you feel as comfortable as possible. You are helping me out by taking part and therefore I am very grateful for your time. You are free to take part in whichever stages of the research you like. If you do not wish to take part at all I will completely respect your decision.

### Benefits of participation

In participating in this research you have the opportunity to share your experiences of assessing without levels. The findings of the research will be shared and you'll be able to gain insight into how the teaching community are adapting to assessing without levels. This will help to share good practice found and could help develop assessment practices in primary schools. You will have the opportunity to read the completed thesis and participated in staff meets where the findings will be presented and discussed.

### Any potential risks

This research hold very low risk to anyone involved. However if issues of a sensitive nature do arise participants will be signposted to the appropriate organisations that can help.

### Data Protection and confidentiality arrangements

The paper copies of the questionnaires will be shredded as soon as I have typed up the findings onto a password-protected device. The voice recordings from the interviews will be stored on a password-protected device. I will then transcribe these interviews and delete the recordings. I will then analyse my quantitative and qualitative findings and highlight key findings from the research.

During the write up of the research, you, your school, your staff and children reserve the right to anonymity. I will use pseudonyms when referring them in my thesis so that nobody will be able to attribute what has been said to a particular person. This is part of a confidentiality agreement I have for the research.

### Participation rights

As mentioned above participation is completely voluntary and you have the right to withdraw your consent to participate at any stage without prejudice.

*Thank you for your time reading this information sheet. If you have any further questions about this research please contact me on [victoria.birmingham@bcu.ac.uk](mailto:victoria.birmingham@bcu.ac.uk) or my Director of Studies on [martin.fautley@bcu.ac.uk](mailto:martin.fautley@bcu.ac.uk). Please fill out the consent forms attached to this information sheet.*

The Government’s “assessment without levels policy” impact on assessment practice in primary schools.

Summary of project

As detailed in the information sheet provided, my name is Victoria Birmingham and I am a PhD student in the School of Education at Birmingham City University. I am writing this to you in order to make sure that you have understood the information provided and are willing to take part in this research. The purpose of this consent form is to protect you, your school, and my University in the research.

To participate I would like you to fill in a questionnaire about summative assessment since the removal of levels. If you were a Year 2 or Year 6 teacher for this academic year I would like to invite you to an interview to talk about summative assessment since the removal of levels in more detail.

Please note, you will not be identified by name or through description in the research, and instead will be anonymised in the write up of the research. All interview recordings will be confidential and stored safely and securely in line with the Data Protection Act (1998) and the British Educational Research Association’s ethical guidelines.

Voluntary participation

Your participation in this research is voluntary and you have the right to withdraw, without prejudice, from the study at any time.

If you have questions

Please do not hesitate to contact me with any questions you may have after reading this at [victoria.birmingham@bcu.ac.uk](mailto:victoria.birmingham@bcu.ac.uk) .

Please read the information sheet provided and respond accordingly below by ticking the appropriate boxes in the table:

	Yes	No
I have received and understood the information about this research.		
I have read and understood the aims of the research.		
I have been offered the opportunity to ask questions about the research.		
I understand that I will be treated anonymously and that my responses are confidential.		
I agree to take part in an interview for the research.		
I agree to the interview I take part in being recorded.		
I understand that I can withdraw at any stage in the research without prejudice and that my responses will remain confidential.		

**Signature of participant:** .....

**Please print your name:** .....**Date:** .....

## APPENDIX 6 PILOT STUDY DATA ANALYSIS THEORY AND PROCEDURE

### Pilot Study Interview Setting Demographics

The pilot study School was a local authority maintained Community Primary School and nursery, within Birmingham. A large majority of students were in receipt of free School meals. In 2017, 60% of students met the expected standard in KS2 SATs for Maths, Reading and Writing compared to a national average of 60%.

### Participant backgrounds

#### Teacher A

Teacher A was a year 2 Teacher who had 24 years of teaching experience, including a break from full time teaching to have children, and taught on a supply basis when her children were young. She had been back to full time teaching for 7 years. She had been teaching in year 2 for 1 year. Previously she had taught in a variety of KS2 years.

#### Teacher B

Teacher B had been teaching for 9 years following a PGCE (Post graduate certificate in education). She had taught in year 2 at the beginning of her career, then in years 5 and 1, and back to year 2. She had been teaching in year 2 again for 2 consecutive years at this point. Her teaching experience was in a few other Schools, and she had taught at this School for the last 2 years.

#### Teacher C

Teacher C had 10 years teaching experience, following completion of the registered Teacher training program in early years. She had experience teaching for terms at a time in a variety of years, working on a supply basis when she first qualified. Her first permanent teaching role was for 2 years in early years. In her next position she taught in year 5 for 2 years, and has taught for the last 5 years in year 6.

#### Head Teacher Pilot study

HT1 came to the profession later than most by volunteering for a children's charity and play schemes. After initially being interested in teaching secondary maths, he trained as Primary

School Teacher. As a Teacher he worked mainly in KS2 with some short spells in KS1. His first position was 5 years, before moving to another School. After a few years, he taught in year 6 while taking the role of deputy head Teacher for 4 years. Following from this he took the headship at the Pilot study School, where he has held the position of head Teacher for 5 years. He does not do any classroom teaching.

#### Pilot study Data Qualitative analysis

Each semi-structured interview was transcribed, verbatim and punctuated by the researcher. This had a dual purpose of ensuring I was familiar with the data before analysing it and to most accurately structure the sentences to reflect the interviewee's meaning. If someone else punctuated the interview and had not been familiar with the topic, meaning could have subtly been changed by that person.

Transcripts were coded into themes systematically in multiple trawls (reading through and considering the mean of what was being said). A theme is here an identifiable topic of discussion that emerged from the data. Although themes were not decided in advance, it is acknowledged that themes were anticipated from the process of conducting the interviews based on the area of research and through repeated listening during the transcription process. Additionally themes were anticipated according to emergent themes in the literature, and the researcher's own experience of terms used within the area of research. However, themes were not confined to those known in the literature, or those anticipated, and no lists of themes were created until they emerged from the data when coding began. Braun and Clarke (2006: 12) name this approach as theoretical thematic coding, outlined as coding that would 'tend to be driven by the researcher's theoretical or analytic interest in the area' rather than inductive analysis, where data is coded '*without* trying to fit it into a pre-existing coding frame, or the researcher's analytic preconceptions.' More detail is provided below in [Thematic analysis](#).

This identified 55 themes. Similar themes were grouped, and overarching or parent themes were identified and shown in Table 0-1 below.

Table 0-1 The initial organisation of pilot study interview themes.

Teacher opinions and preference	Good change AWL	Implementation changes of new curriculum /New curriculum negatives		
		Test effects on children/AWL or New Curr – ve effects on children	High-stakes impact	Late or lack of guidance and support documents/ Extra guidance & clarification
TA and Test together	More freedom in assessment now	High expectations in curriculum	Performance pressure=SATs to justify TA/ Year 6 pressure/TA pressure, too much to cover/Time pressure, too much to cover/ Teach to the test	Confusion with TA/ Teacher confidence & inexperience with new system
Understand expectations this year with experience	Benefits to broader deeper curriculum	Best fit vs secure fit	High work load demands=Assessment not for children but monitoring/ Tracking/Box ticking instead of assessing	Moderation/Contradictions between guidance and moderation
Test over TA	Assessment driven by curriculum	High focus on SATs –ve=Narrowing curriculum/Chn's background & personal circumstance		Uncertainty
TA over test		Opposition to labelling children	Challenges with AWL reading/Decoding vs comprehension	Consistency with levels/AWL inconsistency issues
Using SA to inform teaching			TA to predict SATs	Needs experienced/skilled teachers
Assessing similar to with levels			Lack of trust in teachers/Need for physical evidence	Need for external programmes/ Admit need to still improve
Management separation from teaching and curriculum			Accountability problems/Accountability for a reason in context	Distrust in government policy
Label of AWL negative				Lack of parental understanding
				Teacher struggle with change/resistance
<b>BOLD= Theme created from change in assessment.</b>	Normal= Theme from old assessment system and new.	Blue= Head Teacher theme.		

These themes were identified primarily on printed transcripts with colour coded pens, following from this, all data and coding was transferred to the computer based coding programme Nvivo. All interviews were themed according to the groupings above. They were then further arranged and reordered with further emergence into and familiarity with the data.

The first set of theme codings were generated based on the information the Teachers were trying to relate about their experiences with assessment. They mirrored a lot of themes from the literature. However, they did not delve into why the Teachers viewed assessment this way or why they responded as they did. For this sociological lens was needed to interpret how and why the Teachers responded how they did. The last two trawls consisted of viewing the data this way and different themes emerged.

### Thematic analysis

A sociological angle was taken to explore why Teachers responded how they did to questions regarding assessment. This approach ties in with the thematic approach to coding

outlined above supported by Braun and Clarke (2006: 14) who state ‘...thematic analysis conducted within a constructionist framework cannot and does not seek to focus on motivation or individual psychologies, but instead seeks to theorise the socio-cultural contexts, and structural conditions, that enable the individual accounts that are provided.’ My previous theoretical framework from the first completion of data coding trawl benefited from the language I gained from sociological theory and ideas by Bourdieu and Foucault. These theorists linked to assessment practices within Schools and the effects accountability and performance systems have on Teachers’ assessment practices (Ball, 2012; Pratt, 2016; Pratt, 2018; Simon, 2002).

Bourdieu’s concepts of *field*, *habitus*, *illusio*, *recognition* and *misrecognition* will be used to consider and think about the views expressed by Teachers interviewed. As Bourdieu did not define specifically what he meant by the terms like *field* in his theory, the terms noted above will be interpreted based on the ways Bourdieu used them and made reference to them, and how education research itself has used his work to shape understanding of data. Bourdieu’s theory will not be alone in aiding structuring the analysis. Foucault’s ideas on surveillance and governmentality were be used to problematize views expressed in terms of monitoring and performance management. The language developed by these sociologists were be used as ‘thinking tools’ and allowed me to be specific about my ‘unit of analysis’ as recommended by James (2011). The combination of theories is encouraged by Murphy (2013: 7) to use parts of theory together in a way to potentially form original ideas and analyses just as theorist themselves did the same in creating different and innovative ways of theorising knowledge.

#### Habitus and field

Bourdieu outlines habitus as ‘a system of predispositions acquired through a relationship to a certain field,’ (Bourdieu, 1990: 90) which ‘becomes effective and operative when it encounters the conditions of its effectiveness, that is, conditions identical or analogous to those of which it is the product.’; this also defines field as the social environment or the ‘conditions’ experienced that created the habitus. Bourdieu does not define field explicitly but refers to it in terms of a persons’ ‘history’ (Bourdieu, 1990: 90) and the different situations that have produced the habitus they possess. The habitus is not fixed but embodied by actions and interaction with it. The habitus is not defined as rules, but as

regulation of practice, whereby 'the rule is not automatically effected by itself and that it obliges us to ask under what conditions are rule must operate' (ibid: 76), which Bourdieu (1990: 63) often refers to as 'a feel for the game'. It has to be in the best interests, judged by the agent, to follow the rule but it is not determined that they will.

Habitus was used to encompass the behaviour that regulates the actions of Teachers within the field, defined in this analysis 'as a field site, the structurally identifiable space which marks out the sphere of social activity' (Grenfell and James, 1998: 19). A field was assumed to operate in isolation, as there are many identifiable fields within and outside of that of Primary assessment (Grenfell and James, 1998).

#### Illusio and doxa

Illusio will be used to mean the 'conscious belief in the stakes (enjeux) of the game and the belief that it is worth playing.' (Pratt, 2016: 895). Doxa will be used to describe the dominant discourse within the field that separates out what can or cannot be thought (Pratt, 2016. P.895)

#### Agency

Webb et al. (2002: ix) define agency as 'The idea that individuals are equipped with the ability to understand and control their own actions, regardless of the circumstances of their lives: usually termed 'intentionality' and 'individuality''. Reay (1995: 355) conceptualises agency using Bourdieu's terms by stating that 'While the habitus allows for individual agency it also predisposes individuals towards certain ways of behaving.' This is interesting as the two definitions do not account for an individual's knowledge of their own habitus or the field of which they are in. Therefore, when habitus and field match you do not 'feel the weight of the water' you are in, however, when they do not match agents may become aware of their habitus, and thus display awareness of regulations that were once unnoticed. This may affect actions thus displaying a sense of agency, as the individual becomes aware of their habitus. In this analysis agency will be used to show awareness of the 'game' and a conscious choice of action based on their awareness of field structures at play.

#### Foucault perspective

Foucault's ideas on surveillance and the self-controlling and regulating behaviours they encourage and come to normalize (Allan in Murphy, 2013: 37) will be used to provide a framework for analysing data pertaining to that coding theme. Technologies of surveillance will be discussed in line with Foucault's ideas, as Fejes (in Murphy, 2013: 53) states:

*'In his writing, the focus was on analysing the technologies of power and domination and technologies of the self. The former concerns the practices through which the self is objectified and shaped through dividing practices while the latter concerns the ways in which the self constitutes itself as a subject'*

Ball (2002) used Foucault's concept of the panopticon to problematize surveillance of Teachers in Schools.

## Pilot study data analysis

### Pilot study interview data- themes

Using theoretical framework outlined in Pilot Study Analysis, positions teachers' regulation of their actions, in regards to primary assessment, as their habitus. Therefore, when AwLs was introduced and NCLs guidance removed, the *field* of primary assessment had changed, however, the habitus of the teachers had not.

#### 'Fish out of water'

A sense of displacement was identified through the teachers' comments in between how they 'knew' to assess and AwLs requirements of 'how' to assess.

I find it really hard to assess without levels. I think you become so used to one way of assessing...Teacher A

This indicated that their *habitus* remained in NCLs, while now being positioned in a *field* of AwLs, creating a feeling of being a 'fish out of water'. Participants felt uncertain about how to adapt to AwLs and align their *habitus* with this new *field*.

*I think they need to clarify it, exactly what's expected of us and exactly what these children need to be doing. Teacher A*

*I think they needed to clarify it[assessment expectations] all properly from the very beginning...everybody was just confused. We were all confused as to what was expected. Teacher B*

*It's been very confusing and very, it's felt very negative... Teacher C*

Far from rejecting the new assessment system, teachers were very keen to learn how to operate within it. Teacher B says directly that 'obviously we want to make sure that we are carrying out our assessments properly and in line with the guidelines.' With experience the teachers described feeling more at ease with AwLs.

*I do feel that this year we understand of what is expected of the children and therefore have been able to plan accordingly. Teacher B*

*...it was an uncomfortable experience the first year doing it. This year has been a lot easier and now I feel a lot more comfortable. Teacher C*

This also shows, as Bourdieu describes, that their *habitus* in their new *field* of AwLs was adapting through their continued experience. Teachers felt they could have adapted to AwLs sooner if guidance had been timelier.

*...obviously coming in then at that point in the year's too late...bringing it [ITAFs] in late didn't give us the time to cover what we need to cover. Teacher B*

*...it has become a little easier since they brought out the interim assessments [frameworks] which were obviously brought out quite late in the year Teacher C*

Participants were referring to the Teacher SA guidance, the ITAFs, which were published in March 2016 ahead of the first SATs tests AwLs in May 2016. However, this is ongoing and perhaps unstable, Teacher B is still unsure about whether the assessment policy will change again 'I think everybody is still quite confused with the whole thing and we don't know it's going to stay or whether it will be, whether they'll change it again', indicating a constant sense of instability inhibiting the teachers from understanding the *field* of AwLs.

Stress and performance pressures impact on teaching

As Bourdieu stated, *habitus* is not a rule that has to be followed, but regulations that will be acted upon when the agent judges them in their own best interest. As mentioned above, teachers had started to adapt to AwLs, however, their responses indicated that performance pressures encouraged some teaching practices which participants disagreed with. For example, in removing some children from 'ICT, Science, art, music, PE. The fun things' (Teacher A) to go over areas of maths and English work from the morning, Teacher A contributed towards children being 'drilled' for the purpose of testing, which they disagree with;

*...I think it's sad that at this age they should be out and about and learning from the world around them and not stuck in a class room being drilled for English and maths.*

Concerns were raised that SATs preparations caused their students stress in addition to a narrowing of the curriculum they experienced in favour of SATs tested curriculum content.

*There is so much to cover in such a short amount of time, they don't have the time to have a varied curriculum...some children go to pieces...and you know they're capable of so much more...I think it puts a lot of pressure on teachers because you just feel like you are constantly failing the children, especially when they've asked you to, you know [SATs]. Teacher B*

*I think it's sad then that you're pushing that pressure on children...[a student] he couldn't get past page, question 4 on the test, cause he couldn't fill that answer in, he couldn't do any more of the test and he cried. And I just think, these are little children what, what are we doing? Teacher A*

In their comments neither Teacher A nor B felt they had a choice but to adopt teaching strategies which they felt contributed to their pupils SATs preparation pressures. Teacher B felt that in preparing students for SATs she was 'failing the children' and that it was 'obvious' that the tests upset children. Aside from the stress and negative effect on learning participants disagreed with the secure fit assessment of writing without levels. This was seen to prioritise functional aspects of writing and penalise students for handwriting and spelling problems. For example;

*we've got EAL children, SEN children, I've got one boy who is dyslexic, brilliant use of language but he can't spell so he's always going to be working towards. And then we've got one who is an amazing writer, she's probably one of the best writers I've ever had but she can't spell because she's EAL...in theory you could have a page of writing with all these included, but that has no flair, creativity or imagination in the work. Teacher B*

*...they can't have year 6 level no matter how much effort they put into their work, no matter how hard they try they're never going to be allowed to have it because their handwriting isn't consistently neat and all the same. Teacher C*

Despite Teacher C's disagreement with the ITAF for writing they adopted it to direct content of writing for their students as part of their teaching in Year 6.

I am always encouraging content in their writing to meet the interim standards, but I don't base their actual assessments on them until the end of the year and I can only use evidence from January onwards anyway to evidence the interim standards assessment and I'm assessing from the beginning of the year so. Teacher C

Teacher C's compliance with assessment practices they disagree with, and reliance on the ITAF guidance, may be indicative of their uncertainty regarding the expectations of AwLs. The ITAFs may teach them how to play the game by knowing how to 'teach to the test' (TTTT). Their *illusio* may be explained by how they gained validation as a teacher through successful SATs results.

*It's [SATs results is] a bit of a verification almost for you, that your teacher assessments are very accurate...it highlights the teachers that understand the curriculum and understand their assessments. Teacher C*

There is security for Teacher C in knowing SATs expectations and aligning their teacher assessment as far as possible with this. This additionally shows a managerial logic their performance where an action can be taken, and the output measured; if they perform the action accurately, (following the rules) they will succeed. The action taken here is to align their teacher SA as closely as they can with the ITAFs, the output measured is then the SATs score, and evidence they have performed the action accurately is in how closely teacher SA matches SATs scores. There were signs of conflict with this practice, as the study teacher shared:

*...the requirements of the tests are very different than the teacher, the requirements of the teacher assessment, because the teacher assessment is much wider. The test only tests a selection of skills but the curriculum is much wider...Teacher C*

AwLs was seen as 'tick boxy', using either the ITAFs or curriculum objectives to assess against, as Teacher B said 'for every child for every area'. This supports the analysis of teachers needing to know the new rules AwLs. This was very time consuming for the

teachers, and encouraged test preparation strategies, such as TTTT and coaching on test technique;

*With SATS I feel that as much as it goes against a teachers' professional judgement, we are required in many ways to teach to the test in order to 'prove' that the children have made the progress and are at the level that we are saying that they are. Teacher B*

For Teacher B this mode of assessment was felt to be a requirement to 'prove' their judgments were trustworthy to SLT.

As previously mentioned, all the teachers expressed more confidence assessing the new curriculum once they had experienced it. Through experience and action their *habitus* was becoming more aligned with the *field* of primary assessment, and they feel more like a fish in water without feeling the weight of the water. The amount each teacher has fully bought into the *doxa* of primary assessment seemed different. The sense of each *illusio* in the field varied. For example; Teacher A expressed complete disagreement with the SATs, believing teacher SA, although formed on a rigid tracking system, is far more accurate. Teacher B disagreed with the changes, and questioned those who implemented them, expressing concern that they are damaging to the children's education and happiness.

*I do think [with] teacher assessment we've got a really good idea of where they are without having to have the added pressure of the SATs for the children. Teacher A*

*...if it was based on some fantastic research that shows this is going to really improve their learning and improve their performance than brilliant but...SATs are about accountability of schools rather than assessing children's actual progress... Teacher B*

However, Teacher C, the only Year 6 teacher interviewed, felt successful when following the new curriculum closely, aligning their teacher SA so it agreed with the SATs test results. They associated this with knowing the curriculum through experiencing it. Thus, Teacher C displayed a more obvious *illusio*,

*But if you get your head around it and have a really good understanding of the curriculum and how that fits into the scaled scores system, it makes sense and it becomes a lot easier and a lot clearer and it's not as scary...And if everybody can just get to a consistent point of understanding and we're all assessing fairly, it'll feel as easy as it used to. Teacher C*

Success for Teacher C came from following the new assessment rules. This *illusio*, buying into the game being played, allowed Teacher C to rationalise learning to now play the game in their best interests, as a teacher, but also as an individual with a job. Teacher C certainly expressed an awareness of how working with the system benefited their position, 'I just wish that we...weren't so heavily reliant on them [SATs results] because my performance management is all based on what they get in year six'. Performance management was recently linked to pay in 2014. Bourdieu refers to this as 'structural violence', where by there is a difference in power between social structures with potential impact on an individual's needs such as losing their job or being financially impacted. This influences certain actions, for example to take part in teaching practises which the teachers disagreed with in order to maintain their own reputation and position through their pupils SATs results. As year 6 SATs are used to inform league tables, this structural violence also relates to the school's position. The assessment and accountability system may explain this, the Year 2 teachers did not seem to buy into primary assessment systems to 'play the game' as seen from Teacher C. Adopting TTTT strategies and complying with SATs preparation may be motivated by a desire to protect the school and prevent external intervention by OfSTED. However, working with the system seems to be what is causing stress and unfairness for the children. The risks of not 'playing the game', and aligning with the demands of SATs without levels, could be a lower place in league tables, which may result in less parents 'choosing' to send their children, and result in smaller class sizes and therefore less money for the school, or become subject to OFSTED inspections, and potentially fall into special measures or forced academisation.

Surveillance

As mentioned briefly above, teachers were required to evidence their teacher SA judgments both for external and internal moderation. This is a theme that shows in both years 2 and 6 with teachers saying:

*I know what I'm doing for the next six weeks now in maths, just so I know I've got everything and I can gather evidence before SATs. Teacher A*

*I think there is even more emphasis on making sure that teachers can prove where the children are and have large amounts of evidence to back this up...they've [moderators] really stressed if you have not covered all those, if she has not got all of those things that are on the interims documents then they're not, you can't prove that they're working at that level...Teacher B*

*...as well as the objectives I also assess each final piece of writing against the interim standards so I build a bank of evidence throughout the year, which does inform my assessments...Teacher C*

Moderation of writing formed a large part of the discussion with the teachers around evidencing, teacher B felt that teachers' judgments were not trusted unless backed up with evidence. This is indicative of the 'examination' as described by Foucault; 'It is a normalizing gaze, a surveillance that makes it possible to qualify, to classify and to punish.' The truth of assessment is located in the ITAF objectives and requires examining as part of moderation. Teacher B's additionally commented that they left feedback on all children's' work, not to affect the children's learning, but in case they are monitored. It was not certain if or when student books would be examined by SLT, so teacher B engaged in this performative practice to regulate their actions based on the possibility of surveillance. This self-regulation is reflective of how Foucault applied the theory of the panopticon in our daily lives as a mode of control.

#### Summary of Pilot study themes

The teachers found the transition to AwLs challenging due to uncertainty, a lack of clarity and inconsistencies. This made it difficult for the teachers to adapt to AwLs as they were unable to learn the new rules. With a years' experience now AwLs the teachers all felt more confident, through experience of the new *field* of AwLs the teachers' *habitus* was adapting.

In doing so, this caused conflict with some other teachers as it encouraged the use of teaching practices they disagreed with. The need to know the expectations of AwLs seemed rooted in the use of SATs for school accountability and for measuring teacher performance. SATs were viewed as damaging to children for both KS1 and KS2, causing stress and pressure which, in some cases, were felt to contribute towards a poor result. The assessment of writing was an issue for both key stages because of the secure fit approach. It was felt that assessing writing this way prevented some judgments at the expected level because of spelling and handwriting. Creativity was also seen to suffer, an important aspect of writing to the teachers.

Assessment requirements in school had also added to teacher workload. AwLs required more evidencing and list ticking than they had experience with levels. The work in student's books now formed an additional purpose, it was not just a medium for students to interact with and aid in their learning, it was a tool to monitor, evidence and prove assessment judgments from the teachers.

An important conclusion from this analysis is the extent to which the teachers complied with teaching practices which caused the curriculum narrowing and pupil pressure they objected to. For example, to aim for the attainment target of 85% of KS2 students achieving expected in Maths, a year 6 teacher taught to aspects of the test, thus preventing student exposure to a broad and balanced curriculum. The accountability pressures for themselves and for the school may account for this compliance. It would be unreasonable to expect these teachers to ignore national education policy and for the school to not take part in mandatory testing, but teaching practices in preparing students for SATs can be mediated by schools through the school's own assessment policy. Resisting the pressure to play the game may be a more challenging route but it may well provide the education the teachers passionately wished to provide.

## APPENDIX 7 FULL STUDY REVISED INTERVIEW SCHEDULE

### Teacher Interview Schedule

#### Assessment without levels- First interview

- 1) Tell me about a bit about yourself, how long have you been teaching etc...?
- 2) We're going to first talk about assessing with and without levels for reading writing and maths. Did you assess with levels?
- 3) What were your experiences of assessing maths with levels?

And now without levels?

- 4) What were your experiences of assessing reading with levels?

And now without levels?

- 5) What were your experiences of assessing writing with levels?

And now without levels?

- 6) How do you think the two ways of assessing compare? With levels or without?
- 7) What are your day to day, formative, assessment practices without levels?
- 8) What are your end of term, summative, assessment practices without levels?
- 9) Has assessing without levels affected how you assess noncore subjects?
- 10) Are there any other things you'd like to tell me about assessing that we haven't talked about?

#### Assessment without levels- Second interview

- 1) Thinking about the assessment data that you had this year which do you feel represents the assessment of your class better, teacher assessment or test?
- 2) Do you think the SATs tests and teacher assessment work together? How would you describe the relationship that they have together?
- 3) Are there areas of the curriculum that you feel are best suited to being assessed by a test or being assessed by your teacher assessment?
- 4) With the removal of the mental maths test from last year's and this year's SATs, what are your opinions on that with then introducing the arithmetic test instead? (Year 6 only)
- 5) With the government interim framework, how would you describe the impact it's had on your teacher assessments and in preparing the children for the tests?
- 6) How did you find the moderation process this year compared to last year?
- 7) Thinking about the interim framework being brought in and additional information or resources, is there anything that you feel could have been introduced either at the beginning of the changes in assessment and curriculum or since or even now

that would help you know how to better prepare your children for SATs or to help you with your teacher assessment?

- 8) Is there anything that we haven't talked about either in the first interview that you thought about sense or this interview today that you'd like to add about assessment in general or to do with the SATs or to do with your own teacher assessment or the new curriculum?

### Head Teacher semi structured interview-

Start with their back ground in teaching.

How have you found the transition towards assessment without levels?

What implications have there been in the changes in assessment on you as head teacher?

How did you implement the changes through the school? Policy and practice.

What effects has assessing without levels had on assessment practices in the school?

Have you noticed any differences in how your staff do things since the changes?

What effects has assessing without levels had on children's learning in the school?

Do you know if the summative assessment data you collect for the LA makes a difference to the children's learning? If so how?

What is the terminology you use for assessment without levels in your school to the staff?

What is the terminology you use for assessment without levels in your school to the children?

What measures have you used to inform parents on the changes in policy in assessment? Do you think they understand it?

How have your parents responded to the changes?

## APPENDIX 8 INTERVIEW SETTING SCHOOL DEMOGRAPHICS AND SATS PERFORMANCE

Table 0-2 Pupil demographic data from school research settings.

	<b>Persistent Absence rate</b>	<b>Overall absence rate</b>	<b>Pupils with SEN support</b>	<b>Pupils with English as an additional language</b>	<b>Pupils eligible for free schools meals anytime in the past 6yrs</b>
<i>Pilot study school</i>	22.7%	5.5%	33.1%	27.5%	78.1%
<i>Setting 1</i>	7%	3.4%	7.4%	2.4%	16.7%
<i>Setting 2</i>	3.9%	2.9%	10.7%	4.7%	16.8%
<b>England Primary average</b>	<b>8.3%</b>	<b>4%</b>	<b>12.2%</b>	<b>21.3%</b>	<b>24.3%</b>

(Data gathered from '<https://www.compare-school-performance.service.gov.uk/find-a-school-in-england>')

### Interview Data Settings Demographics

Demographic data were gathered about each school setting by referring to 'compare school performance service' provided by the government for 2017, the year the data were collected. School OFSTED reports and the schools themselves provided further information, for example, if they are two or one form entry (how many classes there are for each year), presence of a nursery. To maintain confidentiality the OFSTED reports will not be included in appendices or reference lists to ensure the anonymity of the participants of the research project. SATs data from the cohort 2016/2017 has also been included with a comparison of the local authority average and national average. Pupil demographic data is based on the percentage of pupils within the school cohort for the year, the pupils eligible for free school meals is calculated to include any child who has been eligible for free school meals in the past 6 years as well as that academic year.

Where KS2 SATs data is discussed, it will be in terms of percentage children reaching the expected standard for the test and the average scaled score. A scaled score of 100 is the minimum score to be awarded, the expected standard on the assessments. Anything below, from 99 to 80, have not met the expected standard of the assessment. Any scaled score of 120 or above has reached a higher than expected standard, referred to also as greater

depth. Scaled scores are not raw test or assessment scores but are, according to the DfE guidance, adjusted raw scores to take into account yearly variation in the difficulty of each SATs test paper. The raw score needed each year to gain 100, the expected standard, or 120, higher than expected standard, will vary. This is stated by the DfE to ensure fair comparison across different cohort scores.

Table 0-3 Pilot study school pupil demographic information.

	<b>Persistent Absence rate</b>	<b>Overall absence rate</b>	<b>Pupils with SEN support</b>	<b>Pupils with English as an additional language</b>	<b>Pupils eligible for free schools meals anytime in the past 6yrs</b>
<i>Pilot study school</i>	22.7%	5.5%	33.1%	27.5%	78.1%
<b>England Primary average</b>	<b>8.3%</b>	<b>4%</b>	<b>12.2%</b>	<b>21.3%</b>	<b>24.3%</b>

(Data gathered from ‘<https://www.compare-school-performance.service.gov.uk/find-a-school-in-england>’)

The pilot study school was a local authority maintained Community Primary school and nursery, within Birmingham. Table 0-3 above shows the demographic information for the pupils in the school. This setting was located in a deprived area, with 78.1% of pupils eligible for free school meals either that year or in the previous 6 years. This is well above the national average of 24.3%. There was also 22.7% of pupils with persistent absence (missing 10% or more of mornings or afternoons or a total absence of 10% of the school year) compared to the national average of 8.3%. The school had an above average number of children requiring support for SEN (special educational needs) at 33.1%, whereas the national primary average is 10.4%, and just above average for pupils where English was an additional language, 27.5% compared to an average of 21.3%. The school catered for Nursery up to year 6, where reception to year 3 were two form entry, and years 4-6 were one form entry.

(Data gathered from ‘<https://www.compare-school-performance.service.gov.uk/find-a-school-in-england>’)

Student performance data (Year 6 SATs)

60% of pupils achieved 'expected' in reading, writing and maths assessments, this is 3% higher than the LA average and 1% lower than the national average. 3% of the cohort achieved 'higher' standard in all three assessments, this is 3% lower than the LA (local authority) average and 6% lower than the national average. The average scaled score for the cohort for reading met the LA average, and was 1 score lower than the national average. The scaled score average for maths matched the LA and national average of 104.

(Data gathered from '<https://www.compare-school-performance.service.gov.uk/find-a-school-in-england>')

Table 0-4 Setting 1 school pupil demographic information

	<b>Persistent Absence rate</b>	<b>Overall absence rate</b>	<b>Pupils with SEN support</b>	<b>Pupils with English as an additional language</b>	<b>Pupils eligible for free schools meals anytime in the past 6yrs</b>
<i>Setting 1</i>	7%	3.4%	7.4%	2.4%	16.7%
<b>England Primary average</b>	<b>8.3%</b>	<b>4%</b>	<b>12.2%</b>	<b>21.3%</b>	<b>24.3%</b>

(Data gathered from '<https://www.compare-school-performance.service.gov.uk/find-a-school-in-england>')

Table 0-4 displays below average pupils requiring SEN support and well below average pupils speaking English as an additional language. 16.7% of pupils were or had been eligible for free school meals in the last 6 years compared to 24.3% national average. The majority of pupils came from a White British background with those who did not coming from a variety of other ethnic backgrounds (OFSTED). School absence figures were just below national averages at 3.4% and 7% for absences below 10% of school time and persistent absence above 10% of schools time respectively.

Student performance data (Year 6 SATs)

In 2017 53% of the year 6 cohort met the ‘expected’ standard for reading, writing and maths in the SATs assessments. This is below the national average of 61% but matches the local authority. No pupil achieved the ‘higher’ standard in reading, writing and maths which is below both the national and local authority average, 5% and 9% respectively. In reading, the cohort averaged a scaled score of 103 which is above the LA average of 102 but below the national average of 104 (where expected is a scaled score of 100). The average maths scaled score of the cohort was 102 which was below the LA average of 103 and the national average of 104.

Table 0-5 Setting 2 school pupil demographic information.

	<b>Persistent Absence rate</b>	<b>Overall absence rate</b>	<b>Pupils with SEN support</b>	<b>Pupils with English as an additional language</b>	<b>Pupils eligible for free schools meals anytime in the past 6yrs</b>
<i>Setting 2</i>	3.9%	2.9%	10.7%	4.7%	16.8%
<b>England Primary average</b>	<b>8.3%</b>	<b>4%</b>	<b>12.2%</b>	<b>21.3%</b>	<b>24.3%</b>

(Data gathered from ‘<https://www.compare-school-performance.service.gov.uk/find-a-school-in-england>’)

Table 0-5 shows the number of pupils eligible for free school meals over a 6 year period is below average at 16.8% and SEN support pupils 10.7% compared to a national average of 12.2%. The majority of pupils came from a White British background, with 4.7% speaking English as an additional language. Persistent absences were 3.9%, well below the national average.

#### Student performance data (Year 6 SATs)

In reading, writing and maths 77% of the cohort met the ‘expected’ standard, this is 20% higher than the LA average and 16% above the national average. 15% of pupils additionally achieved the ‘higher’ standard in all three compared to 6% average in the LA and 9% average nationally. In reading the average scaled score was 108 where 100 was an ‘expected’ standard, this higher than the LA average of 103 and the national average of 104.

In maths pupils averaged a scaled score of 107 which again is above the LA and national average of 104.

## FS2Y6 Teacher G

### Interview 1

Interviewer: Just to start off with, tell me a bit about yourself, how long you've been teaching, what years you've taught in.

Interviewee: Yep so, um, this is my third year teaching now. I came in teaching year six as an NQT, umm, and I've been in year six ever since. When I trained I spent a term in year six, a term in year two and then a term in year 5. So, my experience of the year groups is relatively limited.

Year six is, sort of, all I've ever known really as a teacher. I came in after levels were removed they were sort of midway in the process of removing them during my training year so not much was really mentioned about levels because they knew schools wouldn't be adopting it any more. So, although I know about levels because levels were used when I was in year six that sort of gap between 2000 up to 2013 I know very little about levels and how they've changed and we've been assessing without levels ever since I've been at the school.

Interviewer: So talking about assessing without levels then, that's all you've known?

Interviewee: Yes

Interviewer: How do you find assessing without levels in maths?

Interviewee: not too bad. We use SPTO (student pupil tracker online) for our tracking and it's quite nice because it's broken down by objective and you can mention whether children are far away from the expected, nearly met the expected, or at the expected or at greater depth, so we use that. Last year we used, I forgot the name of it now, depth of learning tracker but the school moved over to S PTO this year, so they are still getting to know that system because it's only, we've only used it since September.

Interviewer: And how do you find assessing reading?

Interviewee: One of the trickiest I would say because with maths it can be quite clear whether a child has grasped the concept through questioning and things like that. Whereas, reading there's lots of elements to it and it can be quite tricky to establish whether they've got it or whether they've got it because the person sat next to them has got it. Reading I find one of the trickiest ones but again we track through SPTO for that.

Interviewer: And what about writing, how to find assessing writing without levels?

Interviewee: Writing, so we use the interim assessment framework for our assessing of writing so we've got the checklist. I don't think that the checklist was produced by the DFE but different teachers, year six teachers, in different schools have produced it and we will assess a piece of writing according to that checklist. So last year we were moderated, so we had a checklist for every single child based on all the writing they've done in their red books and we tick off each element and which paragraph it was found in, as well, just so when we came to be moderated we were able to say 'oh there's a modal verb in this piece of writing in this paragraph' so it was quite time-consuming but we were able to make sure that when we said a child was at the expected or working towards that, they definitely were.

Interviewer: I know you said you don't have much experience or any experience of assessing with levels but thinking of the time where there wasn't anything in place how do you think now assessing without levels compares to that?

Interviewee: It's hard for me to say because I've never assessed using levels but from what I gather there was a general consensus that it needed to change. I'm not sure exactly why but I think they'd sort of, schools had lost their way over the purpose of assessing with levels and I think children were beginning to label themselves with a certain level and I think the government wanted to move away from that but I'm not able to compare because I've never assessed using levels myself.

Interviewer: Do you tell the children in your years where there at?

Interviewee: No.

Interviewer: Thinking of formative assessment, day-to-day, what are your formative assessment practices?

Interviewee: So observation through the classroom, through what children are sharing in the classroom and the work they are producing, questioning to check understanding or to poke and prod and sort of check how deep their understanding is and marking of classwork mainly.

Interviewer: and when it comes to summative assessment is that termly, half termly how does that take shape?

Interviewee: We did and a mock SATs in October which was summative assessment and we will do another one in February but we will do small little assessments based on white rose materials and things like that throughout the year as well. Often at the end of, sort of, a mathematical concept rather than at the end of the half term or a term just when we've covered a topic within maths.

Interviewer: How do you assess non-core subjects?

Interviewee: Mainly through observation and marking and classwork. We don't do any formal assessment through the learning adventure or for music or PE or anything like that. We do a core task in PE at the beginning of the unit and at the end of the unit which assesses, sort, of the development of skills and how far they've come throughout the unit but we don't formally assess foundation subjects.

Interviewer: How do you assess Science?

Interviewee: So last year, I imagine we'll do the same this year, last year at the end of, the end of year six we gave a sort of past paper within science just to check where they were and then we use that to feed into our teacher assessment of science because along with the writing it's teacher assessment at the end of, at the end of the year.

Interviewer: And that's more of a summative assessment?

Interviewee: Yes, yes.

## APPENDIX 10 FULL STUDY CODING PROCESS NOTES

Informed by pilot systematic approach

### Trawl 1- Teacher interviews

For each interview- starting with all the teacher interviews- they were read, then reread with codes drawn out via themes. When one interview was coded it was discussed with my lead supervisor as a check point that the interpretations of the interview content reflected the theme coded for. The rest of the teacher interviews were then read, reread and coded. If an existing code applied from the first or subsequent interviews it was selected, if a new code was required it was created and added to the coding list. Annotations were also made in relation to certain content that didn't quite fit a code or that raised a question or area to explore. Codes were organised and reorganised as the coding progressed. Parent and child themes and relationships between themes were created intuitively as coding continued through the interviews.

### Trawl 2- Teacher Interviews

Once all teacher interviews were coded the codes were analysed and renamed where appropriate to better fit the theme. The interviews were then reread and recoded if needed. This was for two reasons. The first to check the codes still reflect the theme coded to. And second to code aspects of the interviews that were identified as relevant in later interviews as the code book expanded with exposure and immersion into the interviews. These themes may have been present in earlier interviews in the process but not coded yet as the content was becoming more familiar again.

### Trawl 3- Teacher interviews

A third trawl was then carried out. Certain aspects of the coding require different thinking process too, this third trawl, like with the pilot study, focused on consideration of teachers' positions and analysis of their agency, awareness of agency and ontological perspective.

### Trawl 4- Teacher interviews

Re-reading and confirming an agreement with the codes. Some have been reworded or added but on the whole I agree with the coding previously carried out. Questions about

assessing what we value or valuing what we measure are coming up. Also the consideration of the confusion linking to learning theory in my paradigm writing, there's no common vocabulary between schools or schools and the gov't so understand is lacking across the board. Pragmatically a huge interruption has occurred causing a problem that needs to be solved- triggering an inquiry. If greater depth is so important what is the best way to assess it? A test? Or using teacher assessment? IS greater depth just seen as independent?

Trawl 1 SLT (senior leadership team) interviews

As the perspective of these interviews was different because the teachers were in a position of power over policy and practice in the school, they were coded first from fresh without the teacher interview codes. Again they were read and reread then coded as mentioned above.

Trawl 2

Same as above for teachers.

Trawl 3

Same as above for teachers.

Trawl 4

This trawl was to look at any codes which matched the codes from the teacher interviews to use as a comparison for how themes are shared and related between teachers and SLT.

## APPENDIX 11 MY PHD JOURNEY

<p>Compare Teacher Assessment to SATs Assessments to ensure consistency</p> <p>Read into the literature– you shouldn't compare TA to SATs they are importantly different (Halren, 2005)</p> <p>Take a more qualitative turn</p>
<p>Change to qualitative perspective including teachers' voices and views as central</p> <p>Read Bew (2011) and DES (1988A) Now interested in curriculum and how it fuels assessment and what SATs and TA assess from the curriculum</p>
<p>Change to a comparison of objectives assessed at the end of KS1 and KS2 with TA and SATs .</p> <p>Triangulate with results of TA and SATs with teachers views and experience's central to comparisons.</p>
<p>Test construct now added as a focus with international comparisons</p>
<p>Questionnaire added to survey teachers in schools included</p> <p>Questionnaire added to survey third year students as the next generation of teachers</p>
<p>Inclusion of Head Teacher views from a policy and management perspective</p>
<p>Pilot study so it could all change now anyway!</p> <p>Pilot study will be included in the main study findings</p>
<p>Survey will be open in primary teachers. This will give a wider perspective on the changes which the interview qualitative data will give perspective , context and detail.</p>
<p>Interview extended to Prof Paul Black to get a view from a researcher who studied and experienced the last curriculum change.</p>

Re write of research questions to reflect the changing in thinking around assessment while collecting the data.

What is the impact from the curriculum and assessment change on teacher, classroom and assessment practice?

What are the Policy into Practice implications of the curriculum and assessment changes?

Reading has expanded into:

- Teachers' experience of stress as a result of high stakes testing.
- Measurement- what is measurement in education?
- Validity theory and the history of it, mostly American perspective.

Pilot study write up using Bourdieu social theory raised questions of teacher agency and choice.

Started reading into critical realism as a paradigm- struggled with this, had a momentary existential crisis and reviewed all my reading on positionality and paradigm setting on interpretivism, radical constructivism and pragmatism as a theoretical framework.

Reading on Agency lead me to Biesta and Priestley and their work using a ecological theory of agency.

Framed part of my analysis around teacher agency and how to categorise agency in the teachers I interviewed.

Policy chapter researched, finally could place my Neo-liberal writing by viewing policy through this lens.

Struggled with the placing of my quantitative research in my methodology rewrite. Settling after consideration with it being viewed as a separate data set which contain an opportunistic sample of the teachers from my interview settings.

Reference to and the relationship of learning theory and assessment needs to be part of my literature review.