

University of Nottingham
UK | CHINA | MALAYSIA

Principals' and Social Studies Teachers' Assessment Literacy and the Enactment of Indonesian Assessment Reform in Diverse School Contexts

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

August 2025

Yasser Awaluddin

20399594

Supervisors:

Professor Kay Fuller and Dr Michael Adkins

School of Education

University of Nottingham

Acknowledgement

I am deeply grateful to the Almighty God for granting me the strength to complete my doctoral thesis. I also extend my heartfelt thanks to those who have supported me throughout my PhD journey.

First and foremost, I would like to express my profound gratitude to my supervisors, Professor Kay Fuller and Dr Michael Adkins, for their invaluable guidance, timely and insightful feedback and encouragement at each critical stage of the thesis writing. Although they became my supervisors only after the second half of my PhD journey, they quickly grasped the essence of my research and provided precise, effective, and enlightening directions. The feedback they offered, both in writing and during supervisory meetings, challenged me to think more critically and rigorously. Their input helped sharpen the focus of my work and ultimately transformed it into a doctoral thesis.

I would also like to extend my gratitude to my previous supervisors, Dr Vicente Reyes Jr. and Dr Jenny Elliot, for their excellent guidance during the first half of my PhD.

I owe an immense debt of gratitude to the Indonesian Endowment Fund for Education (LPDP) for generously funding my PhD study. My sincere thanks also go to the Indonesia Ministry of Education for granting me permission and support to pursue this doctoral degree. I would like to thank my research survey respondents and especially the interview participants who have contributed their time to this study and generously given me access to their experiences, thoughts, and opinions, shaping the content of the thesis.

I dedicate this thesis to my late mother, Nurasih, and my father, Abdul Malik, for their unwavering prayers and support at every step in my pursuit of knowledge and self-improvement. I would like to express my heartfelt appreciation to my beloved wife, Hasti Rahmaningtyas, for her love, relentless support, constant encouragement, and steadfast belief throughout this journey, even as she pursued

her own PhD. I am also deeply grateful to my dearest daughter, Nasywaa Azizah, for her patience and resilience when her parents pursued their study abroad.

Finally, this long PhD journey was brightened by the warm and enjoyable interactions with my PhD researcher colleagues at the school of education, especially my office mates, Bauyrzhan Kaziyev and Yunus Saracoglu. Our casual discussions and shared moments, during lunch times or outside the campus, brought joy and relief during the demanding phases of our PhD.

Abstract

The purpose of this study is to examine the assessment literacy of principals and teachers, especially social studies teachers, in Indonesia. This investigation is crucial as assessment literacy represents a key factor, determining the success of the ongoing Indonesian assessment reform, under the new national education policy titled *Merdeka Belajar* (Emancipated Learning). Furthermore, the scarcity of empirical studies focusing specifically on the assessment of literacy of social studies teachers and school leaders, provides strong justification of this research.

This study adopts contemporary concept of assessment literacy, emphasising the context of implementation as a key element in evaluating principals' and teachers' assessment literacy. It equally values the technical knowledge and skills of assessment as the core professional capabilities. This assessment concept is particularly relevant to the diverse social, cultural, and economic conditions among regions, spanning from the eastern to the western regions of Indonesia.

Grounded in pragmatic philosophy that emphasises problem-solving, allowing the use of diverse methods best address research questions, this study employed a mixed-method approach, with explanatory sequential design consisting of two phases: (1) a quantitative phase with survey method, followed by (2) a qualitative phase with interview method. Three theoretical frameworks guided the data collection, analysis, and interpretation: social constructivism, theory of practice, and multiplicity of enactment.

The survey revealed moderate levels of assessment literacy among school principals and social studies teachers, with three factors having direct influence: in-service trainings, assessment-related experiences, and assessment coursework in higher education. Meanwhile, local support for principals and teachers had indirect effects on assessment literacy, mediated through in-service trainings and experiences. Interview data analysis demonstrates that principals and teachers had good understanding of the key components of the assessment reform, could identify various reform-related challenges and obstacles within their schools, and employed diverse strategies -including adaptive and improvisational practices- tailored to their specific school contexts. Additionally, paired interview findings highlighted a reciprocal relationship between principals' and teachers' assessment literacy, suggesting that their understanding and enactment of assessment reform mutually influence and reinforce one another.

Overall, this study concludes that Indonesian principals and social studies teachers possess adequate foundational assessment literacy to implement the assessment reform reasonably well in their respective contexts, though further improvement is necessary. A critical finding is that due to the diverse challenges and constraints the school encountered, assessment reform enactment varied significantly across regions and even between schools. This study introduces the term "adapted and

differential enactment of assessment reform" to characterise this phenomenon. This study suggests that such variations should be interpreted as manifestations of principals' and teachers' assessment literacy, rather than as forms of fidelity or infidelity of education policy implementations.

Table of Contents

Acknowledgement ..... **i**

Abstract ..... **iii**

Table of Contents ..... **v**

List of Tables ..... **x**

List of Figures ..... **xi**

Chapter one: Introduction **1**

1. Introduction 1
2. Context of the research 4
3. The education system in Indonesia 5
4. The education reform in Indonesia 6
   - 4.1 The *Emancipated Learning* policy and the transformation of assessment policy and practice in Indonesia 6
   - 4.2 Reflecting the *Emancipated Learning* policy through Emancipatory Curriculum 8
   - 4.3 The enactment process of the Emancipatory Curriculum 9
5. Research Problem 10
   - 5.1 Emancipatory Curriculum and challenges for principals and teachers 10
   - 5.2 Changes in the assessment policy and related consequences for principals and teachers 12
6. Justification of the study 15
7. Research objectives and questions 17
8. Structure of the thesis 20
9. Summary 21

Chapter two: Review of literature **23**

1. Introduction 23
2. Concepts of assessment literacy 26
   - 2.1 Classical concept of teacher assessment literacy 27
   - 2.2 Classical concept of principal assessment literacy 32

2.3 Contemporary concept of assessment literacy 38
3. Empirical literature on assessment literacy levels 43
4. Empirical literature on factors affecting assessment literacy 46
4.1 Assessment-related experiences 46
4.2 Assessment-related courses in higher education 49
4.3 Assessment-related in-service professional development 52
4.4 Local support from colleagues 55
5. Hypothesised model of the relationship between assessment literacy and the predictors 57
6. Empirical literature on assessment policy enactment 59
7. Theoretical framework 65
7.1 Social Constructivism 66
7.2 Theory of Practice 68
7.3 Multiplicity of Enactment 72
8. Summary 74

Chapter three: Methodology and methods 78
1. Introduction 78
2. The researcher's positionality 78
3. Research paradigm of the study 79
4. Research design of the study 82
5. The preparatory stage of the research 85
5.1 Identification of assessment literacy key dimensions 85
5.2 Pilot study 90
6. Research Methods 92
6.1 Phase one 92
6.1.1 Survey 93
6.1.2 Survey questionnaire 94
6.1.3 Population and sample 97
6.1.4 Survey data analysis 99
6.2 Phase two 104
6.2.1 Semi-structured interview 104
6.2.2 Paired interview 106
6.2.3 Interview protocol 106

6.2.4 Interview participants 108
6.2.5 Interview data analysis 110
7. Ethical Issues 115
8. Summary 116

Chapter four: Quantitative findings 118
1. Introduction 118
2. School principals' and social studies teachers' assessment knowledge and skills 119
2.1 Principals' assessment knowledge and skills 119
2.2 Social studies teachers' assessment knowledge and skills 123
3. Factors affecting teachers' and principals' assessment knowledge and skills... 127
3.1 Measurement (CFA) model of social studies teachers' assessment knowledge and skills 127
3.2 Reliability and validity of social studies teacher assessment literacy measures 133
3.3 Structural model of social studies teachers' assessment literacy 136
3.4 Mediation/indirect effect analysis of TAL structural model 141
3.5 Measurement (CFA) model of principals' assessment literacy 143
3.6 Reliability and validity of school principal assessment literacy measure 148
3.7 Structural model of school principals' assessment literacy 150
3.8 Mediation/indirect effect analysis in PAL structural model 153
4. Summary of phase one findings 156

Chapter five: Qualitative findings 158
1. Introduction 158
2. The emerging codes and themes 160
3. Themes concerning principal assessment literacy and the enactment of the new assessment policy 167
3.1 Theme one: The features of the new assessment policy 167
3.2 Theme two: Challenges and potential barriers at local contexts 170
3.3 Theme three: Strategies for assessment reform enactment 174
3.4 Theme four: Factors affecting PAL and the enactment of assessment reform 178
3.5 Theme five: Self-perceived assessment literacy 184

4. Themes concerning social studies teacher assessment literacy and the enactment of the new assessment policy 185
4.1 Theme one: The features of the new assessment policy 185
4.2 Theme two: Challenges and potential barriers at local contexts 188
4.3 Theme three: Strategies for assessment reform enactment..... 192
4.4 Theme four: Factors affecting TAL and the enactment of assessment reform 196
4.5 Theme five: Self-perceived Assessment Literacy..... 201
5. Paired interview analysis results..... 203
5.1 Paired interview one: School A..... 203
5.2 Paired interview two: School B 205
6. Summary of findings of phase two 206

Chapter six: Discussion 208
1. Introduction 208
2. Principals' and teachers' assessment knowledge and skills levels. 209
2.1 Principals' assessment knowledge and skills. 209
2.2 Teachers' assessment knowledge and skills. 212
3. Factors affecting the principals' and teachers' assessment knowledge and skills 215
3.1 Factors affecting principals' assessment knowledge and skills 216
3.2 Factors affecting teachers' assessment knowledge and skills..... 223
4. The enactment of assessment reform 231
4.1 Features of the new assessment policy..... 232
4.2 Challenges of new policy enactment..... 235
4.3 Strategies for successful enactment..... 242
4.4 Self-perceived assessment literacy..... 250
5. The Relationship between Principal Assessment Literacy and Teacher Assessment Literacy..... 253
5.1 Similarities between school A and B..... 254
5.2 Differences between school A and B 255
5.3 Summary of paired interviews regarding the relationship between principal and teacher assessment literacy 256
6. Summary of discussion 257

Chapter Seven: Conclusion **260**
1. Introduction 260
2. Summary of the research 260
3. Answering research questions 262
4. Contributions of the study to knowledge 272
4.1 Theoretical contribution to knowledge 273
4.2 Methodological contribution to knowledge 276
4.3 Empirical contribution to knowledge 276
5. Implications of The Study 277
5.1 Implications for practice 278
5.2 Implications for policy 279
6. Limitation of the study and suggestions for future research 280
7. Concluding Remark 281

References **283**
Appendix **295**
Appendix 1: Survey questionnaire for teachers **295**
Appendix 2: Survey questionnaire for principal **303**
Appendix 3: Interview protocol for teachers **311**
Appendix 4: Interview protocol for principals **313**
Appendix 5: Research ethics approval **315**
Appendix 6: Formal request letter to local education authorities for accessing research site and participants **316**
Appendix 7: Information sheet for survey participants **318**
Appendix 8: Information sheet for interview participants **321**
Appendix 9: Interview participant consent form **324**
Appendix 10: Codes from first cycle coding process **326**

List of Tables

Table 1 Guiding Framework for teacher assessment literacy inventory 87
Table 2 Guiding framework for principal assessment literacy inventory 88
Table 3 Items removed in PALI and TALI 91
Table 4 Items with simplified sentences in PALI and TALI 91
Table 5 Dimensions of principal assessment knowledge and skills 96
Table 6 Dimensions of teacher assessment knowledge and skills 96
Table 7 Teacher sample 99
Table 8 Principal sample 99
Table 9 Cut off criteria for model goodness of fit 103
Table 10 Teacher interview participants 109
Table 11 Principal interview participants 109
Table 12 Principals' assessment knowledge and skills scores 120
Table 13 Principals' scores on each dimension of assessment knowledge and skills 121
Table 14 PALI item difficulty levels 121
Table 15 Teachers' assessment knowledge and skills scores 123
Table 16 Teachers' scores on each dimension of assessment knowledge and skills 124
Table 17 TALI item difficulty levels 125
Table 18 TAL CFA model goodness of fit indices 132
Table 19 TAL item factor loadings 132
Table 20 CR, CV, and DV of TAL and predictors' scale 134
Table 21 Examination of DV TAL using Fornell and Larcker criterion 136
Table 22 Examination of DV TAL using HTMT ratio 136
Table 23 TAL model goodness of fit indices 137
Table 24 Standardised estimates of TAL model structural paths 138
Table 25 Results of hypothesis testing of TAL model 140
Table 26 Indirect effect analysis of TAL model 141
Table 27 Results of hypothesis testing of TAL indirect analysis 143
Table 28 Summary of direct and indirect effect of TAL model 143
Table 29 PAL CFA model goodness of fit indices 147
Table 30 Factor loadings of PAL CFA model 147
Table 31 CR, CV, and DV of PAL and predictors scales 149
Table 32 DV PAL model using Fornell and Larcker criterion 150
Table 33. DV PAL model using HTMT ratio 150
Table 34. PAL structural model goodness of fit criteria 151
Table 35 Standardised estimates of PAL model structural paths 151
Table 36. Results of Hypothesis testing of PALI model 153
Table 37 Indirect effect analysis of PAL model 154
Table 38 Results of hypothesis testing of PAL indirect effect analysis 154
Table 39. Summary of direct effects, indirect effect, and mediation of PAL model . 155
Table 40 Themes and codes from the interviews 162
Table 41. Codes of theme 1 (principal) 167
Table 42 The codes of theme 2 (principals) 171

Table 43. The codes of theme 3 (principal) 174
Table 44 The codes of theme 4 (principals) 179
Table 45 The codes of theme 5 (principals) 184
Table 46. The codes of theme 1 (teacher) 186
Table 47 The codes of theme 2 (teachers) 189
Table 48 The codes of theme 3 (teachers) 193
Table 49. The codes of theme 4 (teacher) 196
Table 50. The codes of theme 5 (teacher) 201

List of Figures

Figure 1. Hypothesised causal relationship between assessment literacy and four predictors 59
Figure 2. Relationship of methods an RQ 92
Figure 3. Principal assessment knowledge and score distribution 120
Figure 4. Configuration of PALI item difficulty levels 123
Figure 5. Teacher assessment knowledge and skills score distribution 124
Figure 6. Configuration of TALI item difficulty levels 126
Figure 7. TAL Measurement (CFA) model before item parcelling 130
Figure 8. TAL Measurement (CFA) model after item parcelling 131
Figure 9. TAL model with std. regression weights 139
Figure 10 PAL measurement (CFA) model before item parcelling 145
Figure 11 PAL measurement (CFA) model after item parcelling 146
Figure 12 Standardized estimates of PAL SEM model 152
Figure 13. Relationship of Themes and statistical results 165
Figure 15. Conceptual framework of understanding principal and teacher assessment literacy 275

Chapter one: Introduction

1. Introduction

Recent education policy reform in Indonesia has marked a deliberate shift away from a highly standardised paradigm in schooling. The central idea of this new policy, called *Merdeka Belajar* (translated here as *Emancipated Learning* and used throughout this thesis), emphasises the need to provide students with greater flexibility in their learning experiences in order to achieve their maximum potential. To realise this vision, schools and teachers are granted increased authority and autonomy to design and administer instructional programme that is responsive to students' diverse learning needs (Utomo, 2024). This reorientation toward flexibility and autonomy in organising student learning, inevitably reshapes how student learning progress and achievement is assessed and evaluated.

The transformation in educational policy which has increased the complexity of assessment practices has been seen across many countries for more than two decades (DeLuca et al., 2019). In Indonesia, under the new policy, the national government's role in regulating student assessment, previously characterised by rigid national standards and high-stakes examinations, have been significantly reduced. As a result, schools and, particularly teachers, now have greater autonomy and flexibility in designing, conducting and interpreting student assessment. They are empowered to make critical decisions regarding students' mastery of competencies, progression to next grade level, and graduation eligibility at both primary and secondary levels.

Given this increased autonomy, principals' and teachers' sound and competent assessment practices have become pivotal to the success of the assessment reform. Sound and competent assessment practices grounded in assessment literacy are essential to ensure that assessment results are valid, reliable, and conducive to inform educational decision making. This study is therefore concerned with exploring the assessment literacy of school principals and teachers, particularly in relation to the ongoing assessment reform under a new curriculum and education

policy. Assessment literacy, as the core construct of this study, is broadly defined in this introductory chapter and explored in more detail in the literature review chapter (Chapter Two).

Popham (2011) defines assessment literacy as the knowledge and skills of assessment that teachers need to possess to measure and promote student learning. Stiggins (2001) elaborates that assessment literacy comprises of two key competencies, "first is the ability to gather dependable and quality information about student achievement; second is the ability to use that information effectively to maximise student achievement" (p.20). For school leaders, assessment literacy is part of instructional leadership with which they can support the improvement of teaching and learning (Stiggins & Duke, 2008) and an important competency required for effective school practices (Stiggins, 2001). Furthermore, it plays a critical role for a principal in influencing educational decision-making (Pfeiffer-Hoens, 2017).

More recent attention to the idea of assessment literacy have expanded the conception, framing it as a contextually negotiated and multifaceted professional and social competence and practices (DeLuca et al., 2019; Willis et al., 2013). This contemporary view highlights that both an education professional's knowledge and skills in assessment and the context where the assessment is practiced shape the educator's assessment literacy. The current study adopted this expanded conceptualisation in examining the principals' and teachers' assessment literacy, as the sociocultural context of Indonesia often becomes the challenges when it comes to the implementation of educational policies (Clifton, 2023). Chang et al. (2014) in their research report to the World Bank referred to Indonesia as one of the largest and most complex education systems in the world.

The significant disparities among regions across the eastern to the western parts of Indonesia, in terms of socio-cultural characteristics, economic welfare, access to educational resources, and overall education quality (Azzizah, 2015; Dharmawan & Suryadarma, 2021; Halim, 2020), make it essential to investigate the enactment of the new assessment reform policy by the school principals and teachers across

different contexts. This contextual analysis is key to understanding their assessment literacy within the modern framework. As such, this research is timely and highly topical in the Indonesian context.

This study employed three theoretical frameworks, namely social constructivism theory developed by Vygotsky (Adams, 2006), theory of practice (Bourdieu, 1977) and multiplicity of enactment theory (Buxton et al., 2015) to interpret the principals' and teachers' assessment literacy from the lens of contemporary conception (DeLuca et al., 2019; Willis et al., 2013) through the enactment of the assessment reform in their schools. These theories and their relevance with this study are discussed in the next chapter.

I have been working as a government official in a training centre for (in-service) teachers and other education personnel under the Indonesian Ministry of Education (MoE) for nearly twenty years when my PhD journey began in October 2021. Following the completion of my master's degree in educational research, assessment and evaluation in 2010 and teacher trainer programme in 2012, I was assigned a role as a teacher trainer. My responsibility centres around teaching and developing training materials concerning the topics of educational assessment and evaluation and school leadership. Further details of my positionality and how it affects the study is provided in Chapter Three. Given my positionality, this study was driven by the desire to support the successful implementation of the assessment reform in alignment with *Emancipated Learning* policy. Therefore, the primary purpose of this study is to contribute to the understanding of assessment literacy of school principals and teachers in Indonesia, recognising their critical contribution to the success of ongoing educational policy reform in Indonesia.

This introductory chapter outlines the topic and focusses of the research followed by the research context. It then describes the *Emancipated Learning* policy, the new curriculum called Emancipatory Curriculum (*Kurikulum Merdeka*), and the assessment reform. Subsequently, this chapter identifies the problems driving the study, followed by the research objectives and questions. Before summarising, this chapter outlines the structure of the thesis.

2. Context of the research

Indonesia is an archipelago country located in Southeast Asia between Indian and Pacific oceans. The over 5,000 km length archipelago is also situated between Asia and Australia. It has more than 17,500 islands, of which only about 7,000 are inhabited (Ministry of Foreign Affairs, 2018). There are eight main islands stretching from west to east namely Sumatera, Jawa (Java), Kalimantan (Borneo), Sulawesi, Bali, Nusa Tenggara, Maluku, and Papua. Indonesia is divided into three time zones, that is UTC+7 in the western part, UTC+8 in the middle, and UTC+9 in the eastern part.

Indonesia is a presidential republic, consisting of 38 provinces, including the national capital region of Jakarta. Each province, including the special region, is administered by an elected governor. The country's administrative structure further divides provinces into municipalities, led by mayors, and regencies. Indonesia operates under a decentralization system, wherein the central government delegates certain levels of authority in administration, fiscal matters, and politics to provincial and municipality/regency governments (Nuradhawati, 2019). As part of this system, public services, including education (early childhood, primary and secondary levels), fall under the jurisdiction of local governments.

According to the Indonesian Centre of Statistics in 2023 the Indonesian population was 278,696,200 where Java island consisting of 6 provinces, including Jakarta the capital city, is the most densely populated island accounting for 56 % of the population. Indonesia is socio-culturally diverse having more than 633 ethnic groups (bps.go.id, 2023) with 652 native languages spoken across the country excluding dialects and sub-dialects (Kemdikbud.go.id, 2018). These socio-cultural differences combined with the significant disparities in economic and human resources development between regions, have led to the inequalities in schools' conditions among Indonesian regions (Azzizah, 2015; Dharmawan & Suryadarma, 2021; Osses-Vargas et al., 2022).

3. The education system in Indonesia

Indonesia uses a 6-3-3 structure in its primary and secondary education system, encompassing six years of primary education (grades 1-6), three years of junior secondary (grades 7-9), and three years of senior secondary level (grades 10-12). The Ministry of Education governs both public and private general and vocational schools, which make up more than 80% of the total schools. Meanwhile, the Ministry of Religious Affairs manages religious-based education units, comprising fewer than 20% of all schools. As reported by the Ministry of Education (Rizaty, 2023), in 2023, there are over 400,000 formal education units in Indonesia (public and private), encompassing early childhood (180,351 schools), primary (149,225 schools) and secondary (71,941 schools) levels. Among the secondary levels, 42,907 are junior secondary schools (grade 7-9), the focus of the current study.

According to the Regional Government Act number 23 of 2014, while the Ministry of Education (MoE) holds full control on the policy and administration of higher education, the MoE and the local education authorities share responsibilities in managing education in the lower levels. The MoE is responsible for formulating the national education policy for early childhood, primary and secondary education, including the national standards of education, national curriculum as well as providing general guidance, monitoring and supervision of the implementation of education throughout the country. The local education authorities manage and oversee the operational implementation of education at early childhood, primary and secondary levels. They are responsible for executing the national education policies and programmes within their respective jurisdictions, including the implementation of the national curriculum. The local education authorities are also encouraged to enrich the national curriculum with locally relevant content and may design initiatives to enhance regional education quality. Additionally, they have the authority to recruit, promote and transfer teachers and other education personnel across the schools within their territory (Rubiyantoro, 2024).

4. The education reform in Indonesia

4.1 The *Emancipated Learning* policy and the transformation of assessment policy and practice in Indonesia

In early 2020, the Indonesian Ministry of Education (MoE) introduced a landmark educational reform initiative titled "*Merdeka Belajar*" (Clifton, 2023), translated here as "*Emancipated Learning*". The guiding philosophy of the policy is similar to Carl R. Rogers' (1969) humanistic concept of Freedom to Learn. Whilst not a direct adaptation, the policy reflects several of Rogers' core ideas, including the following principles (Rogers, 1969):

- Students develop their potential more effectively and quickly when they have the freedom to choose what and how they learn, rather than being dictated to.
- Schools and teachers act as facilitators of student learning, providing ample resources to support this learning. This role requires them to have adequate autonomy.
- Student learning should be largely self-initiated rather than based on a prescribed curriculum.
- Assignments, as a form of formative assessment, should be highly flexible.
- Standardized tests, including national examinations, should no longer carry high stakes for students.

Since its inception, the *Emancipated Learning* policy has sparked considerable debates among education scholars and practitioners due to its significant structural and pedagogical changes it proposes, such as the termination of national examination, granting schools and teachers extensive authority in assessing student learning outcomes, and teacher readiness issues (Clifton, 2023; Hunaepi & Suharta, 2024). The Ministry of Education introduced a series of policy packages through twenty-seven episodes of reform under the *Emancipated Learning* policy umbrella, encompassing different levels of the education system, from early childhood to higher education and various education themes/issues such as curriculum,

assessment, technological infrastructures, funding mechanism and budget structure, and school-to-work link (Utomo, 2024). As such, not all the education reform episodes are relevant to the current study, thus only the ones that have direct relationship with this study are highlighted and discussed.

Among the most prominent and contested changes were those related to curriculum, pedagogy, and assessment (Rukoyah et al., 2023). This study focuses specifically on the assessment policy reforms within the context of junior secondary levels (grades 7-9), examining how schools interpret and enact the new assessment policy introduced in the initial phases of the educational policy reform. The phases (i.e., episodes) of the reform that introduce the new assessment policy are presented in subsequent sections.

One of the earliest and most consequential reform episodes was the discontinuation of National Examination and National Standardised School-Based Test for primary and secondary students. These assessments, long embedded in the Indonesia education system were replaced by *Asesmen Nasional* (National Assessment in English), as the government's primary tool to monitor national education quality (MoE, 2021). The Asesmen Nasional consists of three components namely Minimum Competency Assessment, Student Character Survey, and Learning Climate Survey.

Unlike its predecessors, which were high-stakes assessments used to determine student graduation, the Asesmen Nasional is designed to evaluate school quality rather than individual student performance (Purnamasari et al., 2023). Its focus is diagnostic and developmental, aims to provide data that inform educational improvement without burdening students with exam-related pressure.

In terms of structure, the Minimum Competency Assessment (MCA) represents significant departure from the traditional national exam. Participation is limited to a random sample of students of year 5, 8, and 11 rather than being compulsory for all students in terminal grades (year 6, 9, and 12). Moreover, while previous national tests evaluated a wide range of subject-specific knowledge (e.g., mathematics, science, Indonesian, English, and social studies), the MCA focuses solely on

students' fundamental skills in literacy and numeracy, which are viewed as fundamental for further learning. This shift reflects a broader policy aim: to assess the effectiveness of school in fostering essential learning outcome, rather than judging individual academic achievement.

4.2 Reflecting the *Emancipated Learning* policy through Emancipatory Curriculum

Another significant reform with direct relevance to this study is the fifteenth episode of *Emancipated Learning policy*, which introduces the *Kurikulum Merdeka* (translated in this study as Emancipatory Curriculum). This new curriculum is designed to be succinct and more flexible than its predecessor reducing the quantity of content students are required to learn (Hunaepi & Suharta, 2024). This reduction aims to provide students with greater opportunity to engage more deeply into the concepts being taught, allowing for more meaningful learning experiences (Febriani et al., 2022). Additionally, the new curriculum provides schools with greater flexibility to tailor their educational approaches to suit their students' specific circumstances (Hunaepi & Suharta, 2024). One pedagogical shift introduced by the curriculum is the introduction of a differentiated instruction approach, which encourages schools and teachers to respond to individual student profiles. These distinguishing features, the succinctness, flexibility, and responsiveness to learner diversity, set apart the Emancipatory Curriculum from the previous curriculum.

The Emancipatory Curriculum's emphasis on student-centred learning is intended to promote active student engagement in classroom learning, fostering attributes such as creativity, independence, and autonomy (Anis & Anwar, 2020). Within this framework, teachers are expected to take on more facilitative role, empowering students by allowing them greater freedom to choose their learning methods and strategies. This approach is believed to be more effective in unlocking students' potential (Stiggins, 2001) and nurturing students' self-organising behaviour (Dangwal & Kapur, 2008).

Furthermore, the new national curriculum not only aims to enhance students' knowledge and skill acquisition but also emphasises the development of their

character by integrating local values into the school-based curriculum, aligning with the national philosophy, the *Pancasila,* promoting a motto Unity in Diversity. The curriculum also mandates that schools prioritise students' well-being, recognising the holistic nature of education and the interdependence nature of academic and emotional, social and ethical development.

In addition to these pedagogical and curricular shifts, the Emancipatory Curriculum also signals a transformation in assessment practices. Moving away from uniform, standardised approaches, the new curriculum encourages the adoption of formative, contextualised, and learner-centred assessment approaches. This shift has important implications for educators' assessment literacy as it demands a more nuanced understanding of assessment principles, along with the capacity to design and implement assessment practices that both instructionally embedded and responsive to individual learning needs.

4.3 The enactment process of the Emancipatory Curriculum

In the initial phase of the Emancipatory Curriculum implementation, the Ministry of Education involved 40,000 pilot schools (nearly 10% of all the schools from early childhood to secondary levels) both public and private schools to pioneer the enactment of the curriculum. During this time, the government also invited non-pilot schools to adopt the new curriculum on voluntary basis. By the end of 2023, it was reported that over 250,000 schools, representing around 60% of the total schools nationwide, have applied the Emancipatory Curriculum, either as part of the pilot programme or voluntary participation (Putra, 2023). The MoE have targeted that in 2025/2026 academic year, all schools will have adopted the new curriculum, except for the schools located in remote areas which are expected to be able to implement the new curriculum in 2026/2027.

The rollout of the new curriculum follows a phased approach. Initially, it begins with students in year one and year four of primary schools, year seven in junior secondary schools, and year ten in secondary schools during the first year of implementation. Subsequently, in the second year, it expands to include students

in year two and five of primary schools, year eight in junior secondary schools, and year eleven in secondary schools. By the third year of implementation, all students in pilot schools at every level were expected to fully engage with the new curriculum.

The Ministry of Education provided various support programmes to pilot schools for enacting the new curriculum. These programmes include initial training sessions for principals and teachers; monthly collaborative workshops organised for groups of pilot schools within the same geographic region; and quarterly project management officer (PMO) meetings involving representatives from the Ministry of Education, local school superintendents, principals and teachers from the pilot schools. These support initiatives extend over a three-year period. However, despite encouraging the non-pilot schools to adopt the new curriculum, they did not receive direct assistance programmes from the Ministry of Education. Instead, they were expected to independently utilise available resources provided on the Ministry of Education's official website or collaborate with nearby pilot schools for guidance and support.

To complement these efforts the Ministry of Education developed a dedicated education digital platform to aid in curriculum implementation for all schools. This platform hosts various instructional and assessment resources, including sample lesson plans, assessment plans, teaching modules, and videos demonstrating the application of the new concepts in the curriculum such as differentiated instruction and assessment for learning strategies. Additionally, the platform offers a community-sharing feature for teachers to document and share their successful practices, allowing colleagues to learn from each other's experiences.

5. Research Problem

5.1 Emancipatory Curriculum and challenges for principals and teachers

The previous curriculum was characterised by a high degree of standardisation, which constrained schools' ability to respond effectively to the diverse characteristics and needs of their students despite a highly diverse Indonesian

school context. For instance, the government regulates the format and content of lesson plans, dictates which assessment methods and techniques should be used to evaluate specific competencies and skills, and imposes national standardised tests and examinations. Such a top-down approach, while intended to ensure consistency, has been criticised for diminishing student enthusiasm for learning and undermining teachers' role in determining what is best for their students (Anis & Anwar, 2020).

The Emancipatory Curriculum, under the *Emancipated Learning* principles, represents a paradigm shift aimed to address these challenges. The curriculum advocates for schools and teachers to function as facilitators of learning rather than the sole providers of knowledge, recognizing that learning opportunities exist beyond traditional school environments. It acknowledges that authentic learning opportunities can arise through students' engagement with the wider social, cultural, and local environments. This perspective is embodied in the curriculum's motto, "everyone becomes a teacher, every place becomes a school," drawn from the wisdom of Indonesia's educational figure and inaugural Minister of Education, Ki Hajar Dewantara: the three centres of education: family, school, environment (Ferary, 2021).

However, while the new *Emancipated Learning* policy through its new curriculum appears to be promising with its emphasis on flexibility and reduced bureaucratic constraints, it also introduces new challenges for school leaders and teachers. They now have greater responsibility in developing school-specific curriculum aligned with the principles of *Emancipated Learning*. This task is more demanding because, while in the previous curriculum, schools could rely on government-provided materials, such as lesson plans, learning modules, and the assessment instruments, now they are expected to construct context-specific curriculum, which respond to their local realities.

Under the recent Indonesian curriculum reform, schools must create a customised operational curriculum that takes into account various internal and external factors. These internal factors encompass teachers' competencies, available school

resources and facilities, and the diverse characteristics of students, which include academic, socio-cultural, and economic aspects. Schools are also encouraged to explore potential collaborations with external community entities, such as local government bodies and private institutions (Maisyaroh et al., 2021). Additionally, schools are required to develop student assessment plans that are tailored to the specific context and diverse characteristics of both the school and its students and to embed these plans as an integral part of the operational curriculum. Scott (2016) argues that an evolving challenge related to assessment tasks a principal faces is providing a learning environment appropriate to various students' characteristics, including socioeconomic status, racial, linguistic, and religious backgrounds.

Furthermore, a notable change of the Emancipatory Curriculum policy was the introduction of multi-subject student thematic projects, as a mandatory component of the school operating curriculum. These projects, which blend topics from 3-4 subjects such as social studies, science, mathematics, or others, occupy 25% to 35% of total instructional time in a given academic year. By applying these projects, students were expected to apply knowledge across disciplines, while schools were encouraged to incorporate distinctive social and cultural elements of their local communities into their educational programmes.

Consequently, schools are responsible for developing assessment methods and instruments to effectively evaluate the diverse aspects of students' competencies targeted by these projects. These assessments must capture a range of student learning outcomes, which reflect the curriculum's broader educational goals.

5.2 Changes in the assessment policy and related consequences for principals and teachers

The *Emancipated Learning* policy and the new curriculum led to a transformation in student assessment policy and practices. This transformation requires a certain level of assessment literacy as indicated by the following four situations. Firstly, crafting school assessment plans, that are both distinctive and complex, as part of the school operating curriculum, necessitates a specific level of expertise in

assessment knowledge and skills. It is widely acknowledged that the ways in which school administrators implement assessment practices significantly shapes the educational decision-making within the school and influence students' academic performance, thus playing a pivotal role in overall school effectiveness (Stiggins, 2001). In their mixed-method research, Day, Gu, and Sammons (2016) highlighted the importance of refining assessment procedures in bolstering school effectiveness, as recognised by primary school principals in the UK. Consequently, the capacity to make informed data-driven decisions, a crucial aspect of enhancing school effectiveness, is heavily dependent on the assessment literacy of all stakeholders within the school community.

Secondly, as noted earlier (see section 2.1), the initial phase of the policy involves discontinuing national examination and nationally standardised school tests. The new Minister contended that, arguably, the national examination and standardised school tests primarily evaluate students' lower-order thinking skills, which do not align with the goals of the national education and the demands of the 21st-century workforce (Kemdikbud.go.id, 2020). In addition, the opponent of the high-stakes assessments argued that schools often prioritise achieving high results, sometimes at the expense of students' deeper and more meaningful learning experiences (Nurhidayah, 2013; Wardhani, 2017). Consequently, teachers were less inclined to prioritise fostering higher order thinking skills and meaningful learning in the classrooms in pursuit of high scores on the national exams. Under this assessment reform, schools now have the autonomy to design and execute their own student assessment systems, incorporating a variety of methods and criteria for passing. This empowerment allows schools to interpret assessment data and determine whether students have achieved predefined targets.

Although schools now have more autonomy in designing and administering their own assessments, they still need to uphold assessment standards, theories, and principles to ensure the accuracy and validity of the results. Valid and trustworthy information is crucial for making effective decisions. Therefore, while the policy provides flexibility, it also requires principals and teachers to be creative and accurate in assessing student performances, both necessitate adequate level of

assessment literacy. Therefore, it is timely and essential to understand how principals and teachers perceive and assessment use data (Militello et al., 2013), especially under the mandated assessment reform.

Thirdly, the new curriculum emphasises student holistic development, encompassing literacy, numeracy, and character qualities aligned with national values. The aim is to enhance students' overall competence so that upon completing their education, they do not only exhibit cognitive, reasoning, and critical thinking skills but also embody character traits reflective of national values and local cultures (Irawati et al., 2022). As a result, schools need to conduct a holistic assessment approach capturing wider and more diverse spectrum of students learning achievement. Designing and administering assessments to evaluate such diverse learning outcomes requires a considerable level of assessment proficiency.

Additionally, given the mandated attention on students' well-being in the school curriculum and classroom teaching, principals and teachers should prioritise assessment policies and practices. This emphasis is crucial as assessment practices can profoundly influence students' emotional and mental well-being as well as their educational trajectories and future lives (Ysenbaert et al., 2020).

Lastly, in the *Emancipated Learning* policy, the MoE re-emphasised that schools apply an assessment for learning (AfL) approach, which is "thought of as metacognitive tools designed to support instruction to facilitate the creation of a learning profile for students to track their progress over time" (Hollingworth, 2012, p. 369), in their assessment practices. This approach has emerged as a prominent trend in various countries including Australia, the UK, Canada, New Zealand (DeLuca et al., 2016b), the US and numerous Asian countries (Puad & Ashton, 2020). The Ministry of Education re-emphasised this approach likely because of its congruency with the idea of *Emancipated Learning* policy as Ysenbaert et al. (2020) contended that applying assessment in the form of assessment for learning is more adaptive to classroom diversity.

The fast development of assessment for learning (AfL), a form of formative assessment (Warwick et al., 2015), across many countries over the last two decades is underpinned by considerable empirical evidence of its powerful impact in increasing students' learning outcomes (Black & Wiliam, 1998; Militello et al., 2013). Although, many studies in the last two decades also pointed out that many young teachers reported that they were not fully prepared to practice this assessment (DeLuca et al., 2019). Despite the promotion of assessment for learning in previous Indonesian curriculum, its implementation fell short of expectations due to inadequate attention and oversight from education authorities (Arrafii, 2021b; Puad & Ashton, 2020). Moreover, the conflicting nature of assessment for learning paradigm with the standardised examination system further hindered its adoption (Azis, 2015; Heng et al., 2021). However, recognizing the significance of this assessment approach, the education authority has prioritised its importance and implementation through the *Emancipated Learning* policy. In such assessment policy reform, developing teachers' comprehensive understandings across various assessment practices is critical (DeLuca et al., 2018). Additionally, principals are expected to support the development of teachers' assessment literacy, particularly the application of the concepts of assessment for learning (Renihan & Noonan, 2018).

6. Justification of the study

The above four conditions were a significant shift in the assessment policy and practices in Indonesian national curriculum reform. They require principals and teachers to possess sufficient set of assessment knowledge and skills to successfully execute the *Emancipated Learning* policy. Hasim, Di, and Barnard (2018, p. 164) argued that "Educational policies should consider the existing knowledge and practices of those most directly concerned with its implementation", such as the principals and teachers in the assessment reform in Indonesia. Arrafii (2020) in his doctoral research on English as a Foreign Language (EFL) teachers assessment for learning practices in Lombok Indonesia during the previous curriculum found that teachers' lack of assessment knowledge and skills caused them to fail to grasp the

core idea of the AfL concepts and thus fail to carry out the desired practices appropriately. Therefore, identification and awareness of the current assessment literacy of school principals and teachers by the government allows the education authority to determine more effective and appropriate actions to ensure the successful implementation of the reform.

DeLuca et al. (2019) argued that assessment literacy topics have received minimal attention within research regarding assessment and measurement despite its importance in school assessment practices. Additionally, there has been a lack of research focusing on the assessment literacy of school principals. Most studies regarding school leadership in recent years have tended to address broader aspects such as managerial, transformational, instructional, and distributed leadership. In the Indonesian context, studies exploring the assessment literacy of teachers remain scant. Most of the research identified in Indonesia on teachers' assessment literacy has primarily focused on English teachers, with only a few studies examining science teachers (Aria et al., 2021; Efendi et al., 2018; Rasyidah et al., 2020). Globally, there is a lack of studies specifically investigating the assessment literacy of teachers of different subject areas (Zolfaghari et al., 2016), including social studies teachers.

DeLuca et al. (2019) recommended further empirical studies on the contemporary concept of assessment literacy, which pays more attention to the various influencing factors that shape educators situated and differential approaches to assessment practices. In light of those gaps and the notable importance of educators' capacity concerning student assessment highlighted above, particularly amid the Indonesian assessment reform, the increasingly complex assessment practices, required by the new policy, underscore the need for a comprehensive study on the assessment literacy of both principals and teachers. Evidence from this study is expected to add to the body of knowledge in the field of school leadership concerning principals' assessment literacy. In addition, this study aims to contribute to research on teacher assessment literacy by focusing on social studies teachers. It adheres to the principle that teacher assessment literacy should be grounded in specific content areas, recognising that social studies remains one of the least

examined school subjects in research on teachers' assessment literacy. In particular, Morris (2017) suggested that future studies about principal and teacher assessment literacy should employ both quantitative and qualitative approaches, which may cancel out the inherent limitations within each approach.

The diverse socio-economic and cultural characteristics of regions across Indonesia will potentially pose another challenge for Indonesian school principals and teachers in enacting the mandated assessment reform into their respective contexts. Hence, this study will also shed light how the educators utilise their agency underpinned by their assessment literacy to navigate the policy implementation especially in balancing between the demand of the national policy and their local circumstances. Viewing educators' assessment literacy as a context-dependent and multifaceted practices is important for the policy makers to provide more relevant support for the principals and teachers in their assessment practices (DeLuca et al., 2019).

7. Research objectives and questions

The overall purpose of the study, to contribute to the understanding of assessment literacy of school principals and social studies teachers in Indonesia, is broken down into four more detailed research objectives as follows.

The first objective of this research is to reveal the assessment knowledge and skills of Indonesian school principals and social studies teachers, recognising its pivotal role in the successful implementation of the *Emancipated Learning* policy, more specifically the assessment reform. As schools enjoy more flexibility and autonomy in student assessment, the assessment literacy of school leaders and teachers becomes essential. This competency will enable educators to generate dependable and valid student information, forming the foundation for sound educational decision-making. For principals and teachers to effectively make educational decisions, they require accurate and reliable information on student achievement obtained through various assessment practices (Stiggins, 2001).

Second, this study also examines how four factors: principals' and teachers' assessment-related experiences; assessment courses in higher education; assessment-related professional development programmes; and assessment-related support available in the local context, contribute to principals' and social studies teachers' assessment knowledge and skills. In this study, the assessment-related experiences refer to three aspects: teaching experience; experiences as a facilitator/trainer/speaker in any assessment-related trainings, workshops, and/or other professional learning activities; and experiences as a student test/assessment writer at national/local levels. Furthermore, assessment courses in higher education refer to assessment courses undertaken during pre-service teacher education programme and/or as part of postgraduate studies. Assessment-related professional development programmes mean any formal or informal professional learning and development activities where principals and teachers engaged with assessment-related topics. Lastly, assessment-related support refers to the support concerning student assessment practices that the principals and teachers receive from their local colleagues, including teachers, leaders, teacher/principal professional communities, or school superintendents.

Studies have shown that experience may significantly impact how educators carry out assessment practices (DeLuca et al., 2018; Rasyidah et al., 2020; Zolfaghari & Ashraf, 2015). Other studies also pointed out the importance of sufficient and continuous assessment education both through formal higher education and in-service professional development programmes for both principal and teacher to effectively perform their assessment-related responsibilities (DeLuca et al., 2013; Deneen et al., 2016; Stiggins & Duke, 2008; Zhang & Burry-Stock, 2003). Support from colleagues through formal/informal professional learning, discussion, and sharing experiences in the educators' professional communities is also evident in improving educator assessment literacy (Bostrom & Palm, 2020; DeLuca et al., 2018; Xu & Brown, 2016). However, other studies also have shown different results concerning the four factors highlighted above. This study may provide evidence to related Indonesian policy makers how those factors, especially the three last ones: assessment-related education in higher education; in-service training programmes

in student assessment; and local support such as principal/teacher professional community, are effective or in-effective in improving educators' assessment literacy.

The third objective of this study is to explore how principals and social studies teachers implement the assessment reform mandated by the *Emancipated Learning* policy, particularly within their unique local contexts. Specifically, it examines the strategies they employ to reconcile the requirements of the new assessment policy with the specific circumstances of their students, school resources, and broader society. Research suggests that teachers' theoretical knowledge of assessment does not always translate directly into classroom practices (Darling-Hammond et al., 2019). By investigating the enactment of the assessment reform, this study intends to provide a clearer picture of the educators' assessment literacy beyond merely their level of assessment knowledge and skills. According to Xu and Brown (2016), a specific context in which different dynamics affect educators' assessment-related decision making and action-taking is crucial in understanding and improving assessment literacy.

Finally, this study seeks to examine the relationship between principals' and teachers' assessment literacy. According to Charteris and Smardon (2021), as an assessment leader in school, a principal should be able to identify and evaluate teachers' assessment skills and upskill teachers' data literacy. "Without question, a school leader plays a key role in setting policies and providing opportunities for professional development so teachers can successfully use assessment for learning" (Hollingworth, 2012, p. 367). Both Charteris and Smardon (2021) and Hollingworth (2012) arguments were pointing to the connection between principals' assessment literacy, emphasising school-wide assessment policy and plan, and teachers' assessment literacy, focusing on classroom assessment. With the increasing emphasis on the role of school principals in curriculum and assessment reforms, the findings of this study can provide insights for government programmes concerning school principals. This aspect is particularly significant given the historical lack of focus on student assessment in principal training programmes.

The overarching research question to achieve the main purpose of the research is: *What is the current state of Indonesian principals' and social studies teachers' assessment literacy in enacting the assessment reform in diverse school contexts?*

To provide an answer to the overarching research question, the following specific research questions were addressed in this study:

1. *What is the level of Indonesian principals' and social studies teachers' assessment knowledge and skills?*

2. *How do assessment-related experiences, assessment courses in higher education, in-service assessment professional development programmes, and local support contribute to Indonesian principals' and social studies teachers' assessment knowledge and skills?*

3. *How do school principals and social studies teachers navigate the enactment of mandated assessment reform against the challenges within their unique local context?*

4. *How do the Indonesian principals' assessment literacy and leadership in assessment affect social studies teachers' assessment literacy?*

8. Structure of the thesis

Following this chapter, Chapter Two presents the literature review, synthesising conceptual and empirical literature on principal and teacher assessment literacy. It examines the key variables affecting assessment literacy as well as reviews global perspectives of assessment reform implementation. In addition, Chapter Two describes the theoretical frameworks informing the investigation, namely, social constructivism theory, theory of practice and multiplicity of enactment. It also establishes the contribution of this research regarding existing studies concerning assessment literacy, policy, and practice.

Chapter Three describes the methodology of the research. This study employs mixed methods as the methodology in addressing the research questions. This chapter details the research paradigm (the ontology, epistemology, and axiology), the research approach and design, and methods used, including data collection instruments, the population and samples of the study, as well as the data collection procedure and analytical strategies for both the quantitative and qualitative data. It also discusses the ethical consideration of this research to underscore the rigor and integrity of the process.

Chapter Four presents the quantitative findings from phase one of the study, while Chapter Five describes the qualitative findings from phase two. Chapter Six synthesises these findings discussing them in relation to the theoretical frameworks and prior studies to address the research questions. This integrated discussion underscores the patterns, contradictions, and implications for assessment policies and practices.

The thesis concludes with Chapter Seven, highlighting how the research questions were addressed and summarising the key contributions to the body of knowledge in the field of assessment literacy, policy, and practice. It acknowledges the limitation of the study and presents the implications of this study to the policy and practice of educational assessment in Indonesia.

9. Summary

Chapter One provides the basis for the research by introducing background information and the context of the research. It also covers the positionality of the researcher, the conceptualisation of the main constructs of the study, and the theoretical frameworks guiding the investigation.

This chapter also presents the research problems which emerge from recent transformation in education policy in Indonesia. It is followed by the study's purpose and objectives, and the research questions, which provide direction for the subsequent stages of the investigation. By situating this study within the broader

landscape of research on assessment literacy, this chapter underscores the justification for inquiry.

The next chapter critically examines past literature regarding principals' and teachers' assessment literacy, principal assessment leadership, assessment policy and standards, and assessment practices. The chapter justifies how the current study adds to the knowledge of assessment literacy of school principals and teachers.

Chapter two: Review of literature

1. Introduction

This chapter reviews existing conceptual and empirical literature related to assessment literacy, assessment policy, assessment standards and practices. The aim is to examine prior research studies on those topics, identify the gaps, and identify how this current study might build upon the findings of the studies to expand the body of knowledge in the field of assessment literacy and practice. This chapter also outlines the conceptual foundations of the main constructs in the study, the assessment literacy of school principals and teachers, with particular attention to the contemporary concept of assessment literacy, incorporating contextual, social, and cultural aspects (Willis et al., 2013).

This chapter draws on a range of sources, including theoretical/conceptual works and empirical research published as peer reviewed journal articles, academic books, book chapters, theses, dissertations and policy documents, addressing both the classical and the contemporary conceptions of principal and teacher assessment literacy. Additionally, the literature reviewed in this chapter also addresses the factors affecting assessment literacy and the enactment of assessment reform in various educational settings, and the theoretical frameworks relevant to this study. Each of the topics is explored thematically in separate sections. Gaps found in the research studies reviewed in this chapter will act as the foundational elements to create a warrant for this research that aims to make an original contribution to the knowledge about assessment literacy.

The search was restricted to literature published in English, except for Indonesian policy documents and a few Indonesian journal articles. In the first stage, the literature search started using key terms within the title and abstract, “assessment literacy”, “teacher assessment literacy”, “principal assessment literacy”, combined with the words “classroom assessment”, “student assessment” and “assessment reform”. Subsequently, similar or related words such as “assessment competence”,

"assessment knowledge", "assessment expertise" and "assessment leadership" were also applied.

The literature search in the databases, such as Eric, Scopus and Web of Science, was confined to the sources from 2013 to present, peer-reviewed and full text. The selection of the works, especially the conceptual and empirical literature, published within approximately ten years was to maintain a balance between including moderately up-to-date relevant sources, while reflecting recent advancement and trend of the research topic within a reasonable timeframe. Key international journals were also searched such as *Assessment in Education: Principles, Policy, and Practices; Educational Assessment; Educational Assessment, Evaluation, and Accountability*. Additionally, the reference lists of the downloaded sources, especially those within the last three years, were scanned to identify whether recent relevant sources had been overlooked and once identified, included them for reviewing.

Section two begins by highlighting the importance of student assessment and the strength of assessment for learning (AfL), which plays a central role in the current assessment reform in Indonesia, aimed at improving student learning. It emphasises that assessment literacy is essential for implementing effective and responsive assessment practices. The section continues by defining traditional concepts of assessment literacy, more specifically teacher assessment literacy and principal assessment literacy. The traditional concepts of assessment literacy are then referred to as assessment knowledge and skills, which forms the basis for the broader and more contemporary concept of assessment literacy used in this study. The review of literature also reveals common themes that laid the foundation for establishing the assessment dimensions used to examine the teachers' and principals' assessment knowledge and skills in this study. Additionally, the literature on contemporary perspectives of assessment literacy, viewed through a socio-contextual lens, were explored to provide a framework for examining teacher and principal assessment literacy more deeply.

The following two sections, section three and section four, review empirical studies concerning assessment literacy and its implementation. Firstly, section three critically examines the existing literature on factors that affect teachers' and principals' assessment literacy. These factors include assessment-related experiences, attendance on courses in student assessment at higher education level, participation in in-service professional development programmes, and assessment-related local support from colleagues. The review also incorporates research conducted in Indonesia, published in international and national journals concerning this issue to provide localised perspective.

This is followed by section four, discussing the literature on the enactment of educational policy, with a focus on assessment reform policy and practices from various countries. The international studies on educational policy reform enactment highlighted the approaches of the enactment to realise the policy changes in education as well as the factors that influence the successful implementation.

Next, section five delves into three theoretical frameworks that informed this study: Vygotsky's social constructivism (Adams, 2006), theory of practice (Bourdieu, 1977) and multiplicity of enactment (Buxton et al., 2015). While social constructivism was employed as a lens for the concept of assessment literacy used in this study, the theory of practice and the multiplicity of enactment were applied in the second stage of the study to provide insights into the enactment of the assessment reform under the *Emancipated Learning* policy (described in Chapter One) across diverse contexts within Indonesia. This approach aims to obtain a more comprehensive picture of the principals' and teachers' assessment literacy. The three theories have been put to work as the frameworks for various topics in education research. However, their use in explaining principal and teacher assessment literacy and the enactment of assessment policy remains largely unexplored.

The chapter concludes with section six summarising the reviews and describing the gaps in the literature on this topic and explaining how the current research aims to fill the gaps. The current study's findings are intended to produce new knowledge

and enhance the understanding of educators' assessment literacy and contribute new insight to the literature.

2. Concepts of assessment literacy

The global trend in education reform agenda focuses on improving pedagogical and assessment practices that ensure fair and non-discriminative outcomes for all students (DeLuca et al., 2019). This approach emphasises the importance of upholding student rights to access learning opportunities, make progress and achieve success. As Willis, Adie, and Klenowski (2013, p. 253) argued "developing teacher assessment literacy is part of the ethical and moral responsibility to provide an opportunity for all students to learn". For decades, educational policy makers and educational researchers have prioritised assessment literacy in their agenda not only because student assessment is acknowledged as one of the levers for educational reform effort but also it facilitates teachers to support student learning by creating responsive assessment practices within classrooms (Pastore, 2023).

In essence, acquiring assessment literacy means education practitioners, especially those working in school and classroom settings, have strong understanding of several key aspects in education which include how to collect evidence of learning and how to interpret and use the evidence to improve learning (Livingston & Hutchinson, 2017). Despite the concept of assessment literacy has evolved for more than three decades, it remains a complex and contested concept as pointed out by a large body of literature in Pastore's (2023) systematic review study.

In this current study, the concept of assessment literacy that is limited to knowledge and skills in the area of student assessment is referred to as classical concept of assessment literacy. The next two sub-sections will describe the classical concepts of teacher assessment literacy and principal assessment literacy in more detail.

2.1 Classical concept of teacher assessment literacy

Growing attention to teacher assessment literacy can be traced back more than three decades, arguably beginning with the introduction of Standards of Teacher Competence in Educational Assessment of Students in the United States of America in 1990 (Xu & Brown, 2016). As awareness of the central role assessment plays in student learning and learning outcomes has increased, there has been more research on teacher assessment literacy topics in the last ten years as indicated by the number of research studies about it across many countries as shown in a systematic review study of Pastore (2023).

One of the key purposes of classroom assessment is to use the derived assessment data to improve student learning (Black & Wiliam, 2018; Brookhart, 2011). Hence, quality classroom assessment is required to help students attain better learning outcomes. Indonesia's *Emancipated Learning* policy introduced an assessment reform as part of the changes of the national curriculum. The assessment policy emphasises the use of assessment for learning approach, a form of formative assessment, to obtain a more complete picture of students' learning progress and to improve learning outcomes (Centre of Assessment and Learning, 2022), as it is believed to have positive effects on students' academic achievement (Black & Wiliam, 1998). Assessment for learning has become popular in many countries such as the United State of America (Heritage, 2010), the United Kingdom, Australia, Canada, New Zealand (DeLuca et al., 2016b; Hasim et al., 2018), as well as in many Asian countries (Arrafii, 2021a; Heng et al., 2021; Leong et al., 2018; Puad & Ashton, 2020).

The policy requires teachers to support students in their learning process by gathering more authentic data through various formative assessment tasks, which cannot be provided only through summative tests administered at certain points in time such as end-of-semester and end-of-year tests. The data from the formative assessment results then are used by teachers as the feedback to develop more effective classroom teaching (Centre of Assessment and Learning, 2022). According to Black and Wiliam (2018), summative assessment functions to draw inferences

about the status of student learning and/or their future potential, while the inferences drawn through formative assessment relate to teacher actions that would help students learn better. The Indonesian Ministry of Education (2022a) defines that formative assessment is used as the feedback to monitor and improve student learning while summative assessment is used specifically for reporting student learning achievements.

Assessment for learning requires teachers to go beyond simply evaluating students' content knowledge through grading practices using traditional form of tests, such as paper-and-pencil-tests, which is commonly observed in the Indonesian previous curriculums (Arrafii, 2021b; Azis, 2015; Puad & Ashton, 2020; Zuhdi, 2015). Instead, teachers should employ a variety of methods and techniques, such as performance-based assessments, projects, portfolios, anecdotal records, self-assessment, and peer assessment to evaluate broader aspects of learning (MoE, 2016). Forster (2024) maintained that when classroom assessment involves more intricate and integrated skills such as developing projects, conducting research, or creating portfolio, teachers would be able to foster student learning better by treating these activities as a comprehensive process rather than isolated assessment tasks. Using diverse assessment techniques would allow for a more complete evaluation of learning outcomes (Squire, 2010).

Effectively using a range of assessment procedures and methods and analysing assessment results to inform the teaching and learning processes requires high-level of assessment expertise. According to Brown (2008, p.286) "an assessment-literate teacher creates, chooses, administers, interprets, responds to, records, and reports assessment information in such a way that those decisions can be shown to be adequate and appropriate".

Popham (2011) defines teacher assessment literacy as knowledge and skills teachers need to possess to measure and improve student learning through assessment practices. According to Stiggins (2001, p. 20), teacher assessment literacy consists of two skills, "first is the ability to gather dependable and quality

information about student achievement; second is the ability to use that information effectively to maximise student achievement".

Although expressed in slightly different terms, many definitions of teacher assessment literacy found in research over the past decade have drawn on the definitions proposed by Popham and/or Stiggins (Christoforidou & Kyriakides, 2021; Clark et al., 2022; DeLuca et al., 2016b; DeLuca et al., 2018; Gotch & French, 2014; Looney et al., 2018; Ogan-Bekiroglu & Suzuk, 2014).

Other researchers did not associate their definitions of assessment literacy to those of Popham and Stiggins, yet their definitions share similar elements. For example, Yamtim and Wongwanich (2014, p.2999) defined assessment literacy as "necessary knowledge and skills for compiling data about students' achievement and for effectively utilising the assessment process and outcomes to improve students' achievement". Similarly, Kaden and Patterson (2014, p. 407) described it as "the ability to design, select, interpret, and use assessments appropriately for educational decisions", or Adie, Stobart, and Cumming (2020, p. 439) defined it as "the skills and knowledge teachers require to measure and support student learning through assessment", while Edwards (2017, p. 206) referred to it more simply as 'assessment knowledge and skills'.

While often referred to as the traditional (i.e., classic) definition of assessment literacy (Pastore, 2023; Willis et al., 2013), the concept of teacher assessment literacy as the knowledge and skills of teachers in educational, classroom, or student assessment continues to be widely adopted by many researchers until recently. This enduring focus shows that teachers' mastery of knowledge and skills in assessment remain central even as contemporary concepts of assessment literacy evolve. The knowledge and skills of assessment are vital because "education systems need a core body of formal, systemic, and codified principles concerning good assessment practices" (Xu & Brown, 2016, p. 155).

Teacher assessment literacy is a complex construct. It encompasses a range of areas or themes related to the knowledge and skills, such as the purposes of assessment,

the construction of assessment, and ethics and fairness in assessment. These areas are also a measurable cognitive dispositions (Herppich et al., 2018). Therefore, to more accurately evaluate teacher assessment literacy, it is essential to specify the key knowledge and skills involved in student assessment as highlighted in most of the literature earlier. Identifying the areas of teacher assessment knowledge and skills can be done by examining the instruments used by researchers to investigate teacher assessment literacy and by reviewing standards for teacher assessment competencies across different countries.

DeLuca, LaPointe-McEwan, & Luhanga (2016b) conducted a systematic review of post-1990 teacher assessment literacy measures using education research databases. They identified eight most widely used measures/instruments for evaluating teacher assessment literacy between 1993 and 2012. Analysis of these instruments revealed seven key themes commonly evaluated as measures of teacher assessment literacy, which are assessment purposes, assessment processes, communication of assessment results, measurement theory, assessment ethics, assessment fairness, and assessment for learning. These themes were also reflected in 15 international standards of teacher competencies in student assessment in various countries including the UK, USA, Australia, Canada, New Zealand, and European countries (DeLuca et al., 2016b). Pastore (2023) pointed out that these key themes are still being used in studies evaluating teacher assessment literacy until recently, such as by Oo et al. (2024).

Following the systematic review of assessment literacy instruments, DeLuca et al. (2016a) developed a new instrument on teacher assessment literacy called Approaches to Classroom Assessment Literacy Inventory (ACAI). This instrument demarcates the construct of teacher assessment literacy into four themes: “Assessment Purposes, Assessment Processes, Assessment Fairness, and Measurement Theory” (p. 258). The key themes of assessment found in the above literature informed the development of the assessment literacy instrument in this study.

The traditional assessment themes found in the eight most widely used assessment literacy instruments from 1993-2012 as well as in the subsequent studies focused solely on general knowledge and skills of assessment. However, some authors argued that assessment competencies should be associated with subject-specific skills (Brookhart, 2011; Pastore & Andrade, 2019). Limiting the measurement of assessment literacy to only core knowledge and skills of assessment without considering subject-specific knowledge is potentially problematic, since the approaches and methods of student assessment may differ from one subject to another, because each subject has its own characteristics (Brookhart, 2011). In addition, educators need an assessment pedagogical content knowledge (APCK) to align their assessment practices to instructional goals (Militello et al., 2013). Militello et al. used the acronym "APCK" to emphasise the need of integrating student assessment practices into instructional practices within a particular subject area as one of the key features of formative assessment.

Zolfaghari, Ahmadi and Hui (2016) sought to capture teachers' views of assessment literacy components by interviewing 32 Iranian secondary teachers of various subjects. The findings show the teachers' view that assessment literacy instruments should comprise both general and subject-specific components. Building on this idea, Dunlai and You (2015) and Nikmard and Zenouzagh (2020), have developed instruments specifically intended to measure subject-specific teachers' assessment literacy for EFL teachers.

Xu and Brown (2016) developed a framework of teacher professional development in assessment literacy, encompassing pre- and in-service teacher trainings. The framework, called the Teacher Assessment Literacy in Practice (TALiP), emphasises core assessment knowledge and subject specific content knowledge as the foundational elements in building teacher assessment literacy.

In light of the above literature and aligning with Indonesian Ministry of Education regulations (the MoE Regulation number 23/2016 concerning standards of assessment and the Guidance on Student Learning and Assessment), this study has developed a framework for examining assessment knowledge and skills of teachers.

It incorporates the key themes of assessment identified in the reviewed literature and links them with social studies subject content. This framework then was used to develop teacher assessment knowledge and skill inventory, which will be described in detail in the methodology chapter.

Following the interest in teacher assessment literacy, topics regarding principal assessment literacy and principal leadership in assessment also received more attention from educational researchers and scholars, especially in the last ten years such as Ysenbaert et al. (2020) and Renihan and Noonan (2018). It will be discussed in the next section.

2.2 Classical concept of principal assessment literacy

The central purpose of school leadership for a principal is to lead a school effectively with the primary goal of enhancing student holistic learning outcomes. A report of a longitudinal study (Grissom et al., 2021) involving 22,000 principals in the United States concluded that despite the different roles principal and teacher have in student achievement, the principal's effect is substantial and nearly equal to the effect of effective teachers on student outcomes, especially by providing students with solid teaching and learning.

Ysenbaert, Houtee, & Avermaet (2020) emphasised the pivotal role of school principals in establishing school-level assessment policy and culture to create and develop a shared vision and understanding of student assessment practices, more particularly in the circumstance in which schools have a significant degree of autonomy and flexibility regarding their instructional and assessment practices. This idea resonates with the current conditions of Indonesian schools under the *Emancipated Learning* policy.

Some authors asserted that principal assessment literacy is an essential factor in assisting teachers' classroom practices, more specifically in conducting classroom assessment, which ultimately affect student learning. For example, Militello et al., (2013) and Stiggins & Duke (2008) argued that principal assessment literacy is part

of instructional leadership with which they can support the improvement of teaching and learning, while Popham (2014) contended that it is a critical component required for effective school practices. Assessment literacy is also essential for a principal by which educational decision-making is influenced (Pfeiffer-Hoens, 2017).

According to Stiggins and Duke (2008) assessment-related activities may take up to 30% of teachers' professional time, and it is part of the principals' instructional leadership responsibilities to ensure that the assessment practices are carried out effectively and of high quality to support student learning. Black, Harrison, Lee and Marshall (2003, p. 113) argued that it is the responsibility of the school leaders to provide their schools with "leadership, stimulus, and support" to enhance assessment for learning practices. However, this can only be achieved when the principals are well prepared and possess key competencies in student assessment.

Even though a school principal generally does not carry out student or classroom assessment directly, their knowledge of and skills in student assessment are important in assisting teachers. As an instructional leader, a principal must have the ability to monitor, evaluate, and give feedback on the performance of teachers in carrying out classroom assessments so that they can gain teachers' trust not only as a leader but also as a knowledgeable authority in the assessment practices (Smith & Engelsen, 2013) as trust is one of the essential leadership factors (Louis et al., 2009).

An assessment literate principal can provide teachers with support to develop and use productive classroom assessment and eliminate any barriers hindering the teachers' assessment literacy development, including personal, institutional, and community obstacles (Stiggins, 2001). School leaders need to make hard decisions to promote high-quality assessment that align with best teaching practices (Webber et al., 2013).

Despite the important role of assessment literacy of a principal, few studies investigating the level of principal assessment literacy were found published in

international databases. In the last twelve years, some doctoral studies examining the level of principal assessment literacy were conducted in the United States of America, for example, Lee-Perry (2013), Rosas (2014), Ball (2017), Morris (2017) and Pfeiffer-Hoens (2017).

Lee-Perry (2013), Rosas (2014), and Pfeiffer-Hoens (2017) examined and compared the assessment literacy between principals and teachers in the United States, using assessment literacy instruments developed by Mertler (2005) or Mertler and Campbell (2005) based on the USA-based Standards for Teacher Competence in Educational Assessment of Students. However, Pfeiffer-Hoens (2017) combined the results of both groups due to an extremely low number of principal participants (N=5), reporting an average score of 59.2 (out of possible 100).

Meanwhile, the studies of Lee-Perry and Rosas yielded different results. Lee-Perry (2013) found no significant difference of assessment literacy between the principals and teachers at high schools in Montana. In contrast, Rosas (2014), who investigated principals and teachers at elementary schools in rural areas of Central Valley California, discovered that the principals significantly outperformed the teachers. It is worth noting that the research context of both studies was different in terms of the level of education (i.e., primary and secondary) and geographic characteristics of the areas, both of which might lead to the difference in findings.

Ball (2017) and Morris (2017) investigated the assessment literacy level of principals at primary and secondary schools in Virginia and Florida respectively. The studies documented evidence that the principals' assessment literacy levels in these two regions were similar (i.e., 65,2% and 63%), with no significant difference between primary and secondary schools. The two doctoral researchers employed similar assessment literacy instruments as the ones used by Lee-Perry, Rosas, and Pfeiffer-Hoens.

Notably, those five US studies examined the principals' assessment literacy level using the framework originally designed to evaluate teachers' assessment literacy. Given the nature of their role, principals require knowledge and competencies in

assessment that extend beyond classroom practices. They are also held accountable for implementing mandated assessment policy in their schools. As Stiggins and Duke (2008) contended, a principal must be assessment literate at classroom-level, programme-level, and institutional level so that they can play an influential role as an instructional leader.

As a consequence, principals' assessment leadership requires building school-wide assessment competence (Absolum et al., 2009; Hipkins & Cameron, 2018), in which one of the main responsibilities is to support teachers to have sufficient assessment capability to make informed judgments about student learning and improve teaching practices (Wyatt-Smith et al., 2017). Therefore, principals need to create a balanced assessment system that aligns both the classroom level and institutional level assessment, while helping teachers develop their assessment literacy. To that end, the assessment literacy required of a school principal differs compared to that of a teacher although there are overlapping elements between the two.

Stiggins and Duke (2008, p. 287) introduced ten leadership competencies in assessment to embody an effective role as a leader in school assessment practices:

1. "Understands the principles of assessments for (that is, used in support of) learning and works with staff to integrate them into classroom instruction.
2. Understands the necessity of clear academic achievement targets and their relationship to the development of accurate assessments.
3. Knows and can evaluate the teacher's classroom assessment competencies and helps teachers learn to assess accurately and use the results productively.
4. Can plan, present, or secure professional development activities that contribute to the use of sound assessment practices.
5. Accurately analyses student assessment information, uses the information to improve curriculum and instruction, and assists teachers in doing the same.

6. Can develop and implement sound assessment and assessment-related policies.

7. Creates the conditions necessary for the appropriate use and reporting of student achievement information and can communicate effectively with all members of the school community about student assessment results and their relationship to improving curriculum and instruction.

8. Understands the standards of quality for student assessments and how to verify their use in their school/district assessments.

9. Understands the attributes of a sound and balanced assessment system.

10. Understands the issues related to the unethical and inappropriate use of student assessment and protects students and staff from such misuse."

The ten competencies cover the primary components of a principal assessment literacy as reviewed in the literature above, encompassing classroom level and school level assessment expertise. In addition, they incorporate assessment for learning principles that align with current assessment reform in Indonesia. Therefore, they were used in developing the principal assessment knowledge and skills framework and data collection instrument in this study (see methodology chapter/Chapter Three).

The literature above provided the foundation of the basic concept of principal assessment literacy, defined as the knowledge and skills needed by principals to fulfil their role in student assessment. In this current study, this definition is referred to as the classical concept of principal assessment literacy. In particular, the literature highlighted the link between principal assessment literacy and their role in leading the assessment practices within the school.

Building on principals' broader responsibility concerning the school-wide implementation of student assessment, Absolum, Flockton, Hattie, Hipkins, & Reid (2009) used the term 'assessment leadership' to highlight the principal ability and responsibility to lead a school through changes and in implementing new assessment policy. Renihan & Noonan (2018) also used the term to highlight the link

between principals' responsibility and assessment policy, noting "In the context of assessment reform, there is a strong interest in principals' knowledge and skills concerning assessment leadership" (p. 3).

Morris (2017) defined assessment leadership as follows: "school leaders who establish inquiry-based professional learning environments that promote assessment literacy and employ effective assessment practices with the goal of improving student-learning outcomes" (p. 40). She argued that assessment leadership consists of three elements.

First, inquiry-based learning environment, where a principal fosters assessment literacy of teachers through a culture of inquiry which encourages collaborative learning and shared professional growth. The inquiry-based learning environment helps align principals and teachers in a shared vision concerning effective student assessment practices.

Second, the assessment literacy component of assessment leadership, which include the essential knowledge and skills outlined by Popham (2011). These comprise of understanding assessment purposes, ensuring assessment validity and reliability, adhering to assessment ethics, constructing assessment, aligning assessment with curriculum standards, and using assessment results to improve instruction.

Lastly, assessment practices, which encompass establishing a vision, setting clear learning targets, using assessment data to improve teaching practices, building teachers' assessment literacy and engaging in assessment self-reflection.

A case study of three principals in Singapore (i.e. primary, secondary and junior college schools) by Tay and Tan (2019) used the term when highlighting the principals' student assessment responsibility. They concluded that there are four threshold levels of success a principal needs to apply to be an effective assessment leader: 1. Bring irreversible change by shifting teachers' paradigm toward assessment; 2. Be coherent and integrative by encouraging teachers to develop a coherent assessment framework that benefit students' learning; 3. Being a catalyst

for transforming the direction and value of education through holistic assessment practices; and 4. Be prepared to confront and address potentially discomforting issues. As such, a principal need to make provision in educational assessment (i.e., assessment literacy).

Referring to the literature on the concepts of principal assessment literacy and its relation to principals' responsibility for leading student assessment practices within schools, assessment literacy is a regarded as a prerequisite for principals to be able to play they role as effective assessment leadership (Brookhart & Moss, 2013; Morris, 2017).

2.3 Contemporary concept of assessment literacy

Studies for more than three decades across various countries on teacher and principal assessment literacy have consistently demonstrated that educators' assessment literacy was considered low to moderate (Alkharusi et al., 2011; Kanjee & Mthembu, 2015; Lee-Perry, 2013; Mertler, 2005; Mertler & Campbell, 2005; Morris, 2017; Pfeiffer-Hoens, 2017; Plake et al., 1993; Prizovskaya, 2018; Umam & Indah, 2020; Yamtim & Wongwanich, 2014). These studies are comparable because they used assessment literacy instruments based on a common framework: the 1990 USA-based Standards for Teacher Competency in Educational Assessment of Students.

It is Important to note that the instruments in those studies evaluated educators' general assessment knowledge and skills using test-based approaches (i.e., correct vs incorrect answers) without aligning the instruments with certain school subjects or to participant teaching contexts. Contemporary views of assessment literacy expand the concept, framing it as a context-dependent social practice construct rather than just a set of general knowledge and skills of assessment. The two following sections present recent views of teacher and principal assessment literacy.

a. Socio-contextual dependent teacher assessment literacy

A more recent conceptualisation of teacher assessment literacy incorporates socio-cultural and local context as part of the assessment literacy construct. This idea was introduced by Willis, Adie, & Klenowski (2013), which has since been adopted and expanded by other researchers. Some have applied the concept directly in their empirical studies (Clark, 2015; Coombs et al., 2018; Poskitt, 2014; Schneider et al., 2020; Ye, 2023), while others combined it with an existing assessment literacy framework (DeLuca et al., 2019) or used it as the foundation to develop new assessment literacy frameworks (such as Pastore & Andrade, 2019; Xu & Brown, 2016). This contemporary view emphasises that teachers' assessment decisions and actions emerge from the interaction between their technical knowledge and skills in assessment and social contextual factors, reflecting their assessment literacy.

Willis et al. (2013) believed that assessment literacy is a complex and multi-faceted concept and should be included as a part of learning theory and viewed from the perspective of a sociocultural lens. They defined it as "a dynamic context dependent social practice that involves teachers articulating and negotiating classroom and cultural knowledges with one another and with learners, in the initiation, development and practice of assessment to achieve the learning goals of students" (p. 242).

Willis and colleagues developed the idea of multiple facets of assessment literacy in contemporary assessment practices by applying Bernstein's (1999) theory of vertical and horizontal discourses, which distinguishes between different forms of knowledge as the criteria. Vertical discourse refers to knowledge that has a ''coherent, explicit and systematically principled structure, hierarchically organised … or it takes the form of specialised languages with specialised modes of interrogation and specialised criteria'' (Bernstein, 1999, p. 159). In the context of the current study, the vertical discourse related to assessment within Indonesia's recent educational policy includes national achievement standards, subject-specific disciplinary knowledge and core knowledge of student assessment. These elements form part of the conception of assessment literacy examined in this study.

Horizontal discourse, on the other hand, refers to common-sense knowledge which has the features of local, context dependent and specific, tacit, multi-layered and segmentally organised (Bernstein, 1999). Therefore, since this knowledge is uniquely associated with the context where it is developed, it is difficult to transfer to other settings. Willis et al. (2013) argued that the horizontal assessment discourses align with the concept of assessment literacy from socio-cultural perspective. From this perspective, assessment literacy should be understood differently depending on the "cultural and policy contexts and so is closely associated with theories of learning" (p. 245). Similarly, Poskitt (2014) also emphasised that it should be situated in a framework of sociocultural learning theory, which is constantly negotiated within contextual, social and cultural understanding. In this study, the horizontal discourse reflects Indonesian teachers' awareness of their local contexts including their students, colleagues, local educational policies, and the characteristics of the local society, which is highly diverse across the country in terms of ethnicity, culture, language, as well as levels of prosperity.

An educational system which emphasises the principles and practices of assessment for learning reflects the horizontal assessment discourse as they are more localised and integrating assessment practices with pedagogy and curriculum and represents the sociocultural perspective of learning (Willis et al., 2013). However, teachers cannot disconnect their assessment decision makings and actions from the vertical discourses. As classroom assessment practices moving between horizontal and vertical discourses, teachers need to recontextualise their assessment knowledge into their classroom contexts, having its own characteristics. Therefore, this study considers both the vertical and horizontal discourses of assessment in exploring the teachers' assessment literacy.

Willis et al.'s conceptualisation of assessment literacy and the theoretical foundations underpinning the concept, provide a framework to inform discussions and decisions in educational policy and practices, ultimately striving to improve student learning outcomes. They suggest an alternative view in understanding and evaluating assessment literacy. This perspective is particularly relevant to the

ongoing education reform in Indonesia, especially concerning the assessment policy, which lead the use of this concept in this current study. Arguments of the following three more recent studies reinforce the view of assessment literacy as a multi-faceted and context-specific concept.

Xu and Brown (2016) believed that while teachers are expected to acquire adequate assessment knowledge and skills as outlined in various standards, in practice, those who work in complex environments must make professional assessment decisions influenced by a range of contextual factors that can either support or hinder their practices. Consequently, understanding teacher assessment literacy requires an examination of how their decision makings and actions are shaped by unique dynamic features of their contexts. This idea of socio-contextual factors of assessment literacy was accommodated in the Xu and Brown's (2016) proposed framework of teacher education in assessment literacy called TALiP (Teacher Assessment Literacy in Practice).

DeLuca, Coombs, MacGregor and Rasooli (2019) contended that a contemporary understanding of assessment literacy should account for social-cultural aspects as they shape teachers' actions across various assessment areas leading to differential assessment practices depending on the contexts. In addition, Pastore and Andrade (2019) proposed a three-dimensional model of teacher assessment literacy, comprising of conceptual, praxeological, and socio-emotional components. This model frames assessment literacy as an integration of teachers' knowledge, skills and dispositions, emphasising adaptive competence "that requires the ability to apply knowledge and skills in flexible, creative, and responsive ways in different educational contexts" (p. 134).

b. Socio-contextual dependent principal assessment literacy

In the current Indonesia's era of school autonomy, principals' effectiveness as leaders and managers is partly measured by the ability to appropriately comprehend and translate policy reforms to the unique contexts of their respective schools (Maisyaroh et al., 2021). Day et al. (2010) argued that while basic leadership

practices are necessary, the effectiveness of school principals also depend on selecting and combining methods that align with the specific needs of different school contexts.

While the studies and ideas of the importance of social contextual aspects in comprehending assessment literacy discussed in the previous section relate to teachers, it is plausible that the same conceptualisation is also applicable to the assessment literacy of principals. Charteris and Smardon (2021), for example, specifically used the term sociocultural assessment leadership to emphasise the contextual and situated nature of principal assessment literacy. They argued that principals' assessment leadership is a set of relational dimensions emerging and are influenced by the social and culturally situated contexts of schools and collaboration between leaders, teachers, and students, rather than a set of separated actions of an individual leader.

Renihan and Noonan (2018) focused their research on the impact of school contexts, specifically rural settings, on principal assessment leadership. Their study provided evidence of the principal's understanding of what it means to be an assessment leader and how they perceived the rural contexts as shaping their assessment role. They concluded that there are distinct aspects of the rural contexts that greatly influence the assessment role of the principals, such as grade patterns, parent involvement, social-cultural issues, and heavier administrative burdens, leading to less time for specific leadership functions, including assessment leadership. In addition, most of the rural school leaders agreed that it is the teachers who should bear significant responsibility of assessment practices because the principals have been holding multiple roles as teachers, managers and supervisors as the impact of the rural contexts, which often had limited access to learning resources compared to urban schools. Further, the principals believed that the assessment practices in their schools should be managed collegially and professionally instead of bureaucratically.

Smith and Engelsen (2013), through a case study, reflecting the perspectives of two elementary principals on their leadership role, also highlighted the need of

incorporating school context as part of assessment leadership competencies. They suggested a principal need to build teachers' ownership to the assessment culture of the school by facilitating discussion, conducting trial implementation, and adapting the assessment practices to align with the schools' relevant contexts.

Volante and Fazio (2007) contended that educators often encounter various obstacles when applying their assessment knowledge and skills into practice. The complexity of a context wherein the knowledge and skills of assessment are practiced can restrict their ability from effectively translating the assessment literacy they have learned. Hence, to more accurately evaluate their assessment literacy, it is essential to look beyond assessment technical knowledge and skills and examine how principals and teachers apply assessment policies in their specific contexts.

The emphasis on socio-cultural and contextual factors in understanding assessment literacy highlights the importance of examining actual assessment-related decision makings and actions in particular contexts. This approach provides a more accurate and valid picture of principal and teacher assessment literacy. In this current study, the contemporary view of principal assessment literacy, one that incorporates the social contextual dimensions of assessment practices, is adopted, rather than the traditional perspective that limited to basic concepts and procedures of educational assessment. The literature reviewed in section four in this chapter provide examples how socio-cultural aspects influence assessment practices.

3. Empirical literature on assessment literacy levels

Among the wide range of studies on assessment literacy conducted globally, many were empirical research studies which reveal the levels of teacher assessment literacy in various countries (DeLuca et al., 2019; Xu & Brown, 2016). Most of these studies aimed at measuring assessment literacy levels based on classical concept as defined by scholars such as Stiggins (2001) and Popham (2011).

This is evident in a systematic review study conducted by DeLuca et al. (2016b), who suggested that for over ten years post 1990, there are eight most widely used instruments in studies measuring teacher assessment literacy levels in various countries. These eight instruments primarily focused on measuring assessment literacy levels from traditional frameworks, emphasising general assessment knowledge and skills. A more recent systematic review study (Pastore, 2023) confirmed that contemporary studies on assessment literacy levels continue to rely on the classical concept. Below are the examples of those studies.

All the average assessment literacy levels presented here were put in a scale of one hundred (100) to permit comparisons between these studies and later with the findings of this current study. This was calculated by dividing the average scores of the assessment literacy with the maximum possible scores and multiplying the results by 100. For example, in Alkharusi et al.'s study (2011), the assessment literacy average scores of the in-service and pre-service teachers were 12.55 and 15.3 respectively. By dividing them with the maximum possible score from the instrument (i.e., 35) and multiplying by 100, the comparable average scores now are 35.8 and 43.7 respectively.

Alkharusi et al. (2011) investigated the differences in knowledge of, perceived skill in, and attitude toward assessment between pre-service and in-service of Omani teachers. They found that even though in-service teachers perceived skills and favourable attitude toward assessment were higher than pre-service teachers, their average assessment knowledge level (35.8) was lower than that of pre-service counterparts (43.7). In Thailand, Yamtim and Wongwanich (2014) conducted a mixed methods study, revealing primary school teacher assessment literacy and to gain insight regarding an appropriate approach to develop classroom assessment literacy for primary school teachers based on the teachers' opinion. The findings showed that primary teachers had an average assessment literacy of 48.9. Additionally, Umam and Indah (2020) reported that 19 English for Young Learners (EYL) teachers participated in a case study in Bogor province, Indonesia, had an average assessment literacy score slightly below 60.

In China, a study by Dunlai and You (2015) reported that 39 middle school Chinese EFL teachers from 14 provinces had an average assessment literacy score of 33.22 from maximum 100 score (33.22). Similarly, Ara and Saeed (2020), investigating 500 Pakistani primary teachers, revealed that their assessment literacy level was 50.3.

Studies concerning the level of teacher assessment literacy employing classical concept of assessment literacy were also found in other parts of the world. For example, Kanjee and Mthembu (2015) investigated foundation phase teachers' assessment literacy from three schools in South Africa. The results indicated that nearly half of the teachers achieved basic levels of assessment literacy, while the rest were at below basic levels. Using a scale of 100, they defined the basic level of assessment literacy as the scores ranging from 35-74.

Move further to west of the globe, where Pfeiffer-Hoens (2017) examined the assessment literacy levels of teachers and school leaders from elementary public charter schools in Tennessee and California US. They found that the average assessment literacy level of 58 teachers and five principals was 57. They also reported that among seven assessment literacy dimensions, Using Assessment Results dimension was the highest, scoring 73, while Recognising Unethical Practices was the lowest scoring dimension (28). A study measuring the assessment literacy levels of teachers across different subjects and school levels assignment was conducted in New Jersey US (Prizovskaya, 2018). She reported that the average assessment literacy scores of the teachers were 51, with the highest score of 61 for Choosing Appropriate Assessment Methods dimensions and the lowest score of 39 for Developing Appropriate Assessment Method dimension.

Collectively, these data suggest a consistent trend of low to moderate assessment literacy levels among teachers across time and different educational settings in Asia and beyond.

4. Empirical literature on factors affecting assessment literacy

In addition to examining the levels of principals' and teachers' assessment literacy and their role in enhancing better student learning and outcomes, it is also important to investigate which factors contribute to the educators' assessment literacy. The following sections review past studies on some factors that affect teachers' and principals' assessment knowledge and skills, including assessment-related experiences, assessment courses undertaken in higher education, participation in in-service teacher professional development programmes, and the impact of local support on assessment literacy and practices. The review serves as the foundation for understanding the relationships between these factors and assessment literacy.

4.1 Assessment-related experiences

In this study, assessment-related experiences include principal and teacher experience in teaching, serving as an instructor in assessment-related professional learning and development activities, and contributing as test or assessment item writer. Teaching experience is considered relevant because student assessment is an integral part of curriculum and pedagogy (Livingston & Hutchinson, 2017). Regular engagement with assessment activities throughout the teaching career is argued to develop teacher assessment expertise (Yan et al., 2018). Hence, it is hypothesised that more teaching experience leads to higher level of assessment competence. Additionally, experiences as an assessment instructor/facilitator and as a test items writer are also expected to associate with assessment literacy, for the experiences in these roles reflect recognition of expertise in student assessment or evaluation or at least in some specific aspects of it. Yan, Zhang, and Fan (2018) argued that experiences in test item writing may hone teachers' intuitions in assessment concepts and principles.

Numerous studies have highlighted the link between the length of teaching experience and assessment literacy (DeLuca et al., 2016b; Edwards, 2017; Livingston & Hutchinson, 2017; Yan et al., 2018; Zolfaghari & Ashraf, 2015). The following three

studies, for example, investigated the relationship between experience and assessment literacy. Zolfaghari and Ashraf (2015) conducted a study to explore the relationship between assessment literacy and years of teaching experience and the age among Iranian EFL teachers. Using one-way ANOVA (Analysis of Variance) technique, they revealed significant differences in assessment competencies among teachers with varying category levels of experience (1-5 years, 6-10 years, and above 11 years). Specifically, teachers with extensive experience demonstrated a greater proficiency in assessment-related practices compared to those with moderate or limited experiences. This finding suggests that teaching experience had a positive impact on assessment literacy of Iranian EFL teachers.

A new assessment literacy instrument called Approaches to Classroom Assessment Literacy Inventory (ACAI) was developed to investigate teachers' approaches to classroom assessment (DeLuca et al., 2016a). Unlike other assessment literacy questionnaires that focus on measuring teacher assessment knowledge and skills, the ACAI evaluates how teachers approach assessment practices. Three categories of assessment approaches were developed, associated with summative, formative, and individualised approaches to assessment respectively. The results showed a positive association between years of teaching experience and assessment approaches, especially in the area of assessment fairness. More experienced teachers were found to use assessment data more to adjust student learning compared to the less experienced peers. Perceived skill in assessment also varied by career stage, where more experienced teachers reported greater perceived assessment skills than their less experienced counterparts. The finding regarding perceived skill is consistent with the results of Alkharusi et al. (2011), who observed similar trends in their study.

Metler (2005) compared the assessment literacy of pre-service and in-service teachers using Classroom Assessment Literacy Inventory (CALI). The study involved 67 pre-service undergraduate students in science and social studies from Midwestern University and 101 in-service teachers from nearby school districts. The results showed that the in-service teachers had higher level of assessment literacy than their pre-service counterparts, indicating that teaching experience plays an

important role in developing assessment literacy. Using similar 1990 US teacher assessment standard-based instruments, several studies yielded consistent results that in-service teachers scored higher than pre-service counterparts (Mertler & Campbell, 2005).

Different findings regarding the impact of teaching experience on teachers' assessment literacy were reported in the studies of Dunlai and You (2015) and Alkharusi, Kazem and Al-Musawai (2011). Dunlai and You found that the teaching experience of EFL middle school teachers from 14 provinces in China did not affect their assessment literacy level. Similarly, a cross-sectional study of Omani teachers by Alkharusi and colleagues showed that, while in-service teachers had higher perceived skills and favourable attitudes toward assessment, their assessment knowledge was lower than that of pre-service teachers. When comparing in-service teachers, those with less teaching experience had higher assessment knowledge than those with more experience. This knowledge gap might be caused by pre-service teachers having recently completed educational measurement courses, keeping their knowledge more up to date. In addition, almost half (100 out of 217) of the in-service teachers admitted that they had not taken any educational measurement courses in their pre-service training.

The mixed findings between Zolfaghari and Ashraf (2015), DeLuca et al. (2016a) and Mertler (2005) on one hand versus Alkharusi et al. (2011) and Dunlai and You (2015) on the other hand show that while teaching experience might contribute to assessment literacy, its impact is likely combined with other factors such as assessment-related courses taken at universities. The interplay between teaching experience and formal assessment education in underpinning teacher assessment literacy is evident in a study by Ogan-Bekiroglu and Suzuk (2014). They compared Turkish pre-service physics teachers' level of assessment literacy using responses from both an assessment literacy questionnaire and open-ended questions upon completing an assessment course. They also investigated how the participants integrated theoretical and practical assessment knowledge through an assessment project. The findings revealed that although pre-service teachers grasp assessment key concepts after the course, they struggled to apply them in practice. Among the

dominant reasons, as revealed from the open-ended responses, was that they lacked experience in using diverse assessment methods.

The following section reviews studies that investigated the impact of assessment courses in higher education on assessment literacy.

4.2 Assessment-related courses in higher education

DeLuca et al. (2016b), through their systematic review of various assessment standards across countries, found a growing awareness of the importance of providing teachers with opportunities to cultivate their assessment competency, given the emerging emphasis on assessment for learning. A fundamental assumption underlying this is that assessment education should cultivate educators' assessment literacy. Research studies indicate that assessment-related courses in university lay the foundation of assessment capacity among newly graduated pre-service teachers or strengthen the assessment knowledge and skills of practicing teachers.

A qualitative case study research by Rogers et al. (2020) investigated novice teachers' assessment literacy and practices, during the transition period from completing their teacher preparation programmes through the first year of teaching, focusing on formative assessment. Data were gathered through performance assessment documents at the end of teacher preparation including lesson plans, 15-minute teaching videos, students' work samples, as well as interviews and observations during the first year of service. This study showed a strong alignment between teacher preparation programmes and the assessment practices of novice teachers in their initial teaching year.

Deneen et al. (2016) investigated the impact of a graduate level teacher education courses in assessment on participants' conception of assessment and assessment literacy. The participants of this study were 32 students consisting of a nearly equal number of pre-service and practicing teachers enrolled in a 12-week course on assessment. Teacher assessment literacy in this study was measured by the course

assessment results, while teacher conception of assessment was revealed by interviewing the participants. The interviews were conducted twice for each of the selected participants, at the beginning and the end of the course, to examine the changes in the conception of assessment. The findings highlighted that teacher skills and knowledge enhanced after participating in the assessment course, although teachers' conception of assessment remained largely unchanged. The researchers argued that short term assessment education is insufficient to alter teachers entrenched negative perception shaped by their long personal experience with negative assessments dating back to their childhood. A previous study by DeLuca, Chavez and Cao (2013) supports Deneen et al. (2016) argument as DeLuca et al.'s study showed that a six month long assessment course (twice as long as that in Deneen et al.'s study) was able to shift pre-service teachers' conception of assessment.

Using self-rated levels of confidence in assessment literacy, DeLuca et al. (2013) investigated pre-service teacher confidence levels in student assessment and their evolving conceptions of assessment during a semester-long assessment education course. The participants' open-ended responses were analysed through deductive thematic analysis based on their primary and secondary conceptions of assessment as testing, format, purpose, and process. The changes in participants' confidence levels were analysed through exploratory factor analysis. To determine changes over time and the interaction between participants' assessment conceptions and their confidence, the authors conducted a repeated-measures analysis across both questionnaire administrations after converting the coded responses of assessment conceptions to numerical values. The results indicate, there was a significant shift in teacher candidates' conceptions of assessment from being predominantly testing focused toward a more integrated understanding of varying formats and purposes of assessment, emphasising more on process of assessment. Additionally, the pre-service teachers demonstrated a marked increase in confidence in their assessment capabilities.

Alkharusi et al. (2011) investigated the knowledge on educational measurement of Omani in-service teachers, comparing those who had attended pre-service

educational measurement courses with those who had not. The results showed that while no differences were found in their self-perceived assessment skills, in-service teachers who had pre-service assessment education exhibit higher level of assessment knowledge than those who did not.

While many studies identified the positive impacts of assessment-related education in universities, either in pre-service teacher education programmes and at postgraduate level, on teacher assessment literacy, some researchers also highlighted shortcomings, suggesting that university-based assessment trainings may not prepare student teachers effectively with necessary assessment knowledge, skills, beliefs or practice for various reasons. For instance, the assessment courses may lack crucial elements necessary for teachers' understanding and practical applications, such as reporting achievement and writing constructed-response items (DeLuca & Klinger, 2010), beliefs and values of assessment (Looney, Cumming, Kleij & Harris, 2018), or skills in external evaluation (i.e., large scale assessment) and integrating assessment into instruction (Schneider & Bodensohn, 2017). The misalignment between assessment courses content/syllabi at university with assessment-related policy and assessment standards also has been noted as the contributing factors to the in-effectiveness of some assessment courses (DeLuca & Bellara, 2013).

In addition, DeLuca, Valiquette, Coombs, LaPointe-McEwan, and Luhanga (2018), investigating teachers' approaches to classroom assessment, found no differences in assessment approaches between teachers with and without previous assessment education, where most among the 404 teachers in primary and secondary teachers surveyed across the United States and Canada approached classroom assessment formatively. The study of DeLuca et al. (2018) showed that assessment training in higher education had little or no impact on teacher assessment practices.

For practicing educators, beside formal assessment courses in university, in-service professional development programmes serve as an alternative avenue for enhancing assessment literacy. The following section presents studies that

discussed the impact of in-service professional development on assessment literacy.

4.3 Assessment-related in-service professional development

Researchers asserted that practicing educators need to engage in professional development programmes or have continuing professional learning activities to maintain and strengthen their assessment literacy (Arrafii, 2021a; Livingston & Hutchinson, 2017; Xu & Brown, 2016). The growing focus on assessment for learning, which highlights the integration of assessment with pedagogy and the use of assessment data to inform daily instruction, underscores the importance of providing teachers with opportunities to enhance their assessment competence as a key element in becoming effective educators (DeLuca et al, 2016a). The following studies have pointed out the contribution of in-service professional development programmes on educators' assessment literacy and related practices.

Bostrom and Palm (2020) investigated how expectancy-value theory can explain changes in teachers' practices after participating in professional development programmes (PDP) on formative assessment and examined its differential effects of the programmes on different groups of teachers. The results showed that year-7 teachers experienced increased motivation to adopt more formative assessment practices, mediated by the increase in their high expectancy of success and value beliefs. However, year-4 teachers demonstrated even higher level of expectancy of success, value beliefs and motivation, leading to more significant changes in assessment practices. The disparities between the two groups were attributed to three factors: first, differences in PDP delivery modes, with the year-4 teachers benefiting from greater exposure to expert practical suggestions and colleagues' sharing of success. Second, the difference in the school context, as the year-7 teachers faced additional burden to prepare their students on national exams. Thirdly, discrepancies in level of support from principals, which influenced teachers' motivation and ability to implement formative assessment effectively. These

findings show the importance of programme delivery approaches, contextual factors, and the organisational support in determining the effectiveness of the PDP.

Randel, Apthorp, Beesley, Clark, and Wang (2016) evaluated the impact of Classroom Assessment of Student Learning (CASL), a self-implemented form of professional development programme, on teachers' assessment knowledge and student achievement in mathematics. CASL provided sources including textbook, DVDs, ancillary text, facilitator handbook for approximately one academic year in a real-life situation. While the programme significantly improved teachers' knowledge of assessment and increased student involvement in assessment activities, it had no statistically significant impact on teachers' assessment practices or student achievement in mathematics. The researchers attributed the results to moderate level of implementation fidelity, which was influenced by contextual factors such as competing school district initiatives, student and teacher attrition, and variability in programme adoption across schools. Both Randel et al. (2016) and Bostrom and Palm (2020) underscored the critical role of the school context in shaping the effectiveness of in-service professional development programmes.

An effective professional development programme can change a teacher's beliefs and practices of assessment. Saputra et al. (2020) demonstrated how a six-month professional learning project for secondary EFL teachers in Indonesia transformed their assessment beliefs and practices. Weekly discussions and observation shifted teachers' perspective from viewing assessment as merely for grading and measuring student achievement to recognising its potential for facilitating student motivation to learn, improving learning outcomes, and informing more effective teaching strategies.

Effort to maximise the impact of professional development programmes or professional learning activities on assessment literacy has led to the exploration of innovative approaches. Lysaght and O'Leary (2017) proposed a small-scale site-based form of teacher professional development (PD) in assessment for learning using Design-Based Implementation Research (DBIR). The collaborative PD programme involved teachers, researchers and policymakers to address specific

assessment contextual challenges while fostering mutual professional development. For teachers, DBIR ensures that PD is tailored to their specific problems faced on-site. For researchers, DBIR may synchronise their community service role and research role, and for policymakers, DBIR provides information about the challenges and opportunities of educational policies being implemented. Lysaght and O'Leary's (2017) findings also illustrate how unique school context may affect the effectiveness of professional development programmes.

Similar to Lysaght and O'Leary study, Christoforidou and Kyriakides (2021) proposed teacher professional development called Dynamic Approach (DA), which was based on the notion that a teacher professional development (TPD) should not only enhance teacher's skills but should also have positive impact on student's learning outcomes. The DA consists of four steps. First, designing a TPD that is tailored to fulfil teachers' specific needs. Second, developing an active collaboration between the teacher trainers and the participating teachers, acknowledging teachers taking ownership of their own learning and development. Third, establishing a formative evaluation throughout the process, allowing a continuous review and improvement for both the intervention and implementation. Fourth, identifying the impact of the intervention on teachers' skills and the indirect effects on student learning outcomes. The researchers found that DA approach on improving teachers' formative assessment competencies, not only enhanced teachers' assessment skills but also positively impacted student learning outcomes.

Student learning outcomes were also targeted through a teacher professional development initiative in New Zealand (Poskitt 2014). Poskitt investigated a New Zealand government initiative of teacher Professional Learning and Development (PLD), a combination of professional learning and professional development activities, to enhance teacher assessment literacy. Using Guskey's (in Poskitt, 2014) framework, the study evaluated 5 levels of a PLD impact: participants' reactions, participants' learning, organisational support and change, participants' use of new knowledge and skills, and student learning outcomes. The results indicated a substantial positive outcome across all levels of evaluation.

Brookhart and Moss (2013) emphasised the importance of school principal involvement in professional development programmes. Their research revealed that principals who immerse themselves in a PDP on formative assessment were better equipped to provide teachers with more specific feedback and actionable information and develop a more productive collegial conversation with teachers. In contrast, Morris (2017) in her doctoral research found that in-service professional development trainings had minimal impact on school leaders' assessment leadership. She attributed this to the shortcomings in delivery approaches of the training programmes, which limited their effectiveness.

Beyond the direct impact of in-service professional development programmes (PDP) on educator assessment literacy and practices, the studies above also underscored the significance of contextual factors such as peers and principal support in enhancing the effectiveness of in-service PDP as well as the assessment literacy and practices. The following section delves deeper into the impact of support from colleague teachers and school principals on assessment literacy and practices.

4.4 Local support from colleagues

Local support in the form of collaboration relationships with colleagues, school principal and other members of education community is a contextual aspect that affects educators' continuous capacity growth, including in the area of educational assessment (Pastore, 2023). The literature showed that collaboration among teacher colleagues, participation in professional learning communities, the support from school leaders and school superintendents contributed to the development of assessment literacy as well as in the implementation of assessment reform.

Rezagah (2022) investigated the impact of a newly introduced educational policy, mandating the use of assessment for learning, on Iranian EFL teachers' assessment literacy. The study found that support from and collaboration with school colleagues, through discussion, strategy exchange, feedback sessions with colleagues, improved teachers' assessment literacy in several ways. The

improvement includes navigating between the demand of assessment for learning and assessment of learning, increasing reliability of assessment, promoting fairness and inclusivity in assessment, and building feedback strategies. In addition, the collaboration forms also enhanced practical knowledge of assessment design, focusing on aligning assessment with curriculum goals and classroom practices.

Livingston and Hutchinson (2017) discussed the development of Scottish teachers' assessment capabilities through career-long professional learning. They suggested that educator professional learning communities (PLC) are particularly effective in fostering assessment literacy by enabling the educators to engage in reflective conversations, analyse students' work and collaboratively refine their assessment strategies. However, the effectiveness of professional learning community meetings is contingent upon their organisations and focus. For example, Umam & Indah (2020) reported that a group of EFL teachers in one of the cities in Indonesia reported limited improvement in their assessment literacy as the PLC meetings lacked organisation and clear objectives.

Webber et al. (2013) conducted a large scale, two-year mixed-method study in Alberta Canada concerning student assessment and student learning outcomes, involving more than 3,300 education stakeholders (students, parents, members of teacher unions, school leaders, members of school board, members of school council, local education authorities, and Ministry of Education personnel). One of the important findings was that it is essential for principals to operate with a robust foundation of assessment knowledge and engage actively with their school communities. The study also pointed out to the role of district superintendents in fostering assessment literacy. The superintendents, alongside school principals, were found instrumental in educating parents and the community about the value of assessment as well as ensuring that assessment literacy and instructional improvement remain priorities at the school district level.

A qualitative study by Hollingworth (2012) revealed an assessment leadership strategy of a principal in supporting the implementation of formative assessment. The principal active participation in the same professional development activities as

teachers, not only enhanced his assessment literacy but also enabled him/her to identify directly the specific supports the teachers needed during the transition to new assessment practices. This direct involvement fostered collaborative culture and facilitated the successful implementation of formative assessment.

5. Hypothesised model of the relationship between assessment literacy and the predictors

The review of existing empirical literature on factors affecting teacher and principal assessment competency shows varied impact of the four factors: assessment-related experiences, assessment coursework in higher education, in-service professional development programmes in assessment, and assessment-related local support. Pastore (2023) through her systematic review argued that teacher capacity in student assessment is growing overtime and influenced by the four variables in this study. This present study further examines the influence of these factors, providing contextual insight from Indonesia, particularly regarding the dynamic relationships among the constructs in the model.

This study puts into context the impact of the hypothesised four predictors on assessment literacy through both direct and indirect relationships. All predictors were hypothesised to have direct impact on assessment literacy. In addition, two predictors, in-service trainings in assessment and assessment-related experience, were also hypothesised to mediate the impact of assessment courses and local support on assessment literacy respectively. In other words, it was also hypothesised that assessment courses and local support have indirect impacts on assessment literacy through the mediating effects of in-service trainings and experience. The hypothesis statements of the model are as follows.

H1: Assessment courses → assessment literacy (Courses → AL)

H2: Assessment courses → In-service trainings in assessment (Courses → Training)

H3: In-service trainings in assessment → assessment literacy (Training → AL)

H4: Assessment courses → assessment-related experiences (Courses → Experience)

H5: Assessment-related experiences → assessment literacy (Experience → AL)

H6: Assessment-related local support → assessment literacy (Local support → AL)

H7: Assessment-related local support → In-service trainings in assessment (Local support → Training)

H8: Assessment-related local support → Assessment-related experiences (Local support → Experience)

H9: Assessment courses → In-service trainings in assessment → assessment literacy (Courses → Training → AL)

H10: Assessment courses → Assessment-related experiences → assessment literacy (Courses → Experience → AL)

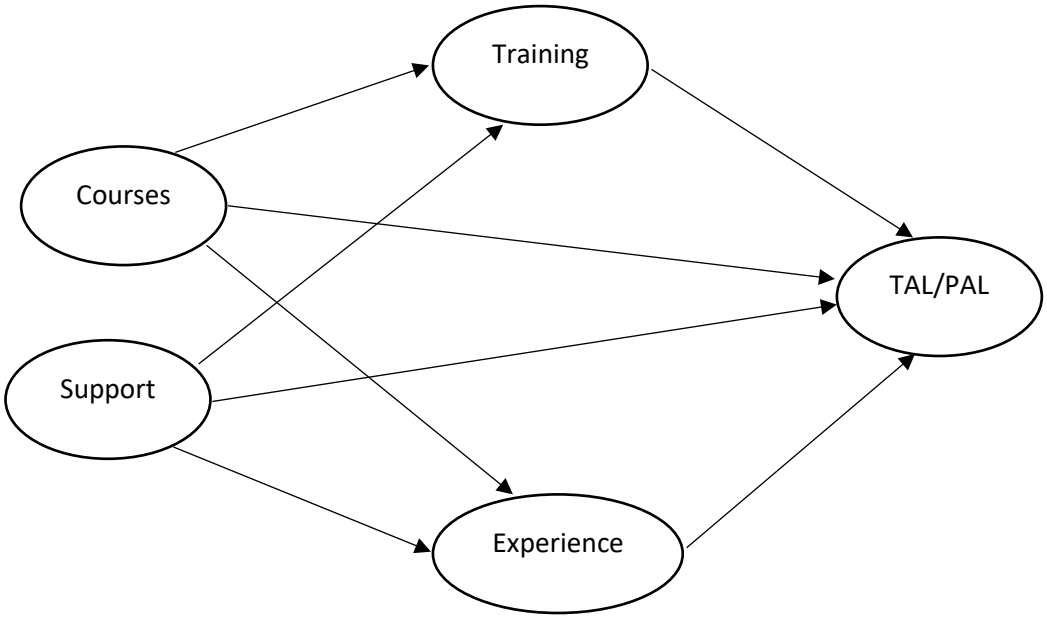
H11: Assessment-related local support → In-service trainings in assessment → assessment literacy (Local support → Training → AL)

H12: Assessment-related local support → Assessment-related experiences → assessment literacy (Local support → Experience → AL)

The following diagram depicts the hypothesised causal relationship between assessment literacy of principal and teacher and the four predictors.

Figure 1

Hypothesised Causal Relationship Between Assessment Literacy and the Four Predictors



Note. TAL = Teacher Assessment Literacy. PAL = Principal Assessment Literacy

The growing recognition of the impact of local contexts on assessment literacy and practices necessitates an investigation of assessment policy implementation in different school contexts to gain deeper understanding of teacher and principal assessment literacy. Consequently, this study also investigates teachers' and principals' implementation of new assessment policy within their school settings, acknowledging the distinct challenges faced by schools. The following section reviews empirical literature concerning policy implementation and assessment reform enactment across countries to frame comparative analysis.

6. Empirical literature on assessment policy enactment

This section examines empirical literature related to the enactment of assessment policy and reform in diverse educational contexts in various parts of the world, focusing on how assessment literacy is influenced by social, cultural, and contextual factors. Drawing on the contemporary view of assessment literacy as a negotiated multifaceted and context dependent social practice, rather than limited to a set of technical knowledge and skills of assessment, the second phase of this study looked further on the level of assessment knowledge and skills of the principals and social

studies teachers, measured in phase one, by exploring their enactment of the assessment reform within their schools. Therefore, the enactment of the assessment policy is crucial in understanding principals' and teachers' assessment literacy. Roger, Reagen, and Ward (2020) believes that educators enhance their assessment literacy through actively engaging in collaborative school environment where they engage with students, colleagues, leaders, and parents in discussing and using assessment tools, goals, beliefs, and strategies. The studies reviewed below illustrate this dynamic in various contexts.

Leong et al. (2018) conducted a review of studies regarding the conceptualisation and implementation of assessment for learning (AfL) in East Asian countries, including Southeast Asian countries such as Indonesia. They suggested that the conceptualisation and practice of assessment for learning may potentially be affected by the local cultural values and beliefs of the context where it is enacted, and as the consequence may have distinctive elements compared to the concept and theory where the idea initially emerged. For example, Azis (2015) found that teachers in Gowa Region of Indonesia valued grading and scoring as the important part of learning and they considered that the practice was necessary to improve student learning and thus was part of the idea of assessment for learning. This practice might reflect Leong et al's (2018) argument that AfL practice is highly situated and contested. This suggests that local values can blend with or reshape assessment reform to suit specific educational contexts.

Merely possessing adequate assessment knowledge and skills does not guarantee educators can fully practice assessment effectively (Xu & Brown, 2016), as contextual factors influence assessment practices (Fulmer et al., 2015). For instance, this was evident in a study by Ara and Saeed (2020) regarding the assessment literacy and practices of assessment for learning of Pakistani primary teachers. They found a weak relationship between the teachers' level of assessment knowledge and practices of assessment for learning. In addition, even though differentiated teaching has been a prominent strategy in Singapore recently due to the increasing student diversity, 'assessment for learning' practice, which normally is a two-sided coin with differentiated teaching approaches, remains underutilised

within classrooms for most students, educators and wider community put a high value on examination results (Heng, Song & Tan, 2021). Heng et al. demonstrated that educators' and other school stakeholders' paradigm on assessment plays a significant role in determining how assessment is understood and practiced.

Oo et al. (2024) conducted a systematic review of 28 studies from various countries on school-based assessment reform implementation. They analysed the approaches used to implement the reform and factors that affect the success. They identified diverse strategies of the implementation approaches such as shifting from summative to formative assessment, incorporating performance-based assessment, and using structured frameworks to guide teacher assessment practices. These approaches were claimed to be effective in enacting assessment reform as well as in enhancing assessment literacy. Seven critical factors were identified related to the success of implementation namely teacher disposition and literacy, principal leadership, student engagement, parental involvement, stakeholder collaboration, school cultural context and supportive policy. This finding underscored the multifaceted nature of assessment reform implementation and the necessity of addressing contextual and systemic factors.

The impact of supportive policy, especially from local education authority, was also evident in a study by Zulaiha et al. (2020), who conducted a study on Indonesian junior secondary EFL teachers' assessment literacy. One of the findings highlighted that even though the teachers had a good grasp of assessment for learning principles and were aware of its positive impact on student learning outcome, their practice did not reflect the knowledge as they tended to focus on grading to meet the demand of local policy of student minimum standard mastery levels.

School-level assessment policy alongside principal assessment leadership also potentially brings successful implementation of a school assessment reform. This was evident in a two-year project on assessment reform conducted in Queensland Australia (Adie et al., 2021). The project aim was to shift assessment culture and practice away from summative assessment and high-stakes testing toward feedback-only assessment without grades, which prioritise student learning and

well-being. This project was highly challenging as the assessment innovation was against the existing high-stakes accountability system and community expectations which value summative assessment and grades. To promote the shift in assessment culture, the school leaders involved the teachers in the planning and design of the new assessment practices, which foster the sense of ownership among the teaching staff. The school also organised teacher capacity building through ongoing professional development activities (workshops, training sessions, and discussions) and collaborative teacher meeting where they discuss their experiences, share resources or collaborate on best practices to implement new assessment strategies. The leaders also fostered a culture of trust and risk-taking where teachers feel trusted and supported to trial new approaches. The school also engaged parents to create a shared vision and expectation. They did the latter by organising an information session and clearly communicate the assessment practices and explained the rationale behind focusing on learning and reducing test-related stress. It is worth noting that this project was undertaken in a high performing school.

Herzog-Punzenberger et al. (2020) explored secondary school teacher assessment practices in culturally diverse classrooms across Austria, Ireland, Norway, and Türkiye. The primary dimension of cultural diversity identified by the teachers was proficiency in language of instruction. This led the teachers to adapt their assessment strategies to address the linguistic challenges such as by providing additional time, alternative formats, or support materials. In addition, teachers used other strategies such as formative assessment, peer assessment, and tailored grading practices to accommodate diverse students' need. However, the teachers' initiatives were mostly individual effort rather than a school-wide policy. The study also highlighted that the government high-stakes assessment regimes which exist in these countries limited the effectiveness of the teachers' culturally responsive assessment practices.

Choi (2017) investigated South Korean elementary teachers' responses to educational policy mandating the use of performance assessment and the application of higher-order thinking. She found varying degree of implementation

among the teachers: profound implementation (integrating the assessment reform effectively), transitional implementation (mixing the traditional and mandated assessments reform approaches), superficial implementation (complying with the surface level of the assessment reform) and reluctant implementation (complying minimally due to reluctant attitude). Those in the superficial and reluctant groups were struggling with the deeper aspects of the reform such as higher order thinking and using assessment for learning. She underscored that supporting policy from national level to school level are crucial in promoting the assessment of reform, such as providing relevant professional development programmes.

The Indonesian teachers' and principals' assessment knowledge and skills level measured through the questionnaire in the first phase of the current study as well as the factors that influence the assessment knowledge and skills, complemented with the exploration of the assessment reform enactment in the second phase delineated a more comprehensive picture of the educators' assessment literacy. In addition, given the highly complex differences among Indonesian regions and schools across the country, which may affect the effectiveness of assessment reform as mandated by the *Emancipated Learning* policy, investigating the enactment of the assessment transformation effort in diverse educational contexts in Indonesia becomes more relevant. Previous studies concerning formative assessment implementation in previous curriculum (2013 curriculum) have been conducted in different regions in Indonesia.

A case study in a rural area of Lombok Island in Indonesia about EFL teachers' views and practices of mandated formative assessment highlighted three characteristics of classroom-based assessment (CBA) of the teachers: first, that assessment of students' behaviour and attitudes are central to the CBA, second, the practice of teacher-centred approach remained, and thirdly, students were hold accountable for the assessment results (Puad & Ashton, 2020). Despite the curriculum policy have suggested that holistic approach includes the assessment of student's behaviour and attitudes, the practices were problematic and raised issues about assessment validity (Popham, 2014). As admitted by the teachers, their subjective

judgment of favourable or otherwise inappropriate behaviours would affect (increase or decrease) the students' final grade.

The second characteristics in Puad and Ashton's study was related to the view that only teachers hold the knowledge authority and thus are capable of judging students' capabilities. Such view is also common in Indonesian neighbouring countries such as the Philippines (Cagasan Jr. et al., 2016) and Malaysia (Sardareh, 2016) as well as in many East Asian countries (Leong et al., 2018). As the consequence, they considered student peer- and self-assessment, two of the main features of assessment for learning, as ineffective since students are perceived as lacking knowledge to do so. Regarding the third characteristics, the teachers believed that students' academic achievement reported is merely the evidence of student capacity and effort and it depends on the student's and their parents whether to improve it or not.

Arrafii (2021a) investigated barriers and opportunities teachers from rural area had in promoting assessment for learning in their classrooms in West Nusa Tenggara, one of the provinces in the eastern part of Indonesia. The findings show that the barriers outweighed the opportunity and were mostly related to limited learning resources and student and teacher readiness. Arrafii also pointed to lack of government support and parents' lack of concern due to high economic pressure as another identified barriers. Similarly, teachers from urban and more developed region than those in Arrafii's study also experienced readiness issues related to teachers' understanding in enacting complex classroom assessment tasks (Umam & Indah, 2020) despite they have more access to various learning resources.

In another study, Arrafii (2021b) documented evidence that EFL teachers in West Nusa Tenggara Indonesia valued and practiced both orientation of assessments, learning and improvement (i.e., reflecting assessment for learning) and curriculum-driven orientation (i.e., reflecting assessment of learning). However, their emphasis on learning and improvement was stronger than the focus on curriculum-driven practices. This finding is understandable as although the teachers recognised the benefits of implementing assessment focusing on learning and improvement

orientation for students learning, they were also under pressured to prepare their students to succeed in the national examination. Similar circumstances were also faced by teachers in many East Asian and Southeast Asian countries (Leong et al., 2018).

Most studies on Indonesian assessment practices have focused on the previous curriculum (2013 curriculum). These studies provided valuable insights but may not fully reflect the current dynamic under the *Emancipated Learning* policy. The current study will bring new insight into assessment practices in Indonesian schools under the new educational policy covering broader area of Indonesia.

Having reviewed the existing relevant literature concerning the classical and contemporary conceptions of assessment literacy, the affecting factors, and the enactment of assessment policy/reform in various countries, the following section presents the theoretical frameworks used to interpret the findings of the research.

7. Theoretical framework

This section discusses three theoretical frameworks informing this research: social constructivism theory, theory of practice and multiplicity of enactment. The social constructivism theory (Vygotsky, 1978) highlights the significance of social interaction and cultural context in the construction of knowledge and understanding. This theory closely aligns with the contemporary concept of assessment literacy as the main interest of this study, which similarly emphasises the importance the social-cultural context in shaping assessment literacy.

The theory of practice (Bourdieu, 1977) and multiplicity of enactment (Buxton et al., 2015) were selected because the modern conception of assessment literacy extends beyond individual general assessment knowledge and skills to incorporate professional and social practices as another key dimensions of the concept. The two theories provided frameworks to help understand the principals' and teachers' assessment literacy through the enactment of the assessment reform in their respective schools. The basis of the two theories on the interplay between human

agency and structural features of a particular context in shaping one's practices and behaviour make them highly relevant with the current study.

While both theories similarly draw on the idea of agency-structure dialectic in developing their concepts, they have distinct emphases in viewing and understanding variations in human actions and decision-making. Theory of practice places greater emphasises on understanding the underlying reasons or causes behind differences in people decision making and choices of actions. On the other hand, the multiplicity of enactment theory uses the structure-agency dialectical relationship leading to educators' differential practices to advocate for an alternative paradigm that recognises the complexity of teachers' educational decisions and practices and challenge the prevalent assumption of strict programme adherence and fidelity of implementation.

Together, these three theories provide the lens through which to interpret the study's findings, revealing the principals' and teachers' assessment literacy. The following sections explain the main concepts of the theories and its application on the findings of the investigation.

7.1 Social Constructivism

The contemporary view of assessment literacy as a social-contextual dependent professional competence and social practice is inherently shaped by a social and contextual environment, having its own "reservoir" (Bernstein, 1999 p. 159), which refers to total sets of social, cultural, and linguistic resources of a society that influence how knowledge is contextualised. This view aligns closely with Lev Vygotsky's social constructivism theory, which posits that knowledge and understanding develop within social environment and cultural contexts through social forms and processes (Adams, 2006).

Vygotsky (1978) emphasised that social interaction is essential in human cognitive development. He argued that learning is fundamentally a social process where knowledge is co-constructed through interaction with others in forms such as

collaboration, dialogue, and shared activities. As social interaction, dialogue and collaboration are the heart of cognitive development and knowledge construction, language as a cultural product, becomes the crucial element of the theory.

In the context of this research, the principals' and teachers' interaction with colleagues, leaders, and other schools' stakeholders are argued to influence the development and state of their assessment literacy. This study explored the interaction among the leaders and educators which occurred both within and beyond school buildings through either informal and spontaneous dialogues and discussions, less formal professional learning sessions such as lesson study activities and workshops, or formal professional development programmes.

Vygotsky also introduced a concept he called zone of proximal development (ZPD), which he defined as "the distance between the actual development level as determined by independent problem solving and the potential development as determined through problem solving under adult guidance or in collaboration with more capable peers" (Vygotsky, 1978, p. 86). In essence, this concept refers to the gap between the cognitive state learners can achieve in isolation compared to what they can accomplish with support from and in collaboration with more knowledge others. Vygotsky argued that knowledge is most effectively acquired through interaction with those who possess greater expertise and by utilising scaffolding as the tool. This study draws on the concept of ZPD to interpret its findings, particularly in elaborating how the assessment literacy of the principals and teachers improves through collaboration and shared activities with the presence of more capable colleagues.

Another core concept in social constructivism theory is the context dependent nature of knowledge. This theory acknowledges that different cultures and society construct knowledge differently. The historical, cultural, and social context influence how people interpret experiences and construct understanding. As such, what works in one context may require adaptation and adjustment to work in another context. In this study, this principle helps explain how different regions in

Indonesia, having their own unique social-cultural characteristics, contribute to shape the principals' and teachers' assessment literacy.

While social constructivism theory was initially associated with student/children's learning, it is also a relevant lens to investigate adult knowledge development (e.g., teacher and school principal assessment literacy). The current research applies social constructivism theory on the latter educational setting. Vygotsky's theory highlights two forms of interaction: horizontal peer-to-peer interactions (i.e., among colleagues) which foster elaboration and metacognitive awareness, and expert-novice vertical interactions, which facilitates the deep integration of knowledge details (Pritchard & Woollard, 2013). In the context of adult learning, peer to peer collaboration among colleagues has been shown to play a significant role in developing educators' competencies.

Besides serving as a conceptual lens in viewing the core construct of this study, the principals' and teachers' assessment literacy, social constructivism theory was employed to interpret findings as responses to the research questions, particularly the findings that highlighted the interaction and collaboration among the leaders and teachers in developing their assessment literacy.

7.2 Theory of Practice

The review of the literature demonstrates how the contemporary concept of assessment literacy reflects the negotiated and multi-faceted context-dependent professional and social practice. Adopting this contemporary concept has guided this study to employ Pierre Bourdieu's Theory of Practice (Bourdieu, 1977) as one of the theoretical frameworks, used particularly in making sense of the principals' and teachers' enactment of the assessment reform. In relation to the design of this study, which is explanatory sequential mixed methods, this theory is mostly related to the second (qualitative) phase of the study and addressed the third and fourth research questions.

Bourdieu introduced a theory about the dynamics of power, habitus, and social practices, which is called "theory of practice". The main idea is how subjective human experience, and objective social structures interact and influence each other within cultural and social contexts. This theory is relevant to this study, particularly for the second phase, investigating the enactment of assessment reform of the *Emancipated Learning* policy in junior secondary education. Specifically, it examines how the principals and teachers implement assessment reform, where they negotiate the demands of the national policy with the challenges emerged from their local school contexts.

Theory of practice key concepts include habitus, capital, field, and practice (Bourdieu, 1977). These concepts provided a robust framework for interpreting how the assessment policy reform is understood, exercised, and legitimised by school principals and teachers, especially considering the highly diverse social, cultural, and often resource-constraint settings across regions in Indonesia. In particular, this theory helped understand the underlying causes of differences in the enactment of assessment reform by principals and teachers across different geographical areas in Indonesia.

Habitus refers to a set of dispositions, deeply ingrained habits, perceptions and ways of thinking, that individuals develop overtime and shaped by their experiences, social and family background, education, and cultural environment. It shapes how individuals view the world, interpret events, and make decisions. This study argues that the principal and teacher assessment-related habitus are formed through their personal experiences including assessment experience as a student, prior teaching and assessment roles, professional education, their social-cultural influences, as well as the assessment-related support from colleagues and leaders, regional characteristics, and their cultural values. In addition, a principal long-experience in a bureaucratic educational system in Indonesia may habitually prioritise management and administrative compliance over innovative leadership practices (Lee & Hallinger, 2012). This study explores how the principal's and teacher's habitus influence their assessment-related decision-making and actions.

Habitus is closely associated with the second concept 'capital' (Harker, Mahar, & Wilkes, 1990).

Bourdieu introduced the concept of capital beyond economic sense including cultural capital (such as education, linguistic skills, and tastes), social capital (network and social connections), and symbolic capital (prestige, recognition or credentials). These multiple types of capitals interact with each other and influence one's power, acquisition of economic capital, and social status in a given field. People accumulate and use these forms of capital that align with or challenge social structures within different fields, such as education.

In Indonesia, there are issues of inequalities of socio-economic conditions among different parts of the country, which subsequently lead to the disparity in education quality among regions (Azzizah, 2015). This study examined how the principals and teachers from different parts of Indonesia enact the assessment reform differently, which might be linked to the gap in education quality. The schools from urban areas, especially those from western parts of Indonesia (most specifically in Java Island) typically benefit from greater cultural capital such as exposure to educational resources and trainings; and social capital such as professional networks and connections with educational policy makers, researchers, and experts. In contrast, those from rural and eastern parts of Indonesia often lack such advantages. This research explores how the differences in these forms of capitals might lead to different assessment approaches and practices in enacting the assessment reform.

A field is a social space, including in education, politics, science and art, where people compete for resources and status. Each field has its own set of rules, values, and norms, creating a structured environment in which people act. Within these fields, people use their habitus and capital to navigate, compete and try to success. In this current study, Bourdieu's concept of field is reflected in two social arenas. First the social-cultural contexts where the principals and teachers work, having their unique characteristics, values, and norms; and secondly the educational assessment realm in which there are differences or even conflicting assessment

paradigms, embraced by educators. These two types of fields are believed to simultaneously affect how assessment reform in Indonesia was enacted.

The last concept used in this study is practice, which is the result of the interplay between habitus, capital, and field. Through practice, the interaction between social structures and individual agency can be understood. Harker et al. (1990) argued that only through examining a particular practice, a field (e.g., education, specifically assessment) can be delineated. In this study, the concept of practice is interpreted as the principals' and teachers' practices of assessment in enacting the assessment reform. It is the core attention of the investigation, as it demonstrates the assessment literacy of the principals and teachers. This current study examined how the disparities in the leaders' and teachers' habitus and possession of capitals as well as the diverse ethnic and socioeconomic landscape (i.e., field) in Indonesia interact to collectively produce different forms of assessment reform enactment. This study argues that these variations might represent an inequality in educational practices from the point of view of theory of practice.

Bourdieu's idea about the dialectic of structure-agency proposed that social structure (the patterned system of relationships, norms, and institutions that shapes society) and individual agency (one's capacity to act and make choice) are interconnected and affect each other. The interplay between the two aspects determines human behaviour. Each person's ability to exercise agency in a field is influenced by their habitus and the amount and type of capital they have. On the other hand, social structure affects the extent to which individuals can exercise their agency. This notion means that people are not merely passive product of their environment, but they actively contribute to the reproduction or transformation of structures through everyday practices (Harker et al., 1990).

From the perspective of theory of practice, the variations of assessment practices by the principals and teachers in translating the new assessment policy in their schools reflect the interaction of their habitus, capital, and field, and might be interpreted as a form of inequality. However, multiplicity of enactment theory offers an alternative insight in making sense the disparities of the assessment

reform enactment the principals and teachers from diverse contexts in Indonesia performed, suggesting that the varied policy implementation may represent contextually appropriate adaption rather than an inequality. The following section discusses multiplicity of enactment theory in detail.

7.3 Multiplicity of Enactment

To get a comprehensive picture of principals' and teachers' assessment literacy as a socio-contextual dependent professional and social practice, which is negotiated and multi-faceted, also led to the use of theory of Multiplicity of Enactment proposed by Buxton, Allexsaht-Snider, Kayumova, Aghasaleh, Choi, and Cohen (2015) as another theoretical component of the framework of this study.

Buxton et al. 's (2015) theory of multiplicity of enactment was drawn from a longitudinal project conducted over three years that explored the relationship between science teachers' level of participation in a series of professional learning projects and the teachers' level of enactment of the pedagogical models learned from the project, within their classrooms. Buxton and colleagues maintained that while teachers' level of engagement in the project is important to their classroom implementation, the teachers' enactments of the pedagogical models were shaped by the interaction of their agentic moves and the structural forces of their school and classroom contexts.

This theoretical framework, which particularly explains the findings of the second stage of the current study, is underpinned by practice theory's structure-agency dialectic (Bourdieu, 1977). This current study argues that the theory's focus on the structure-agency dialectic shaping teachers' enactment of pedagogical practices upon participating in a series of professional learning activities, is closely related to current educational reform in Indonesia, emphasising schools' and teachers' autonomy and flexibility in providing education services for students. Autonomy is a key feature of agency defined as an individual ability to act independently, make decisions, and exert control over one's thought, behaviour, and environment (Bandura, 1986).

Buxton et al. (2015) claimed that the standard paradigm of programme adherence and fidelity of implementation in translating the pedagogical models from the professional development programme into the teacher professional practices was inadequate to explain the complex interaction and decision-making of teachers. This standard paradigm of linear relationship between professional development programmes and teaching practices assumes teachers as passive objects instead of active agentic actors engaged within social processes. They observed the teacher participants made a wide range of choices in the implementation of project practices into their classrooms. The researchers described that despite the same education policy and similar school contexts, the teachers exhibited different levels of tolerance for pushing against the structural obstacles. Those tolerances were manifested in the multiplicity of enactment of the project pedagogical practices.

Buxton and colleagues conceptualized the teachers' variety of choices, decisions and adaptation of the pedagogical models within the structural features and power of their teaching context to best suit their students' need as an agentic choices and actions. The researchers do not view the teachers' selection and adaptation of the pedagogical practices as a sign of weak performance or lack of fidelity of implementation. Instead, they saw it as teachers taking ownership of the practices that were more flexible, sustainable, and responsive to their classrooms. They asserted that teachers' multiplicity of enactment of the pedagogical models was the result of the interplay between the teachers' agency and the structural forces within their own contexts.

In the current study, the theory of multiplicity of enactment is used as the lens to explain the enactment of assessment reform in diverse school contexts in Indonesia. The researcher holds the idea that principals and teachers are agentic actors who apply their agency to negotiate structural features, including political and social-cultural demand and pressure both at micro and macro level (Xu & Brown, 2016), which resulted in the multiplicity of enactment of the new assessment policy. In addition, Coombs et al. (2018) suggested that despite uniform policies and similar professional trainings, students may experience assessment differently depending on their teachers. From the perspective of this theory, even when the leaders and

teachers operate under the same education policy, they might enact mandated policy differently due to their differences in agentic choices in relation to their respective unique contexts.

Whilst Buxton et al.'s theory of multiplicity of enactment which emerges from structure-agency dialectic was based on their research on teacher enactment of science pedagogical models learned from their engagement in a series of professional learning programme, this current study applies this theory as the lens to explain the differential enactment of school principals and social studies teachers of the assessment reform policy across diverse school contexts in Indonesia, exhibiting highly varied conditions.

The multiplicity of enactment theory provides an alternative framework for interpreting varied implementation of assessment reform across regions as opposed to the conventional paradigm of educational policy compliance and fidelity of policy enactment. This theory recognises the different forms of enactment as the principals' and teachers' agentic decisions based on their awareness of their local contexts and students' best interest. Importantly, these context-responsive enactments reflect principals and teachers' assessment literacy, particularly its contemporary conceptualisations viewing it as professional and social practices alongside assessment technical competences.

8. Summary

Hasim, Di, and Barnard (2018), in their study of Malaysian context, underscored the importance of paying attention to the existing knowledge and practices of stakeholders directly involved in educational policy to ensure a successful implementation of the policy. Given that school principals and teachers are the two key players in enacting the *Emancipated Learning* policy in Indonesia, it is important to investigate their assessment literacy since student assessment is one of the central focuses of the curriculum change.

Review of past literature on assessment literacy presented in this chapter, combined with Indonesian government regulations concerning student assessment, have provided a basis for the current study to define specific dimensions of teacher and principal assessment competencies. These dimensions then were used as the framework to initially evaluate the current level of assessment knowledge and skills of Indonesian school principals and teachers respectively through assessment literacy instruments developed in this study. In particular, assessment scholars advocated linking the dimensions of teacher assessment knowledge and skill to a specific school subject rather than being treated as a set of generic skills and knowledge of student assessment. The assessment literacy instrument for teachers in this study was developed specifically for social studies teachers.

Existing literature also identified factors that influence principal and teacher assessment literacy. Among the dominant factors that are believed to affect assessment knowledge and skills are assessment-related experience, courses in student assessment taken in tertiary education level, participation in in-service professional development programmes in student assessment, and local support from colleagues. However, findings on the impact of these factors on assessment literacy were varied across different education settings. Thus, this study sought to examine the combined impact of these four factors on the assessment literacy of Indonesian principals and teachers. Insights gained from this investigation may serve as valuable information to national and local education authorities in Indonesia for developing effective in-service professional development programmes, integrating more relevant assessment-related curriculum in pre-service teacher education programmes and postgraduate programmes, and establishing support structure such as professional learning communities.

Recent literature (from 2013 onward) on assessment introduced a new perspective of assessment literacy, viewing it as a negotiated, multi-faceted, context-dependent professional and social practice of educators rather than a fixed set of technical skill. The proponent of this recent concept argues that assessment literacy should be viewed from socio-cultural perspective and align with social constructivism theory, which believes that knowledge and understanding develop (more effectively) within

social and cultural contexts. This study adopted this contemporary framework and used it to further investigate the principals' and teachers' assessment literacy through the enactment of the assessment reform within their schools.

A review of past literature on the implementation of assessment-related policy, focusing on assessment for learning concept, highlights a variety of approaches and strategies aimed at applying the policy successfully. These strategies included employing a framework to guide teacher assessment practices, providing additional times and alternative formats of assessment, and tailoring grading practices to address diverse need of student population. The literature also pointed to some challenges that hindered the reform enactment, such as cultural barriers within schools, insufficient support from local authorities, and issues of teacher readiness.

Understanding the principals' and teachers' decision makings and actions regarding assessment in enacting the national mandate within their respective context is essential for evaluating the educators' assessment literacy. To analyse the dynamics, this study employed two theoretical frameworks, namely theory of practice and multiplicity of enactment. These theoretical frameworks account for the interplay between the educator's agency and the power and forces of the structures within their social and institutional contexts, providing a nuanced interpretation of how the assessment national policy is enacted differently across schools.

A careful review on international education journal databases such ERIC, Web of Science, and Scopus revealed no studies about assessment literacy of Indonesian principal and limited studies on Indonesian teachers' assessment literacy beyond the field of English as a Foreign Language (EFL). While a few studies have examined science teachers, no research specifically focuses for social studies teachers. The current study seeks to address these gaps by contributing new insights on the assessment literacy of Indonesian school principal and social studies teachers. This study will particularly add to the breadth of knowledge concerning the educators' assessment literacy across countries and school subjects. In addition, the instruments developed in this study for evaluating social studies teachers and

principal assessment literacy can also serve as useful tool for future relevant research.

Finally, it is hoped this current study will contribute to the existing knowledge base in this area by empirically examining how four factors (assessment experience, in-service professional development, assessment courses in higher education, and local support from colleagues) contribute to the principal and teacher assessment literacy. While previous research has produced mixed findings regarding the individual impact of four factors: assessment experience, in-service professional development, assessment courses in higher education, and local support from colleagues as the contributing factors on principal and teacher assessment literacy (e.g., the relationship between teaching experience and assessment literacy), this study explores their simultaneous effect. This study hypothesised that each of the four factors have a direct impact on assessment literacy, while in-service professional development and assessment experience also mediate the effect of assessment courses and local support on assessment literacy.

By addressing these gaps and building on existing research, this study aims to advance the understanding of teacher and principal assessment literacy within diverse educational settings and to inform policies and practices to support principals and teachers in implementing assessment reform. The next chapter presents the methodology and methods used in this study to achieve the study's aim.

Chapter three: Methodology and methods

1. Introduction

The previous chapter (Chapter Two) presented the conceptual and theoretical overview of principal and teacher assessment literacy, emphasising how contextual aspects influence the concept. The chapter also reviewed relevant empirical literature and identifies the key gaps within the literature, establishing the rationale to undertake this research.

Chapter three describes the researcher's positionality and research paradigm of the study in conjunction with ontology, epistemology, and choice of methodology. This chapter also presents the research design and research methods employed in this study to seek answer to the overarching research question: *What is the current state of Indonesian principals' and social studies teachers' assessment literacy in enacting the assessment reform in diverse school contexts?*

Besides presenting the research paradigm, design, and methods, this chapter also explain the rationale of the selection. Following this, details of data collection instruments, data collection procedure, as well as the population and sampling strategy, and data analysis are provided. Lastly, it discusses the ethical consideration of this research.

2. The researcher's positionality

In chapter one, my background that influence and drove the research and its topic was briefly presented. My previous education and working experience as a teacher trainer within the Indonesian Ministry of Education in the field of educational assessment and school leadership affected my research interest.

As part of the Ministry of Education, I was motivated to contribute to the success of Indonesia's education transformation policy through my research. On the other hand, as a teacher trainer who frequently interacts directly with principals and

teachers from many parts of Indonesia participating in the training sessions, I believe that it is essential to consider the school leaders' and teachers' experience and thoughts of the reform as they are the locomotive of change in their respective regions. By conducting this study, I intended to provide data-driven insights and recommendations, ensuring that the education reform, particularly in the area of assessment, are implemented effectively, taking into account the diverse characteristics of Indonesia's regions.

From my experiences throughout the training sessions attended by participants coming from various regions in Indonesia, I have found evidence based on the teachers' accounts that their instructional and assessment practices were heavily influenced by the conditions of the students, the schools, and the society where they serve. However, I also believe that teachers' pedagogical and assessment competencies, partially influenced by their previous education and trainings, play a significant role in shaping the quality of their practices.

Therefore, I argue that understanding the principals' and teachers' assessment literacy requires equal attention to the two key elements: the assessment knowledge and skills and the contexts where the assessment is practiced. This view has led to the adoption of the contemporary conception of assessment literacy and the theoretical frameworks of the study, discussed earlier in the previous chapter, and justified the pragmatic research paradigm in conducting the current research as detailed in the following section.

3. Research paradigm of the study

The choice of research paradigm refers to the researcher's way of viewing the social world and making sense of its intricacy (Patton, 2015), in relation to the research investigation. It is also known as worldview or research philosophy (Creswell, 2014), encompassing a researcher's ontological and epistemological stances (Crotty, 1998).

Ontology concerns the fundamental assumptions about the nature of reality, distinguishing between views that reality exists independently from human mind with singular truth and those who regard reality as socially constructed (shaped by social contexts), and thereby allowing multiple meanings (Bryman, 2016). These ontological positions in turn lead to distinct epistemological stances regarding how knowledge about reality is acquired, whether knowledge is discovered through objective and reliable empirical observation or constructed through human's subjective interpretation (Denzin & Lincoln, 2018). In addition to the ontological and epistemological considerations, research paradigm is further shaped by axiological commitments concerning the role of researchers' values in research. The axiological positions range from those who maintain neutrality and value-free research to those who acknowledge the inevitable influence of researchers' values throughout an investigation (Lincoln et al., 2018).

Based on the ontological and epistemological assumptions, there exist two opposing perspectives within the spectrum of research paradigm: positivism and interpretivism/constructivism (Bryman, 2016; Creswell, 2014; Crotty, 1998; Lincoln et al., 2018). Positivists believe that reality exists on its own, observable through senses and that it is independent from the human mind (ontological perspective) and thus requires objective, neutral, and reliable protocols of inquiry to discover (epistemological and axiological perspective) (Bryman, 2016). In contrast, interpretivists/constructivists assume that social reality is relative and thus permits multiple meanings and that it is constructed through humans' interpretation (ontological perspective), and therefore, it is actively and subjectively built, instead of discovered, through social interaction, experience, culture, and historical contexts (epistemological and axiological perspective) (Crotty, 1998).

The adopted research paradigm fundamentally informs the selection of research methodology, which refers to as systematic approaches employed in conducting research. Consistent with their underlying philosophical assumptions, positivist researchers typically adopt quantitative methodology, such as experimental design and survey in generating knowledge, whereas interpretivist scholars predominantly

apply qualitative methodology, such as in-depth interviews and case study analysis (Creswell, 2014).

In between these two contrasting paradigms exists pragmatism, which believes that reality exists as both objectively (value-free) independent of human mind and subjectively (value-laden) constructed through human interaction , and thus knowledge of reality is constantly negotiated, interpreted, or debated (Cherryholmes, 1992; Creswell & Plano Clark, 2018).

The focus of pragmatic paradigm is on addressing research problems, thus allowing various approaches to provide solutions to the problem. Using this paradigm, a researcher has the flexibility to choose and combine methods, procedures, or techniques, allowing them to get the best understanding of the research problem (Teddlie & Tashakkori, 2010). According to Morgan (2007), there is a clear connection between the view of pragmatic paradigm regarding the nature of knowledge and the technical choices of methods to produce the knowledge. Consequently, for pragmatists, knowledge should be examined using whatever tools possible (Patton, 2015), both quantitative and qualitative tools.

As a PhD researcher with a background of both as an education practitioner and a government official, the underlying motivation in conducting this study is to generate actionable knowledge that can enhance the effectiveness of education policies, particularly the assessment reform policy. By producing such knowledge, I intend to offer real-world solutions in implementing the assessment reform policy, by engaging school leaders and teachers, the front liner of the reform, as the active contributors of the knowledge.

For that reason, I did not limit myself to a single approach (i.e., quantitative or qualitative) but instead used multiple relevant methods to achieve the study's objectives. This aligns with the pragmatic paradigm which advocates pluralistic methods and emphasises problem-solving and the practical utility of research. Bielsa (2010) and Morgan (2007) highlighted that pragmatic paradigm is particularly suitable for education practitioners-researchers and policy-stakeholders as it

facilitates the production of knowledge that is both theoretically grounded and practically actionable to educational policy context.

Additionally, my recognition of assessment literacy as a multifaceted context-dependent professional competences and social practices demands an understanding of both measurable competencies (i.e., the assessment technical knowledge and skills) and contextual factors as school leaders and teachers need to negotiate between the demand of national assessment reform and the unique characteristics of their respective education context. Therefore, the principals' and teachers' assessment literacy cannot be fully understood through positivist and interpretivist lenses alone, thus require both quantitative and qualitative insights.

Driven by my positionality, research interest and motivation, as well as the nature of the main construct of interest in this study (i.e., assessment literacy), this study draws on pragmatic research paradigm. Ontologically, this study operates on the premise that assessment literacy encompasses both as measurable skills and as a socially mediated practice. Epistemologically, assessment literacy is derived from both objective and reliable statistical analyses alongside the principals' and teachers' accounts. Methodologically, the pragmatic paradigm of this study justifies the adoption of a mixed-method approach: a quantitative analysis to identify the principals' and teachers' assessment knowledge and skills and quantitative exploration to understand the enactment of assessment reform within their local context, together leading to a comprehensive understanding of the principals' and teachers' assessment literacy. The following section presents the mixed-method approach used in this study.

4. Research design of the study

The pragmatic paradigm of this study underpins the use of a mixed-method approach, combining or integrating quantitative and qualitative methods (Creswell & Plano Clark, 2018), allowing the researcher to work back and forth between the two opposing research approaches (Morgan, 2007). The combination of both methods is powerful as it maximizes the strengths while, on the other hand,

neutralizing and overcoming the limitation and weaknesses of each (Johnson & Onwuegbuzie, 2004; Teddlie & Tashakkori, 2010).

Mixed-method research acknowledges and operates on the principle that the social world is neither exclusively quantitative or qualitative, nor does it privilege one element over the other, but rather constitutes a complex mixed entity, regardless of the researcher's predominant disposition toward numerical or non-numerical data (Cohen et al., 2018). The adoption of a mixed-method approach is highly suitable in understanding the multifaceted nature of assessment literacy in this study, requiring both quantitative measurement of the assessment technical knowledge and skills and qualitative elaboration to understand the contextual factors affecting the assessment reform enactment by the school principals and teachers in their schools.

In social and education research, there are three primary mixed-method designs namely convergent parallel design, explanatory sequential design, and exploratory sequential design (Creswell & Plano Clark, 2018). This research employed an explanatory sequential mixed-method design, beginning with quantitative data collection and analyses (phase one), which results are then followed up in qualitative data collection and analysis (phase two). By using this design, I intend to use the qualitative results to explain in more detail the findings of the quantitative phase regarding principal and teacher assessment literacy. The results of the quantitative phase was also used to inform the selection of participants and the questions to ask in the subsequent qualitative phase (Creswell & Plano Clark, 2018).

It is important to note that before phase one was conducted, a preparatory stage was performed. The objective of this stage was to develop new assessment literacy instruments for principal and teacher. The study preparation is discussed in detail in the next section.

In phase one, this study used assessment literacy questionnaires to answer research questions one and two:

RQ1: *what is the level of Indonesian principals' and social studies teachers' assessment knowledge and skills?*

RQ2: *how do assessment-related experiences, assessment courses in higher education, in-service assessment professional development programmes, and local support contribute to Indonesian principals' and social studies teachers' assessment knowledge and skills?*

At the end of the questionnaires, the survey respondents were invited to participate voluntarily in an interview session. Those who were willing to participate provided their email address so the researcher could reach them out.

In phase two, this study collected and interpreted qualitative data regarding principal and teacher enactment of assessment new policy through semi structured interviews and paired interview methods. This phase is dedicated to answer research questions three and four:

RQ3: *how do school principals and social studies teachers navigate the enactment of mandated assessment reform against the challenges within their unique local context?*

RQ4: *how do the Indonesian principals' assessment literacy and leadership in assessment affect social studies teacher assessment literacy?*

Through the interviews, the researcher drew on the survey results to explore two main issues. First, the study examined the extent to which the principals' and teachers' assessment knowledge and skills informed the assessment-related decision making and actions, specifically how they balanced the unique contextual features of their schools with the demand of implementing the national policy of assessment reform. Second, it investigated whether engagement in the enactment processes contributed to the development of their assessment literacy. In addition, the paired interview analysis provided a deeper insight into the relationship between principal and teacher assessment literacy, detailing the specific forms of support principals provided for the teachers conducting student assessments.

Creswell and Plano Clark (2018) argue that besides the timing of data collection (i.e. concurrent or sequential), mixed-method research should also identify the emphasis of the type of databases used (i.e. quantitative or qualitative) and the strategy of integrating both databases (i.e., merging, connecting, or embedding). This study puts equal importance on both types of data from the first and second phases since both results are built upon the other to provide a clearer understanding of the phenomena under study. To integrate both sets of data, the results of phase one were followed up (i.e., connected) in phase two. By connecting both types of data, it is expected that more profound and more comprehensive analyses and interpretations can be achieved, leading to a better understanding of principal and teacher assessment literacy, the influencing factors, and how the education context and its complexity in Indonesia affect the assessment reform enactment process.

5. The preparatory stage of the research

The preparatory stage aimed at developing two distinct assessment literacy instruments, as no suitable measurement tools have been found to examine the two main constructs in this study: principal assessment literacy and social studies teacher assessment literacy. Zolfaghari et al. (2016) argued that limited studies have been undertaken concerning assessment literacy areas on different subject-specific field, including social studies, while review of literature in this study revealed no data collection instruments specifically tailored to principal assessment responsibilities as discussed in the previous chapter (Stiggins & Duke, 2008). This preparatory work was carried out through two steps: identification of assessment literacy key dimensions and pilot study. Each of the steps are presented as follows.

5.1 Identification of assessment literacy key dimensions

Two sources have been used to inform the development of the frameworks for examining assessment knowledge and skills of principals and teachers: two government official documents (i.e., ministry regulations) and literature focusing on

the components of assessment competency for teacher and principal. Consulting the Indonesian government regulations concerning assessment literacy components in addition to the analysis results of the literature are essential as the components may have different importance or meaning in different cultural settings (Pastore and Andrade, 2019). These frameworks then were used to develop principal and teacher assessment literacy inventory as part of the assessment literacy questionnaires.

The two government documents were: Ministry of Education Regulation of Assessment Standards for Pre-School, Primary, and Secondary Education (MoE, 2022a) and Guideline on Learning and Assessment (Centre of Assessment and Learning, 2022). Whilst the literature focusing on teacher assessment literacy dimensions were: Pastore and Andrade (2019); DeLuca et al. (2016a); Xu and Brown (2016); Brookhart (2011); and Mertler and Campbell (2005).

The analysis of the official documents and research literature identified several key assessment knowledge and skills dimensions for teachers present in both types of sources, namely assessment purposes (summative, formative, and diagnostic), assessment processes (constructing, administering, scoring, analysing, interpreting, and communicating assessment results), basic measurement theory (validity and reliability), and assessment ethics and fairness. Simultaneously the analyses of the two sources resulted in assessment literacy framework for teachers, consisting of six key dimensions and the related aspects drawn from the literature.

The following table presents the guiding framework for the teacher assessment literacy inventory (TALI) and the number of proposed items for each assessment dimension.

Table 1

Guiding Framework for Teacher Assessment Literacy Inventory

No.	Dimension	Aspects	Number of proposed items
1.	Purpose of Assessment (T-PA)	• Diagnostic • Formative • Summative	5
2.	Constructing and Administering Assessment (T-CAA)	• Developing and selecting appropriate assessment methods and instrument • Matching assessment with learning objectives	5
3.	Analysing and Interpreting Assessment (T-AIA)	• Test and item analysis. • Interpreting assessment results • Identifying problems of assessment instruments	5
4.	Communicating Assessment Results (T-CAR)	• Feedback strategies • Communication with other teachers • Communicating assessment results to parents	5
5.	Validity and Reliability of Assessment (T-VRA)	• Assessment validity • Assessment reliability	5
6.	Ethics and Fairness in Assessment (T-EFA)	• Implementing fairness principle in classroom assessment • Implementing ethics principle in classroom assessment	5
	Total		30

Note. The dimensions are derived from government documents and relevant literature. The aspects are derived from relevant literature. The alphanumeric codes in parentheses (e.g., T-PA) are the labels for each dimension.

As for principals, ten principal assessment competencies framework proposed by Stiggins and Duke (2008) was consulted in the development of the guiding framework for principal assessment literacy inventory (PALI). This study linked and clustered the ten competencies into five key dimensions of principal assessment knowledge and skills in relation to the student assessment regulations in Indonesia. The table below presents the five key areas drawn from the official documents and

their association with the ten competencies proposed by Stiggins and Duke (2008) and the proposed number of items for the PALI.

Table 2

Guiding Framework for Principal Assessment Literacy Inventory

No.	Dimensions (from Gov. documents)	Aspects (from Stiggins and Duke, 2008, p.287)	Number of proposed items
1.	Assessment purposes (P-PA)	1. "Understands the principles of assessments for (that is, used in support of) learning and works with staff to integrate them into classroom instruction". 2. "Understands the necessity of clear academic achievement targets and their relationship to the development of accurate assessments". 5. "Accurately analyses student assessment information, uses the information to improve curriculum and instruction, and assists teachers in doing the same".	5
2.	Support teachers to improve their assessment literacy (P-SUP)	3. "Knows and can evaluate the teacher's classroom assessment competencies and helps teachers learn to assess accurately and use the results productively". 4. "Can plan, present, or secure professional development activities that contribute to the use of sound assessment practices". 5. "Accurately analyses student assessment information, uses the information to improve curriculum and instruction, and assists teachers in doing the same".	6
3.	Communicate assessment practices and results to school community and local authority (P-CAR)	7. Creates the conditions necessary for the appropriate use and reporting of student achievement information and can communicate effectively with all members of the school community about student assessment results and their relationship to improving curriculum and instruction. 8. Understands the standards of quality for student assessments and how to verify their use in their school/district assessments.	5

4.	Ethics and fairness in school assessment policy and practices (P-EFA)	10. Understands the issues related to the unethical and inappropriate use of student assessment and protects students and staff from such misuse."	6
5.	Assessment policy that aligns with the national assessment reform effort (P-POL)	6. Can develop and implement sound assessment and assessment-related policies. 8. Understands the standards of quality for student assessments and how to verify their use in their school/district assessments. 9. Understands the attributes of a sound and balanced assessment system.	5
	Total		27

Note. The dimensions are derived from government documents. The aspects are derived from Stiggins and Duke (2008, p.287). The alphanumeric codes in parentheses (e.g., P-PA) are the labels for each dimension.

The guiding frameworks and the proposed number of items then were used to develop the PALI and TALI items respectively. To establish the content validity of the instruments, the researcher conducted expert and peer review. The review panel consisted of two fellow Indonesian teacher trainers, one university lecturer specialising in assessment, and one experienced school superintendent. In addition, two Indonesian UK-based PhD students who were undertaking research concerning educational leadership and policy were also involved. The reviewers evaluated all the items in both instruments for its alignment with the constructs being measured, its relevance to the current educational and assessment policy in Indonesia, and the clarity and appropriateness of wording.

Following revisions based on the feedback and suggestions, both instruments were tested in a pilot study to further validate its effectiveness. The following section describes the pilot study for the research data collection instruments and its results.

5.2 Pilot study

The pilot study of Principal Assessment Literacy Inventory (PALI) and Teacher Assessment Literacy Inventory (TALI) involved 161 respondents in total, comprising of 104 teachers and 57 school principals. The objectives of the pilot study were three-fold. First, to examine statistically whether the PALI and TALI in general are the accurate measures for principal and teacher assessment knowledge and skills. Second, to identify any inappropriate item(s) in measuring PALI and TALI and then either revised or dropped from the scale. Third, to get feedback from the pilot study respondents regarding their perception of the scale and their written comments of any changes needed to improve the scale.

The psychometric properties of PALI and TALI were examined using Rasch model for item and test level analyses. Rasch model is a statistical method that is used to test whether a measurement instrument, including a test, is unidimensional, that the items in the test are functioning well in measuring an intended construct (Bond & Fox, 2015). In this study, the intended constructs are the principal assessment literacy and teacher assessment literacy. The criteria in Rasch analysis to measure an instrument's unidimensionality are the infit and outfit in mean square (MNSQ) values, where the expected values range between 0.5 – 1.5 and values close to 1 is better. These criteria should be combined with their standardized indices (ZSTD) with expected range between -2.0 - 2.0 where values close to zero are better (Linacre, 2012). These four indices should be used simultaneously.

The Rasch analysis indicated that most items in PALI and TALI scales fell within the expected ranges, supporting their validity for measuring assessment literacy. However, two PALI items (P-SUP6 and P-SUP5) misfit for the scale as they exceed the threshold with ZSTD index above 2.0 and the MNSQ is > 1.5. Similarly, two TALI items (T-VRA4 and T-EFA5) also misfit, showing high ZSTD values (>2.0). Additionally, the strong reliability coefficients of PALI (0.89) and TALI (0.93) further support the instruments' effectiveness.

The researcher also received some important written feedback from the pilot study's respondents. For example, some items were redundant (ex. T-PA5 with T-PA1 and P-EFA1 with P-EFA6), some items were ambiguous or too complex (ex. T-CAA2), and the number of items in the scale was a bit too many. Additionally, some items were considered not relevant with its dimensions (ex. P-PA5 and P-POL4)

To address the results of the pilot study in improving the two scales, the researcher removed redundant and/or ambiguous items and misfit items. Five items were removed from each scale, leaving twenty-five items in TALI and twenty-two items in PALI. Further, items with long and complex sentences were simplified without changing the meaning. Clear and unambiguous questions and options encourage respondents to answer correctly and help reduce measurement error (Creswell, 2012). The tables below present the changes made in PALI and TALI scales/instruments following the pilot study.

Table 3

Items Removed in PALI and TALI

No.	Instrument	Removed items	Total items removed
1.	PALI	P-PA5, P-SUP5, P-CAR3, P-EFA1, and P-POL4	5
2.	TALI	T-PA5, T-CAA2, T-AIA2, T-VRA4, and T-EFA5	5

Note. Items removed due to misfit, redundancy, and/or not relevant with its dimension

Table 4

Items with Simplified Sentences in PALI and TALI

No.	Instrument	Sentence simplified items	Total items
1.	PALI	P-PA3 and P-EFA5	2
2.	TALI	T-PA2 and T-VRA2	2

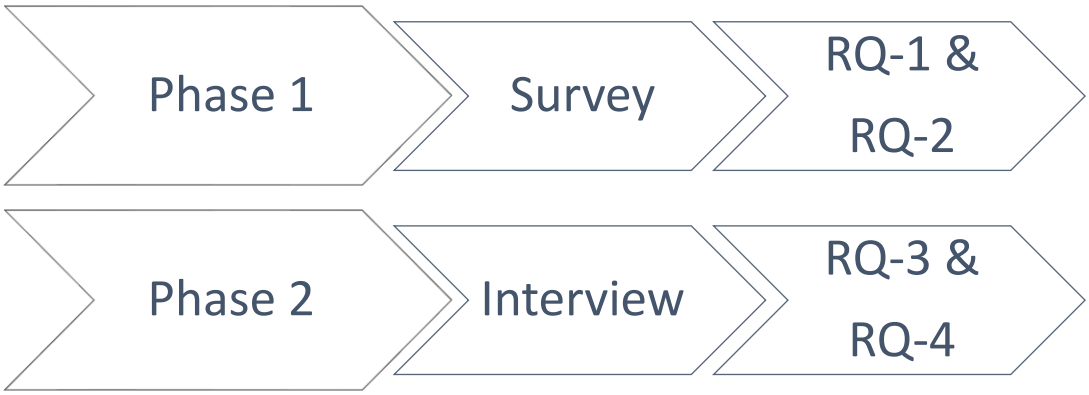
Through this pilot study, this research developed objective, neutral, and reliable quantitative data collection instruments designed to measure principal and teacher assessment knowledge and skills as the observable and measurable element of the

assessment literacy construct under investigation. These instruments were utilised in the first phase of the study. The research methods employed both in the first and second phase of the study are elaborated in the following section.

6. Research Methods

This section describes the research methods employed in phase one and phase two including the data collection instruments, and data analysis. The below diagram depicts the methods of the study and the related research questions.

Figure 2
Relationship of Methods and RQ



The method in each phase are discussed in two separate sections.

6.1 Phase one

The survey method was used in phase one to investigate school principals' and social studies teachers' assessment knowledge and skills levels (RQ-1) and the factors that were hypothesised to affect the assessment knowledge and skills (RQ-2). The findings from this phase reflect Bernstein's vertical discourse of assessment literacy viewed as a multifaceted context-dependent professional competences and social practices (Pastore & Andrade, 2019; Willis et al., 2013).

6.1.1 Survey

Survey research is an approach in which researchers gather data from a sample or an entire population to describe the population's attitude, behaviour, opinion, or characteristics (Creswell, 2014). In this study, a survey was used to gather data from sample of principals and teachers at junior secondary schools across several regions in Indonesia to draw inferences about the current level of assessment knowledge and skills of Indonesian school leaders and social studies teachers. The survey was administered online, taking advantages for its quick distribution and the familiarity and extensive uses of the internet by people today, including teachers and school leaders.

The online survey method was suitable for this study as it allows the researcher to gather information from large and geographically dispersed population in a time and cost effective manner (Fowler, 2014). Additionally, the online mode of the survey was intended to encourage candid responses as it provides the respondents with flexibility to complete the survey in their comfortable time and place, while ensuring anonymity. The respondents' anonymity and the privacy when responding to the online questionnaire may reduce bias in the data due to social desirability (Adams & Lawrence, 2019). By this approach, the survey data is expected to be more reliable and reflect the approximate 'true' state of the principals' and teachers' assessment knowledge and skills and its relationship with the predictor variables.

Three basic types of survey are cross-sectional, which collect data at one point in time, trend survey collecting similar data on the same population on more than one occasion, and longitudinal survey collecting data on the same individual to measure change over time (Sullivan & Calderwood, 2016). In a quantitative survey, a questionnaire is used to gather quantitative data, which is then analysed to answer research questions or test hypotheses (Bielick, 2016). This research began with a quantitative survey method using questionnaires with close-ended questions to get cross-sectional quantitative data. The data from the survey questionnaires were

analysed to answer research questions one and two regarding principal and teacher assessment knowledge and skills and the factors affecting assessment literacy.

6.1.2 Survey questionnaire

This study developed and employed two assessment literacy questionnaires, for teachers and principals respectively (see Appendix 1: Survey questionnaire for teachers and Appendix 2: Survey questionnaire for principal), as the survey instruments. Microsoft Form survey software provided and approved by the University of Nottingham was used to create the online version of the questionnaires.

Two main sections made up both questionnaires. Section one consists of items regarding background information and the four proposed assessment literacy predictor variables. For the background information, the participants were asked to identify the status of their school (pilot or non-pilot) and the length of assessment reform implementation (first or second year), their gender, age, municipality/regency; provinces; and degree attained. This information provides the readers with detailed information of the study's participants. More importantly, the background information was used for data analysis, such as the schools' status (pilot/non-pilot) and the regions (i.e., the municipality/regency) where the schools are situated.

This was followed by items concerning the four predictor variables: experience, courses, in-service training, and local support. All these four variables were measured using 10-point rating scale items. The aim was to provide respondents with a wide range of responses and flexibility to obtain more accurate data. First, assessment-related experiences were measured using three items: years of teaching experience; serving as an instructor in assessment-related professional development programmes at school, regional, and/or national level; and contributing as test or assessment item writer at the same three levels.

Secondly, three items regarding the variable of assessment-related courses in higher education were asked next, encompassing the number of course credit/unit

undertaken in higher education; the perceived relevance of the course content with their current assessment responsibilities; and the perceived effectiveness of the course delivery to support their assessment practices.

The subsequent three items measured the assessment-related in-service professional development programme (PDP) variable: the number of training hours of PDP they have participated in; the perceived relevance of the PDP content with their current assessment responsibilities; and the perceived-effectiveness of the PDP training toward their assessment practices. The questionnaire items regarding experience, courses, and in-service trainings for principal and teacher participants are identical.

The last part of section one consists of three items which make up the assessment-related local support variable. The local support for teacher variable consists of the items: support from teacher colleagues, support from school principal, and support from local teacher professional community. Whilst, for the principals, the three items are: support from teacher colleagues, support from the superintendent, and support from local principal community.

Section two of each questionnaire measured principal and teacher assessment knowledge and skills through Principal Assessment Literacy Inventory (PALI) with 22 items and Teacher Assessment Literacy Inventory (TALI) with 25 items respectively. Framing under assessment for learning concept, both instruments adopted scenario-based assessment inventory developed by Mertler and Campbell (2005) where each real-world classroom scenario is used for a group of items measuring various assessment dimensions.

A principal need to develop a balanced system between the classroom and institutional levels while also helping teachers build their assessment literacy. Since principal's assessment leadership requires building school-wide assessment competence (Absolum et al., 2009; Hipkins & Cameron, 2018), the assessment literacy of principals covers aspects beyond classroom assessment practices. Therefore, the twenty-two items in PALI cover the dimensions of purposes of assessment, supporting teachers to improve assessment literacy, communicating

assessment, assessment ethics and fairness, developing and implementing assessment policy. There are five assessment scenarios in PALI, each scenario is followed by four or five items with three-option responses. The dimensions of PALI and number of items are as follows.

Table 5

Dimensions of Principal Assessment Knowledge and Skills

No.	Dimensions	Number of items
1.	Purpose of assessment (P-PA)	4
2.	Assessment support for teachers (P-SUP)	5
3.	Communicating assessment results (P-CAR)	4
4.	Ethics and Fairness in Assessment (P-EFA)	5
5.	Assessment Policy (P-POL)	4
	Total	22

As for the teacher respondents, building on the view that teacher assessment literacy should be embedded with a subject content knowledge (Xu and Brown, 2016; Brookhart, 2011), this study situates the data collection instrument of teacher assessment knowledge and skills within social studies subject for junior secondary students. This was done by incorporating social studies topics into the scenarios and the items concerning the primary assessment dimensions. There are five assessment scenarios in TALI, and each scenario is followed by four to five questions, each with three response options. The dimensions and the number items of TALI are presented in the following table.

Table 6

Dimensions of Teacher Assessment Knowledge and Skills

No.	Dimensions	Number of items
1.	Purpose of Assessment (T-PA)	4
2.	Constructing and Administering Assessment (T-CAA)	4
3.	Analysing and Interpreting Assessment (T-AIA)	4

4.	Communicating Assessment Results (T-CAR)	4
5.	Validity and Reliability of Assessment (T-VRA)	4
6.	Ethics and Fairness in Assessment (T-EFA)	5
	Total	25

The following section specifies the survey's target population and the sample selection criteria.

6.1.3 Population and sample

The target population of this study were school principals and social studies teachers at junior secondary schools (years 7-9), whose schools were implementing the new national curriculum, the Emancipatory Curriculum. According to the Ministry of Education, in 2024, slightly more than 40,000 junior secondary schools, comprising pilot and non-pilot, public and private schools, across 38 provinces have implemented the Emancipatory Curriculum. In a quantitative research, sample should be representative of the population. The sampling process started with "defining the target population, and identifying accessible population, which the researcher has reasonable access" (Collins, 2016, p. 281).

Due to some constraints related to participants' accessibility, time, and resources, convenience and snowball sampling techniques were used as they offered feasible approaches to collect actionable data without compromising the study's objectives. The sample of school principals and social studies teachers were selected conveniently from five provinces as they might be available and willing to participate. After securing permission from local education authorities, I distributed the survey links to some principals and teachers in the five provinces who were the alumni of my institution's training programmes. The principals and teachers then were kindly asked to forward the questionnaires to their colleagues within their network.

According to Creswell (2012), when probability sampling is difficult to get, researchers can use non-probability sampling, in which participants are selected because they are available and represent characteristics the researchers seek to investigate. Even though a non-probability sampling technique was used in this study, the distribution of the sample across the five provinces in Indonesia provided confidence as an adequate representation of the population. Furthermore, the sample selection was guided by its relevance to the study's conceptual and theoretical framework (Patton, 2015). The context-dependent nature of assessment literacy and the multiplicity of enactment theory adopted in this study necessitated the researcher to deliberately include participants representing Indonesia's socio-cultural diversity, economic development levels, and geographical setting (i.e., rural and urban areas).

Accordingly, this research targeted samples from five Provinces located in five (from eight) main Islands namely Java, Sumatera, Kalimantan, Sulawesi, and Papua. The provinces are East Java, Riau (in Sumatera), East Kalimantan, South Sulawesi, and West Papua. These provinces were selected as, first, they represent the eastern, central, and western parts of Indonesia. Secondly, they represent the diverse characteristics of Indonesian regions in terms of social, cultural, and economic prosperity aspects, and thirdly, the researcher has either direct or indirect access to communicate with the principals and teachers and/or the local education authorities (regional level). The final sample consists of participants coming from more than these five provinces (seven provinces for teachers and eight provinces for principals) due to the snowball sampling technique.

In total 645 respondents completed the online surveys comprising of 257 principals and 388 social studies teachers. The number and geographical distribution of the principals and social studies teacher respondents is depicted in following two tables:

Table 7
Teacher Sample

No.	Provinces	Number of samples
1.	Papua & West Papua	14
2.	North and South Sulawesi	78
3.	East Borneo	56
4.	East Java	105
5.	Riau (in Sumatera)	134
	Total	388

Table 8
Principal Sample

No.	Provinces	Number of samples
1.	West Papua	7
2.	North and South Sulawesi	51
3.	East Borneo	68
4.	East Java	58
5.	Riau, South & West Sumatera	73
	TOTAL	257

Data from the sample of this study were analysed using the following quantitative data analysis techniques.

6.1.4 Survey data analysis

The data collected in phase one through the survey questionnaires were analysed using descriptive statistics and structural equation modelling (SEM). Before proceeding to the quantitative data analysis, issues such as data cleaning and missing data in the databases need to be addressed (Creswell, 2012). The survey database was inspected to clean the data from values outside the accepted range using a technique of frequency distribution. The online questionnaires were designed so that only respondents with complete data were included in the analysis.

This was done by making the response to each item required before moving to the next item, resulting in no missing data in the database.

After verifying the survey data are free from missing values and out-of-range entries, the data analysis commenced with descriptive statistics, including the central tendency (mean), variability (standard deviation), frequency and shape distribution of the data with graphical representation (Gravetter & Wallnau, 2017). Descriptive statistics is an essential element in the analysis, for it examines basic statistical assumptions for inferential statistics in testing hypotheses or answering research questions (Steedle, 2016). This analysis results provided background information regarding the sample of the study: the respondents' gender proportion, the average age of the respondents, the proportion of higher education degrees, and the proportion of province of origin. The descriptive statistics also show the numbers and percentages of pilot/non pilot schools and schools with first/second year implementation of the reform; both are important in interpreting the findings. Further, the descriptive statistics analysis described the frequency of data points in different categories and the normality distributions of the items making up the four predictors of assessment literacy: experience, courses, in-service training, and local support. The latter results are essential for the subsequent multivariate data analysis (structural equation modelling) (Hair et al., 2010).

To answer the first research question, "what is the level of Indonesian principal and social studies teacher assessment knowledge and skills?", the mean score and standard deviation from the PALI and TALI, respectively, were calculated. These data describe the average assessment knowledge and skills of the respondents and how spread out the scores around the average value (Gravetter & Wallnau, 2017). The normality of the assessment literacy score distributions is also examined, complemented with graphic representation of the distributions. The normality of the score distribution of the assessment literacy is important as many statistical analysis and educational assessment rely on the assumption that data are normally distributed, allowing fair comparison between groups and valid interpretation respondents' ability (Gravetter & Wallnau, 2017). To provide more detail information, the mean scores and standard deviation of each dimension of

assessment literacy were also calculated. In addition, item difficulty analysis using Classical Test Theory (CTT) (Kline, 2005) was performed to examine how each item and the complete instruments are functioning toward measuring assessment knowledge and skills.

Structural Equation Modelling (SEM) analysis with maximum likelihood estimation method using AMOS software (Byrne, 2010) were performed to answer the second research question, "How do assessment-related experience, assessment courses in higher education, in-service assessment professional development programmes, and local support contribute to Indonesian principals' and social studies teachers' assessment knowledge and skills?". Structural equation modelling (SEM) is a multivariate statistical analysis which examines complex causal relationship among unobserved (latent) constructs, while at the same time accounts for the measurement of the latent constructs by observed (measured) variables (e.g., items in a questionnaire/test) (Kline, 2016).

There are two primary approaches in SEM, the covariance-based (CB-SEM) and partial least square (PLS-SEM) (Hair et al., 2019). While PLS-SEM emphasises to maximise the explained variance of outcome variables (i.e., PAL/TAL), with the purpose of theory development, and useful for highly complex models with many constructs, CB-SEM focus on analysing covariance structures of the observed variables to test a hypothesised model, with the purpose of theory testing, and evaluates how well the model fits the observed covariance matrix (Hair et al., 2019). I argue that the CB-SEM is more appropriate with the purpose of this study and the characteristics of the hypothesised models, which are relatively parsimonious.

In a full SEM, the relationship model combines confirmatory factor analysis (CFA), which examines the extent to which some observed variables reflect the latent constructs, with the structural model (path analysis), testing the causal relationship among the latent constructs. The causal relationships among the constructs in a model were represented by structural paths, while the relationship between the constructs and the corresponding observed variables (indicators/items) which

reflect the constructs are termed factor loading, and both structural path and factor loading are estimated using regression weight/coefficient (Kline, 2016).

The ability to examine complex relationship among variables enables structural equation modelling (SEM) to address complex real-world problems in social and educational fields, which often involve multiple variables with intricate relationship. For that reason, this study employed SEM to analyse the relationship between principals' and teachers' assessment literacy with four predictor variables: experience, assessment-related coursework, assessment trainings, and local support, both direct and indirect relationships. The SEM model of principal assessment literacy (PAL) and teacher assessment literacy (TAL), depicting the relationship between the four predictors and assessment literacy, have been presented in Chapter Two.

Before analysing the causal relationships in PAL and TAL models, the validity and reliability of the scales need to be established. This was done first by examining all the measurement scale of the latent constructs (the PAL, the TAL, and the four predictors) within both SEM models using confirmatory factor analysis (CFA) (Kline, 2016). In addition, the composite reliability (CR) and construct validity (CV) of the measures were examined (Cheung et al., 2024).

Several model goodness of fit statistics was used to evaluate the measurement of the constructs through CFA. The fit statistics are χ^2/df (Kline, 2016), Comparative Fit Index (CFI), Tucker Lewis Index, Root Mean Square Error of Approximation (RMSEA) and Standardised Root Mean Square Residual (SRMR) (Hu and Bentler, 1999). The chi-square (χ^2) is the statistics used to evaluate the extent to which a hypothesised CFA model matches the data from the sample, reflecting the approximate reality in the population. A χ^2 probability (p) value > 0.05 indicates that there is no significant difference between the hypothesised model and the sample data, meaning that the model fits the actual data well. However, this p value is sensitive to sample size, where large sample size tends to produce small p value (<0.05). To address this sample size issue, degree of freedom (df) of a model is incorporated to adjust the

model fit interpretation. A χ^2/df ratio less than 3 suggests a model fits the data well even though the p value <0.05 (Kline, 2016).

The comparative fit index (CFI), and Tucker-Lewis Index (TLI) are fit criteria using a baseline model to which a proposed model is compared, where a value of one show that the proposed model fits the data perfectly/well. The RMSEA and SRMR both measure the deviance level of a model to a perfect model fit, thus values closer to zero indicate a good model fit (Goretzko et al., 2024). The following table presents the acceptable/minimum cut-off criteria for a good model fit (Hu & Bentler, 1999; Kline, 2016).

Table 9

Cut Off Criteria for Model Goodness of Fit

Fit Indices	Cut-off value
χ^2 /df	< 3.0
CFI	≥ 0.90
TLI	≥ 0.90
RMSEA	< 0.06
SRMR	< 0.08

Note. χ^2 /df = chi-square per degree of freedom. CFI = Comparative Fit Index. TLI = Tucker-Lewis Index. RMSEA = Root Mean Square Error of Approximation. SRMR = Standardised Root Mean-squres Residual

Lastly, as both models hypothesised direct and indirect relationship between the four predictors and assessment literacy, the indirect analysis of the model were also conducted. In SEM, indirect analysis examines whether a variable mediates the impact of an independent (predictor) variable to a dependent (outcome) variable. In this study, two variables: assessment-related in-service trainings and assessment-related experiences were hypothesised mediating the impact of assessment courses and local support to principals' and teachers' assessment literacy.

6.2 Phase two

In phase two, semi-structured interview and paired interview methods were employed to address research questions three and four. Findings from phase one were elaborated more deeply through the interviews and paired interview analysis. Each of the methods, the interview instrument, and qualitative data analysis are presented in the following sections.

6.2.1 Semi-structured interview

An Interview is conducted to capture people's perspective using open-ended questions based on the assumption that the perspective is valuable, meaningful and can be made explicit (Patton, 2015). The open-ended questions used in an interview allow participants to respond and reveal their opinion, values and experience without any constraints from the researcher's perspective or past research findings (Creswell, 2012). According to Patton (2015), an open-ended interview can be carried out using three approaches: informal conversational interview, general interview guide, and standardized open-ended interview.

First, in an unstructured informal conversational interview, questions are asked spontaneously during the natural flow of interaction between participants and interviewer, often as part of observation fieldwork. Second, a general interview guide approach uses a framework, in which issues and topics to explore are determined in advance. Using the framework, an interviewer can freely ask questions within an issue/topic, decide the sequence of issues to be addressed, and wording the questions spontaneously while keeping the conversation focused on the predetermined issues or topics.

Third, in a standardised structured open-ended interview, the exact wording and sequence of the questions, including probing questions and transition sentences, are carefully crafted before the interview. This method ensures each interview gets the same question in the same order, thus can reduce the variation in data collected, especially when several interviewers with varying interview skills are

involved. When studying multiple sites, this method provides comparability across sites.

This research used a combination of the second and third interview approaches. The decision to combine the second and third approaches allowed for the best use of limited interview time available and a more systematic and comprehensive interview by focusing on the issues to be explored determined in advance, while at the same time providing space for unique individual experiences and contextual differences to emerge. The combination of the second and third interview approaches led this study to employ a semi-structured interview (Bryman, 2016). It is semi-structured because it involves a set of predetermined open-ended questions under several issues regarding the assessment reform enactment, while allowing for emerging follow-up or probing questions to explore the issues in more depth based on the participants' responses. The use of probes allows the interviewees to either elaborate or clarify their answers (Creswell, 2012).

Given the adoption of the context-dependent and multifaceted conception of assessment literacy in this study (DeLuca et al., 2019; Willis et al., 2013), semi structured interviews provided valuable means of capturing unique experiences, thoughts, and opinions of Indonesian principals and teachers in enacting the assessment reform in their diverse local contexts. Multiple perspectives from the interviewees shed light on the various enactment of the assessment reform.

The school leaders and teachers who participated in phase one and expressed their willingness to voluntarily participated in the phase two of this study and met the pre-determined criteria (explained later in section about population and sample), were invited to a series of one-on-one interviews, either face to face or online through a video conference software, depending on the participants' preference. Among the interview participants, there were some pairs of principal-teacher coming from the same schools. These pairs facilitated further analysis to address the fourth research question.

6.2.2 Paired interview

The pragmatic paradigm of this study justifies the use of various approaches, methods, or techniques, allowing the study to get the best understanding of the research problem (Teddlie & Tashakkori, 2010), and that knowledge is best explored through any relevant methods (Patton, 2015). Accordingly, in the phase two, this study also employed a paired interview method, particularly to address research question four. This approach was necessary because research question four requires matched principal-teacher participants from the same schools.

To perform the approach, this study isolated and concurrently analysed the interview data of the principals and teachers within the same schools. This approach was used to uncover their individual assessment conceptions and practices, and to determine the extent of alignment between them. It is noteworthy that the researcher did not purposefully invite the pairs of principal-teachers from the same schools. The pairs emerged when the researcher identified that there were some principals and teachers voluntarily participating in phase two coming from the same schools. In this research, the paired interview approach enabled the exploration of the interaction between principal and teacher in the enactment of the assessment reform, which is important to understand their assessment literacy.

The interviews were guided by an interview protocol developed in this study, which is explained in the following section.

6.2.3 Interview protocol

An interview protocol was used to guide the interviewing process (Creswell, 2012) (see Appendix 3: Interview protocol for teachers and Appendix 4: Interview protocol for principals). The interview protocol was developed based on the research questions and informed by the theoretical frameworks of the study. For example, the social constructivism theory (Vygotsky, 1978) shaped the questions regarding the collaborative dynamics and support system among teachers and leaders, both within and beyond school settings. The multiplicity of enactment theory (Buxton et al., 2015) guided the formulation of questions concerning the challenges and

barriers encountered during the implementation of the assessment reform and the adaptive strategies to address them. The theory of practice (Bourdieu, 1977) particularly informed the questions probing the participants about how their past education, professional trainings, and experience impacted their engagement with the assessment reform.

The questions were designed to elicit the dynamics of the assessment reform enactment, which, according to Willis et al. (2013) and DeLuca et al. (2019) reflect the context-dependent and multifaceted nature of the principal and teacher assessment literacy. Three types of questions were used during the interview: experience-behaviour, opinion-values, and knowledge questions (Patton, 2015). Experience-behaviour questions that elicit behaviour, experience, actions, and activities were asked during interview since phase two aimed to unpack the assessment enactment by principals and teachers as the result of the interplay between the national assessment reform mandate and the unique school local contexts. Some opinion-values questions were also raised to reveal the interviewee's opinion about to what extent they are able to enact the assessment policy and whether the enactment process strengthens their assessment literacy. In addition, knowledge questions were also necessary to confirm the interviewee's understanding of the policy.

Several main issues relevant to the focus of the investigation were used as a guide in writing of the open-ended questions. The main issues include: the challenges of policy enactment; difficult aspects of the new assessment policy; the support available in implementing the reform; the strategies used during the enactment process; adjustment, adaptation, and modification of the assessment reform enactment; assessment of student behaviour; and the perceived successful enactments of the assessment reform. The issues used to guide during the interview process acted as the analytical framework (Patton, 2015), describing how principals and teachers navigate the national assessment policy and their local education circumstances. In addition, questions concerning the impact of professional experience, assessment coursework in higher education, in-service professional development programme in assessment, and local support were asked to elaborate

the influence of these predictors examined in the quantitative phase. Probing and follow up questions were also raised during the interview to clarify or deepen participants' answers.

Unsurprisingly, due to the nature of the open-ended questions in the interviews, new relevant issues emerged and identified during the interview process such as potential barriers to the implementation, strengths of the new policy, and shared experiences during the enactment process. These new emerging issues along with the pre-determined issues guiding the interviews both were later used to cluster all the codes identified from the transcripts and were considered to be the main themes of the qualitative findings.

The interview protocol was used to gather data from the following interview participants.

6.2.4 Interview participants

A purposive sampling scheme defined by specific criteria were used in the qualitative phase (Creswell, 2012). In this mixed-method research, 14 participants were selected to interview in the second phase from the first phase (Creswell, 2014). In selecting the interview participants, the researcher sought a balanced proportion of the interview participants in terms of their scores (high/moderate/low), region characteristics (rural/urban), and the duration of assessment reform implementation (first/second year). This approach aimed to ensure that the interview results reflect data from a wider range of participants' profile, enriching the information obtained (Patton, 2015).

The following two tables present the teacher and principal interview participants.

Table 10

Teacher Interview Participants

No.	Participant (gender)	Provinces	Urban/rural	Pilot/non	1st/2nd year	AL score[a]
1	Teacher 1, (female)	North Sulawesi	Rural	Non-pilot	1st year	80% (20/25)
2	Teacher 2, (female)	North Sulawesi	Urban	Pilot	2nd year	88% (22/25)
3	Teacher 3, (female)	East Java	Rural	Pilot	2nd year	84% (21/25)
4	Teacher 4, (female)	East Java	Urban	Pilot	2nd year	76% (19/25)
5	Teacher 5, (female)	West Papua	Rural	Pilot	2nd year	80% (20/25)
6	Teacher 6, (female)	Papua	Rural	Pilot	2nd year	88% (22/25)
7	Teacher 7, (female)	East Borneo	Urban	Pilot	1st year	68% (17/25)
8	Teacher 8, (male)	Riau	Urban	Non-pilot	1st year	32% (8/25)

Note. [a] AL score = Assessment Literacy score

Table 11

Principal Interview Participants

No.	Participant	Province	Urban/rural	Pilot/non	1st/2nd year	AL score[a]
1	Principal 1, (male)	North Sulawesi	Urban	Pilot	2nd year	50% (11/22)
2	Principal 2, (female)	East Java	Urban	Pilot	1st year	81% (18/22)
3	Principal 3, (male)	East Java	Rural	Non-pilot	1st year	64% (14/22)
4	Principal 4, (male)	West Papua	Rural	Pilot	2nd year	86% (19/22)
5	Principal 5, (female)	North Sulawesi	Rural	Pilot	2nd year	68% (15/22)
6	Principal 6, (male)	East Borneo	Urban	Pilot	2nd year	77% (17/22)

Note. [a] AL score = Assessment Literacy score

To categorise the teachers' and principals' scores, this study applied this rule: High = >80%, Moderate = 51% – 80%, Low = < 50%.

In the following chapters, the presentation (Chapter Four and Five) and discussion (Chapter Six) of findings referring to interview participants, use 4-digit alpha-numeric identification codes as follows:

- Role (P = principal, T = teacher)
- List order identification (1 to 6 for principal, 1 to 8 for teacher), followed by a hyphen.
- Length of policy implementation (1 = first year, 2 = second year)
- Assessment knowledge and skills category level (H = high, M = moderate, L = low)

For example, P4-2H refers to a principal (P), listed fourth (4), second year of implementation (2), and scored high (86) (H) on assessment knowledge and skills.

The qualitative data from the interview participants were analysed through the following procedures and techniques.

6.2.5 Interview data analysis

The interviews conducted face-to-face were audio-recorded and those that were conducted online through the University of Nottingham's Microsoft Teams were recorded using the recording feature in the software, following participants' consent. Each interview lasted between approximately 45-60 minutes. The researcher listened to each of the interview recordings several times and transcribed them verbatim manually. After finishing transcribing each interview, the researcher listened to the recording again to ensure the accuracy of the transcription. The steps of listening to the recordings and transcribing them manually helped the researcher get general initial understanding regarding the assessment reform enactment processes within each school. The researcher, in fact has started the qualitative data analysis early in this stage through the listening and transcribing iterative process (Creswell, 2014).

All the interviews (fourteen in total) were conducted and thus transcribed verbatim in the Indonesian language. To maintain credibility and transparency of the interview data analysis, a member checking (Saldana, 2016) approach was conducted. This was done by emailing all the complete transcripts to the relevant participants to get feedback and correction if necessary. The participants were welcomed to talk to the researcher if they thought something was missing or incorrect from the transcript. The researcher then would refer back to the recording and make adjustments and/or notes to the transcripts when necessary.

In addition, three transcripts (one from teacher and two from principals) were translated into English for review by the research supervisor. However, since the initial transcripts were messy due to the nature of verbatim transcription, prior to translation the researcher carefully edited the transcripts to enhance the readability without compromising the participants' intended meaning. The editing process included removing unnecessary repetitive words and, where necessary, inserting clarification words within [square bracket] to incomplete statements. This ensured accurate translation while preserving original meaning.

After all the transcribing process completed, the analysis of the interview data then proceeded to reading each of the transcripts carefully to get deeper sense of the data by linking them with the research questions and the theoretical frameworks (Bryman, 2016). While reading, initial codes were assigned, labelling segments of the transcripts, either using terms produced by interviewees or created by the researcher (Creswell, 2012). Important notes are also jotted down in the margin to follow up further. The researcher used NVivo software provided by the University to help with the coding process.

To increase the trustworthiness of the data, at the beginning of the coding process, the researcher and one of the supervisors conducted intercoder comparison on the three English-translated transcripts to determine agreed codes of the interview data. The agreed codes then were used as the references for the subsequent coding process of the rest of the transcripts. However, new codes emerged from the rest

of the transcripts when the available agreed codes could not represent the relevant or important situation described in the interviews.

Content analysis of the interview transcripts was employed to analyse the interview data. It was an attempt to make sense of qualitative data by identifying consistencies and meaning in the form of patterns or themes (Patton, 2015). This was done by reviewing and organising participants' responses from the interviews and labelling them to generate codes (Creswell, 2012). This analysis yielded detailed descriptions of participants' experiences, understanding, knowledge and perspectives, described under pre-determined issues and emerging issues of various assessment reform enactment.

The study conducted three cycles of coding. No particular coding methods were used as a guidance in the initial stage (coding the first three transcripts), which Saldana (2016) called emergent data-driven (inductive) coding choices strategy. However, before the first cycle coding proceeded, a literature (book) of coding method written by Saldana (2016) was consulted. Several coding schemes in the literature were found to be relevant with the coding process performed in this stage such as versus, descriptive, process, emotion, dramaturgical and structural coding methods. These coding techniques were subsequently used to guide the coding processes.

This study occasionally utilised versus coding method as the researcher was looking to compare and contrast the assessment policy and practices in the previous curriculum and the current curriculum. "Versus Coding is appropriate for policy studies, evaluation research, critical discourse analysis, and qualitative data sets that suggest strong conflicts or competing goals" as the codes identify dichotomous or binary terms in conflict with each other within the phenomena under investigation (Saldana, 2013, p. 115). This coding approach was performed first, to compare the claims the government made on the assessment policy and practices of the current curriculum over the previous curriculum from the practitioners' perspectives. Secondly, this comparison was also used to see whether the previous

assessment policy actually equips school practitioners in enacting the policy changes.

To some extent, this study also applied dramaturgical coding method, which approaches participants' account as social drama (Saldana, 2016), often characterised by the emergence of conflict within a society. "Dramaturgical coding is appropriate for exploring intrapersonal and interpersonal participant experiences and actions…", as it "…provides a deep understanding of how humans in social action, reaction, and interaction interpret and manage conflict (Saldana, 2013, p.123). During the interview, some of the respondents revealed the conflicts and obstacles they faced while enacting the assessment policy and the coping strategies to deal with the challenges. In addition, they also disclosed about their emotions and attitudinal impact of the implementation. These narratives, encompassing the struggles, adaptive tactics, and affective responses, collectively constitute key analytical elements of dramaturgical method.

However, structural coding method was the most relevant method to describe the coding approach in this study during the process. This is because many features of structural coding method were applied in this study (Saldana, 2016), including:

- Content-based or conceptual phrase representing a topic of inquiry to a segment of data that relates to a specific research question used.
- Appropriate for virtually all qualitative studies, but particularly for those employing multiple participants, standardised or semi-structured data-gathering protocols, hypothesis testing, or exploratory investigations to gather topics lists or indexes of major categories or themes.
- Similarly, coded segments are then collected together for more detailed coding and analysis to frame the interview.
- Coding and initially categorising data corpus to examine comparable segments' commonalities, differences, and relationships.
- More suitable for interview transcripts than other data.

In the second cycle coding, related and similar codes from the first cycle were clustered into initial themes. These initial themes reflect the pre-determined and emerging issues guiding the interview processes. After all codes were organised into preliminary themes, each theme was carefully reviewed to ensure that all codes within are relevant to the assigned theme. Any codes found to be irrelevant under certain theme were either relocated to more appropriate categories or accommodated within newly created theme. This second cycle coding concluded with the establishment of refined themes based on the coherence and thematic relevance across the codes.

In the third cycle coding, the initial themes and the related codes were reviewed and compared to identify any patterns. The themes with relevant connection and which have overlapping elements were merged, forming broad themes concerning the enactment of assessment reform by the principals and teachers in their schools. All the themes are presented in detail in Chapter Five (Qualitative Findings).

These themes represent the main qualitative findings from the interviews and provide answers to research question three of the study. The themes were also used to further interpret and elucidate the quantitative findings from phase one, represented by research questions one and two. This marks the integration of the quantitative and qualitative data reflecting the key feature of a mixed method research (Creswell & Plano Clark, 2018).

Specifically, to explain the relationship between principal assessment literacy and teacher assessment literacy (i.e., research question four), further analysis through paired interview was performed. The researcher identified and isolated the interview transcripts from matched principal-teacher pairs coming from the same schools. The themes and its relevant codes were identified, analysed and interpreted to understand the relationship between principal and teacher assessment literacy.

Simultaneously, all the themes addressed the four research questions and provide a thorough understanding of principal and teacher assessment literacy.

7. Ethical Issues

This research adhered to the University of Nottingham Code of Research Conduct and Research Ethics. Before conducting the study, ethical approval from the School of Education Research Ethics Committee was acquired (see Appendix 5: Research ethics approval), ensuring that the researcher was aware of his research ethical conduct responsibility and the ethical implications for the participants, government officials, social and cultural values.

Secondly, approval to access the research sites and participants were sought. The schools are under local government administration in the Indonesian education system instead of the MoE. Therefore, permission to get access to principals and teachers from local education authorities were requested through a formal letter (see Appendix 6: Formal request letter to local education authorities for accessing research site and participants) explaining the purpose of the research, the participants, activities conducted during data collection, types of data collected, and potential benefit for the schools, participants, and the local education authority.

Before the data collection began, the participants were informed of the study's purpose and potential benefits. They were assured that participation was voluntary with the right to withdraw at any time. Confidentiality and anonymity were guaranteed, and it was emphasised that their responses would pose no risk. Online survey participants provided consent by clicking on a statement, placed at the end information sheet (see

Appendix 7: Information sheet for survey participants), before proceeding to the questionnaire, while the interview participants signed an informed consent form before the interview (see

Appendix 9: Interview participant consent form).

The use of an online questionnaire allowed the participants to respond at any time within 6 weeks, so that it did not disrupt their regular activities, while the interviews were scheduled at a convenient time for the interviewees. All data collected was

treated according to a mandatory Data Management Plan (DMP) under university data management policy, which were reviewed during the research ethics application.

At the data analysis and final report stage, a full range of findings were disclosed, which became the basis for the quantitative and qualitative analyses and interpretation. The survey respondents had the opportunity to communicate with the researcher to discuss any unclear instructions or items in the questionnaire. To check for the accuracy and credibility of the interview data, triangulation (i.e., between principals and teachers) and member checking methods were performed (Creswell, 2012). Additionally, interview participants may have a copy of the study results should they wish.

8. Summary

This study adopted a pragmatic research paradigm encompassing the ontological, epistemological, and methodological views in developing knowledge. As the consequence of the adopted research philosophy, this study employed a mixed-method multiphase design in addressing the research questions. This study conducted three phases in the following order: mixed-method, quantitative, and quantitative. The justification of the research approach is also provided.

Phase one aimed to develop two quantitative data collection instruments for principal and teacher. In doing so, document analysis and a pilot study were conducted. The resulting data collection instruments then were used in the second phase. The second phase were intended to answer research questions one and two. Results from phase one, then followed up in the third phase. The third phase were designed to provide answers to research questions three and four, whilst deepening the findings from second phase. Structural Equation Modelling (SEM), Confirmatory Factor Analysis (CFA), and Item difficulty analysis were used in the quantitative analysis, whilst content analysis was performed in the qualitative phase.

The population of this study was school principals and social studies teachers at junior secondary school, who were implementing new national curriculum including the assessment reform. Six hundred and forty-five respondents in total participated in the quantitative phase and fourteen participants in the qualitative phase. In interacting with the research participants, this research followed the guideline set out in the University of Nottingham Code of Research Conduct and Research Ethics.

The following chapter (Chapter Four) presents the findings of phase one of the study, which addressed research questions one and two.

Chapter four: Quantitative findings

1. Introduction

This chapter presents the findings of the quantitative phase (phase one) of the study, addressing research questions one and two:

1. RQ1: *What is the level of Indonesian principals' and social studies teachers' assessment knowledge and skills?*

2. RQ2: *How do assessment-related experiences, assessment courses in higher education, in-service assessment professional development programmes, and local support contribute to Indonesian principals' and social studies teachers' assessment knowledge and skills?*

The presentation of the findings follows the order of the research questions. First, the chapter presents the principal and teacher assessment knowledge and skills levels, including its dimensions, described earlier in previous chapter. Second, the findings of the impact of the four predictor variables on assessment knowledge and skills are presented, encompassing the instruments' validity and reliability tests, the causal relationship model between the assessment knowledge and skills and the predictors, and the analysis of the indirect effect within the models.

These findings offer a picture of the current state of assessment knowledge and skills among school leaders and social studies teachers in Indonesia, as well as the nature and influence of relevant factors that affect their knowledge both directly and indirectly. Furthermore, the findings serve as the basis for further investigation of the principals' and teachers' assessment literacy in enacting the assessment reform in their respective schools, particularly in relation to the diverse challenges they encounter in their local contexts. This is especially important given the critical role of assessment literacy in supporting the success of the current education reform in Indonesia.

2. School principals' and social studies teachers' assessment knowledge and skills

This section provides findings about the assessment knowledge and skills levels of principals and social studies teachers, as measured using PALI (Principal Assessment Literacy Inventory) and TALI (Teacher Assessment Literacy Inventory), consisting of 25 and 22 multiple-choice items respectively. The overall assessment knowledge and skills scores of school principals and social studies teachers are presented first, followed by the scores of its assessment dimensions to provide a detailed picture of the school principals' and teachers' assessment literacy.

As the assessment knowledge and skills inventories for principals and social studies teachers are newly developed instruments, a test item analysis was also conducted to evaluate how well the items were functioning to measure the principal and teacher assessment competencies. The item analysis was conducted using Classical Test Theory (CTT), focusing on the item difficulty analysis (Kline, 2005). The results serve as the initial examination of the reliability and validity of the instruments before going into more specific reliability and validity tests in the subsequent sections.

2.1 Principals' assessment knowledge and skills

The average principals' assessment knowledge and skills levels is 54, with a mean score of 11.83 out of 22 and standard deviation of 3.38. Standardising the average score using 100-scale (i.e., 54 out of 100) is necessary to enable later comparison with two different data sets: the teacher scores in this study and the teacher and principal scores from other studies, despite heterogenous assessment literacy instruments with varying scales used across studies. The score range among the principal survey respondents were wide, with the minimum score of 3 and the maximum score of 19. Table 12 below presents the scores.

Table 12

Principals' Assessment Knowledge and Skills Scores

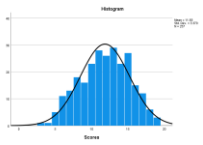
Group	N	Min	Max	Mean	SD[a]	Skewness
PAL	257	3	19	11.83	3.38	-0.15

Note. [a] SD = Standard Deviation

The principals' assessment knowledge and skills scores were normally distributed with the skewness value of -0.15 and is depicted in the following histogram.

Figure 3

Principals' Assessment Knowledge and Score Distribution



The principals' assessment knowledge and skills encompass five assessment areas/dimensions, which have been described in the methodology chapter. The average scores of each of the assessment dimensions are as follow (each correct item was scored 1 and 0 if incorrect):

Table 13

Principals' Scores on Each Dimension of Assessment Knowledge and Skills

Assessment Dimension	Σ item	Max Score	Mean	% Mean	SD
Purpose of assessment (P-PA)	4	4	2.55	64%	1.06
Assessment support for teachers (P-SUP)	5	5	2.43	48%	1.17
Communicating assessment results (P-CAR)	4	4	2.54	63%	1.08
Ethics and Fairness in Assessment (P-EFA)	5	5	3.22	64%	1.26
Assessment Policy (P-POL)	4	4	2.42	60%	1.05

Among the five dimensions, assessment support for teachers (P-SUP) was the lowest scoring (48%), while purpose of assessment (P-PA) and ethics and fairness in assessment (P-EFA) were the highest, both scored 64%.

To reveal whether the PALI and its items are a good measure for the principals' assessment knowledge and skills, the following table below describes the results of the item analysis, presenting the item difficulty levels and categories, detailing how the items and the inventory as a whole measured the school leaders' assessment knowledge and skills.

Table 14

PALI Item Difficulty Levels

Assessment Dimension	Item	Item difficulty	Difficulty category
Purpose of Assessment (P-PA)	P-PA-1	0.62	Moderate
	P-PA-2	0.75	Easy
	P-PA-3	0.84	Easy
	P-PA-4	0.34	Moderate
Assessment support for teachers (P-SUP)	P-SUP-1	0.50	Moderate
	P-SUP-2	0.64	Moderate
	P-SUP-3	0.44	Moderate

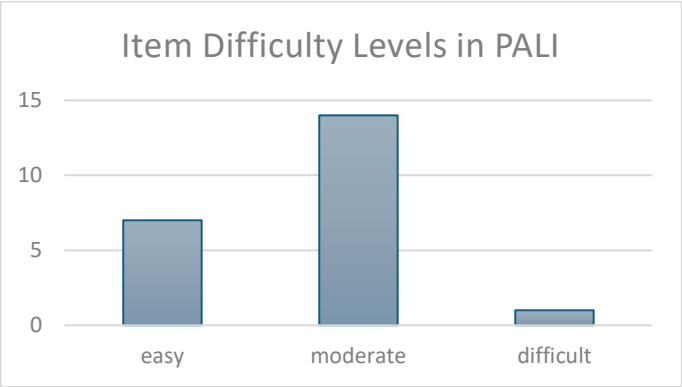
	P-SUP-4	0.60	Moderate
	P-SUP-6	0.24	Difficult
Communicating Assessment Results (P-CAR)	P-CAR-1	0.53	Moderate
	P-CAR-2	0.79	Easy
	P-CAR-4	0.81	Easy
	P-CAR-5	0.41	Moderate
Ethics and Fairness in Assessment (P-EFA)	P-EFA-2	0.74	Easy
	P-EFA-3	0.70	Moderate
	P-EFA-4	0.81	Easy
	P-EFA-5	0.54	Moderate
	P-EFA-6	0.45	Moderate
Assessment Policy (P-POL)	P-POL-1	0.65	Moderate
	P-POL-2	0.42	Moderate
	P-POL-3	0.72	Easy
	P-POL-5	0.63	Moderate

Item difficulty describes the proportion of test takers who answer a test item correctly, ranging from 0 to 1, where values close to 1 mean very easy items and values close to 0 indicate very difficult items (Kline, 2005). In order to reflect the test takers' ability better (in this study, the principals' and teachers' assessment knowledge and skills), it is commonly recommended that more of the test items should have moderate level of difficulty, ranging from 0.30 to 0.70, than those with higher or lower difficulty level. Some items with higher difficulty (lower than 0.30) and lower difficulty (higher than 0.7) levels are still needed to increase the sensitivity of the test to discriminate between different capability levels of the test takers. By including such items, the test will be more sensitive to identify those having very high capability levels and those with very low capability.

Below is the diagram depicting the distribution of item difficulty levels in PALI.

Figure 4

Configuration of PALI Item Difficulty Levels



As shown from the table and diagram above, nearly two thirds of the items in PALI were at moderate level of difficulty (14 out of 22 items), while seven items were considered easy, and only one was difficult. Therefore, given the distribution of the items' difficulty levels, the PALI is arguably an accurate measure of principal assessment knowledge and skills.

2.2 Social studies teachers' assessment knowledge and skills

The average teacher assessment knowledge and skills levels was 59%, with a mean score of the 14.73 out of 25 and standard deviation of 3.74. As with the principals, the score range among teacher respondents was wide, with the lowest score of 4 and the highest score of 24. Table 15 below presents the scores.

Table 15

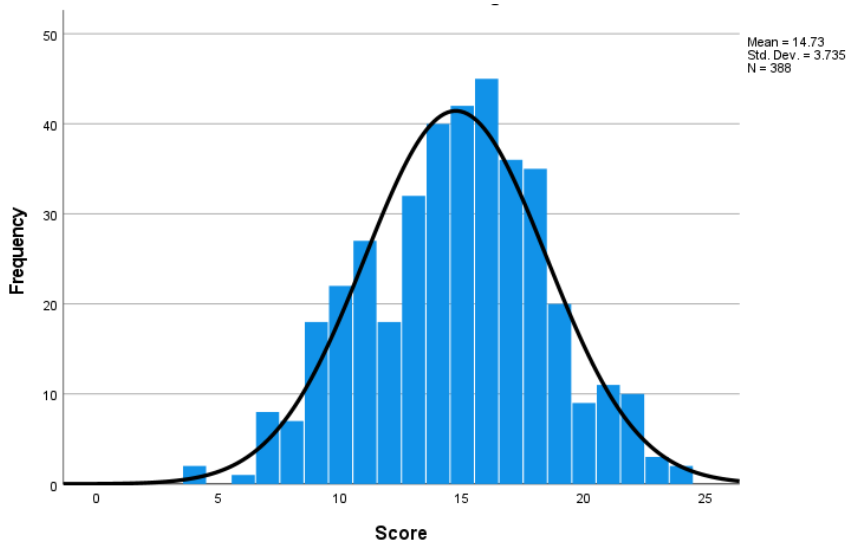
Teachers' Assessment Knowledge and Skills Scores

Group	N	Min	Max	Mean	SD	Skewness
Teacher	388	4	24	14.73	3.74	-0.11

The TAL scores were normally distributed as the skewness level (-0.11) was within the acceptable range of a normal curve, which is -2.0 to +2.0 (Hair et al., 2010). The scores distribution is depicted in the following histogram.

Figure 5

Teacher Assessment Knowledge and Skills Score Distribution



The teacher assessment knowledge and skills in this study covers six assessment areas/dimensions, explained in detail earlier in Chapter Three. The average scores of each of the assessment dimensions are as follow (measured with dichotomy items where each correct item was scored 1 and incorrect item was scored 0):

Table 16

Teachers' Scores on Each Dimension of Assessment Knowledge and Skills

Assessment Dimension	Σ item	Max Score	Mean	% Mean	SD
Purpose of Assessment (T-PA)	4	4	2.6	65%	1.10
Constructing and Administering Assessment (T-CAA)	4	4	2.8	70%	1.05
Analysing and Interpreting Assessment (T-AIA)	4	4	1.4	35%	0.94
Communicating Assessment Results (T-CAR)	4	4	2.2	55%	0.97
Validity and Reliability of Assessment (T-VRA)	4	4	2.1	53%	0.88
Ethics and Fairness in Assessment (T-EFA)	5	5	3.5	70%	1.2

Among the six TAL dimensions, analysing and interpreting assessments (T-AIA) was the lowest scoring dimension, with an average score of 1.4 out of 4 (35%), while constructing and administering assessments (T-CAA) and ethics and fairness in assessment (T-EFA) were the highest scoring, both were 70%.

Turning to the TALI item analysis, the following table presents the results of the item analysis, including the item difficulty levels and categories, providing information of how the items functioned in measuring the assessment literacy of social studies teachers.

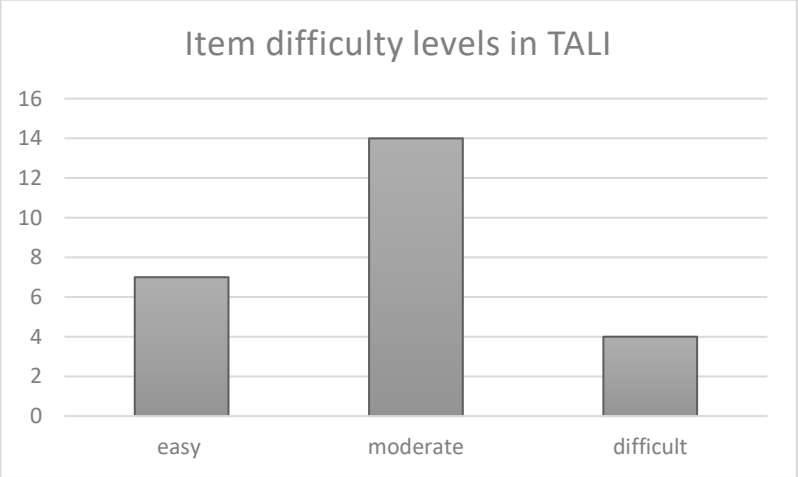
Table 17
TALI Item Difficulty Levels

Assessment Dimension	Item	Item difficulty	Difficulty category
Purpose of Assessment (T-PA)	T-PA-1	0.50	Moderate
	T-PA-2	0.70	Moderate
	T-PA-3	0.67	Moderate
	T-PA-4	0.77	Easy
Constructing and Administering Assessment (T-CAA)	T-CAA-1	0.86	Easy
	T-CAA-3	0.59	Moderate
	T-CAA-5	0.61	Moderate
	T-CAA-6	0.76	Easy
Analysing and Interpreting Assessment (T-AIA)	T-AIA-1	0.40	Moderate
	T-AIA-3	0.16	Difficult
	T-AIA-4	0.21	Difficult
	T-AIA-6	0.64	Moderate
Communicating Assessment Results (T-CAR)	T-CAR-1	0.60	Moderate
	T-CAR-2	0.65	Moderate
	T-CAR-4	0.70	Moderate
	T-CAR-5	0.28	Difficult
Validity and Reliability of Assessment (T-VRA)	T-VRA-1	0.82	Easy

	T-VRA-2	0.44	Moderate
	T-VRA-3	0.63	Moderate
	T-VRA-5	0.25	Difficult
Ethics and Fairness in Assessment (T-EFA)	T-EFA-1	0.62	Moderate
	T-EFA-2	0.72	Easy
	T-EFA-3	0.73	Easy
	T-EFA-4	0.74	Easy
	T-EFA-6	0.69	Moderate

The distribution of item difficulty levels in TALI is depicted in the following diagram.

Figure 6
Configuration of TALI Item Difficulty Levels



As can be seen from the table and diagram, just over half the items in TALI were at moderate difficulty level (14 items out of 25), fewer items are considered easy (7 out of 25) and difficult (4 out of 25). Therefore, based on the configuration of the items' difficulty levels, the TALI is a good measure for the social studies teacher assessment literacy.

The influence of assessment-related experiences; assessment courses in higher education; in-service trainings in assessment; and assessment-related support from colleagues, professional community, and superior on assessment knowledge and skills is described in the following sections.

3. Factors affecting teachers' and principals' assessment knowledge and skills

In this study, as discussed in chapter two, four factors were hypothesised to affect teachers' and principals' assessments literacy: assessment-related experiences (experience), assessment courses in higher education (courses), in-service professional development programmes (trainings) in assessment, and assessment-related local support (support). Each of the four predictors in TAL and PAL model were measured by three items respectively (see chapter 3). The impact of these four factors was analysed using Structural Equation Modelling (SEM) in AMOS 28.0. Before conducting the SEM analysis for the proposed models, a number of statistical tests were performed including several reliability and validity tests of the measures in the study to provide foundation for the subsequent analyses (SEM).

Cheung et al. (2024) recommended incorporating multiple criteria in examining the reliability and validity of a study's instruments. In a study applying a SEM analysis, the first step to examine the reliability and validity of the measures is by establishing the measurement model through Confirmatory Factor Analysis (CFA). When the CFA model fulfils the overall fit criteria indicating the proposed model fit with the data well, it establishes basic requirements of reliability and validity tests (Cheung et.al, 2024).

3.1 Measurement (CFA) model of social studies teachers' assessment knowledge and skills

Assessment knowledge and skills is a complex concept consisting of several key dimensions. These dimensions, to a relative extent, may associate with each other. Therefore, though not always, when an item is intended to measure one dimension, for example, purpose of assessment, it may also, to a certain degree, unintentionally measure another dimension such as analysing assessment results or it has dual factor loading. For example, an item in TALI previously was written: "*A teacher analysed mid-semester assessment results and use the results to improve the lesson plans that she has developed at the beginning of the semester. The purpose this action is … (answer options: a. formative, b. summative, c. diagnostic)*". This item

contains two dimensions and might load to both dimensions. This item was removed from the scale.

Consequently, in the measurement model (the CFA) of assessment knowledge and skills, an item intended to measure one dimension might also load to/measure another dimension (i.e., having cross-loading), or its residual variance might correlate with residual variance of other items within the same or different dimension (Goretzko et al., 2024).

This issue can be addressed using item parcelling technique. Parcelling clustered items (i.e., calculating the composite/average scores of a group of items) aimed at measuring a particular dimension/factor may produce a new simpler CFA measurement model that reduces unique variances of its items that may lead to correlated residual variances or dual item loading to trivial and non-influential magnitude (Little et al., 2013).

Item parcelling technique was applied in both PAL and TAL models, particularly in their respective measurement (CFA) models, to simplify the models' structures and prevent under-identification. Given the moderate-low number of respondents relative to the number of parameters to be estimated in the models, parcelling clustered items within each of the PAL and TAL dimensions reduced model complexity. Without this technique, the CFA models would contain excessive parameters, rendering by the sample size in this study insufficient to provide robust analysis for the relationships among variables in the models. Little et al. (2013) maintained that item parcelling in SEM can provide valid, reliable, efficient manifest (observed) variables for a factor with good measurement properties when carefully crafted.

To examine the TAL measurement model, second order CFA was first conducted (Tayag & Ayuyao, 2020). This step identified that 4 of the 25 items in TALI have non-significant p values (i.e., p values higher than 0.05): T-VRA5 (p value = 0.16), T-VRA2 (p value = 0.76), T-AIA4 (p value = 0.17), and CAR2 (p value = 0.18). These four items were then removed in the subsequent analysis. Four other constructs in the TAL

CFA model (Experience Courses, Training and Support) were measured by three items, as discussed in methodology chapter, wherein all of them statistically significant.

The 21 statistically significant items in CFA TAL were parcelled based on the 6 dimensions of the teacher assessment knowledge and skills. The composite score (i.e. the average score) of the clustered items in each dimension was calculated and become the indicator for TAL. As a result, the TAL construct has six indicators. The following two diagrams depict the measurement (CFA) model of TAL before and after using item parcelling.

Figure 7

TAL Measurement (CFA) Model Before Item Parcelling

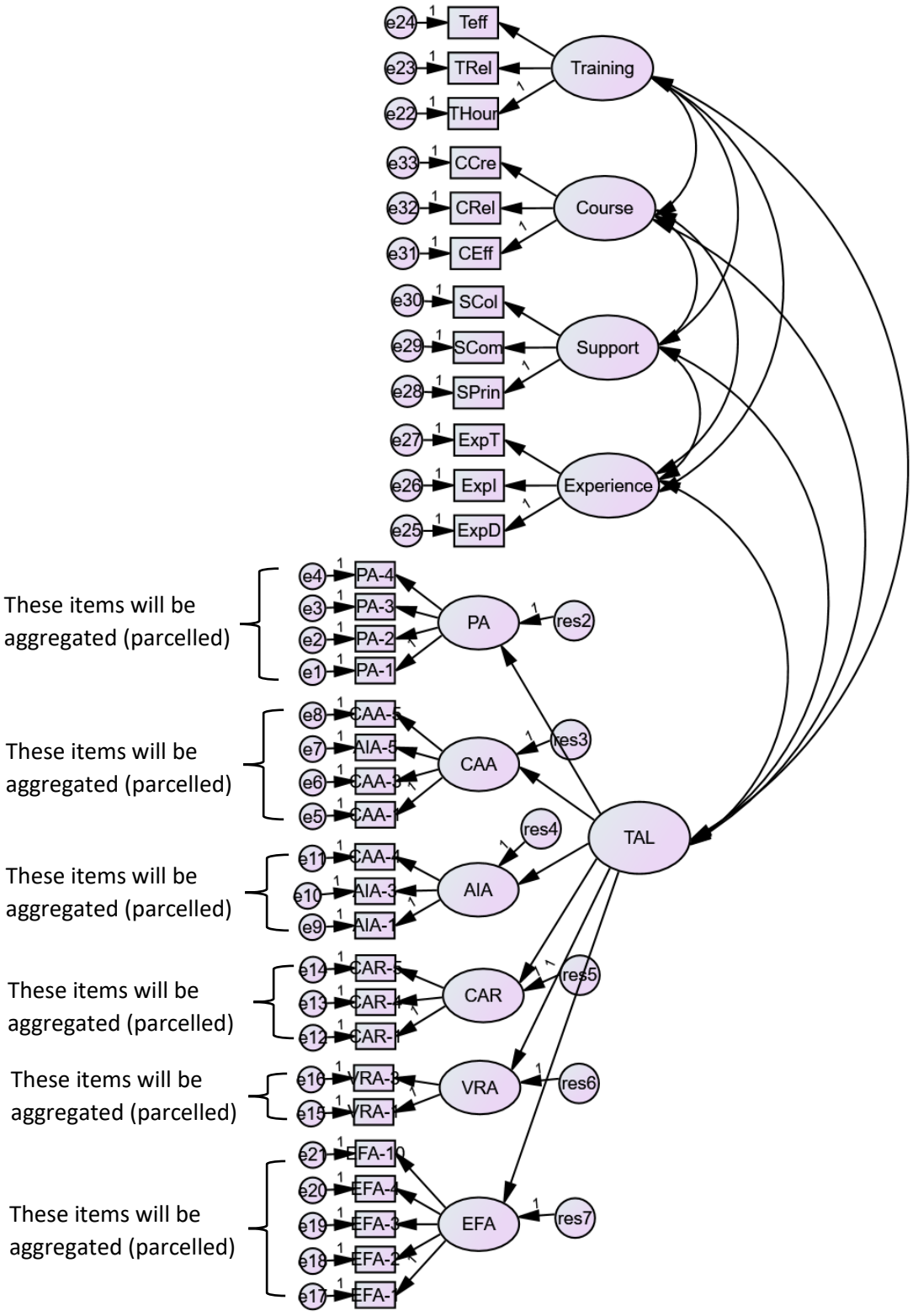


Figure 7 shows the TAL CFA model without parcelling technique. This unparcelled model incorporates all individual items within each assessment literacy dimension,

for example, PA1, PA2, PA3, PA4 for the purposes of assessment (PA) dimension. This makes the CFA model highly complex.

Figure 8

TAL Measurement (CFA) Model After Item Parcelling

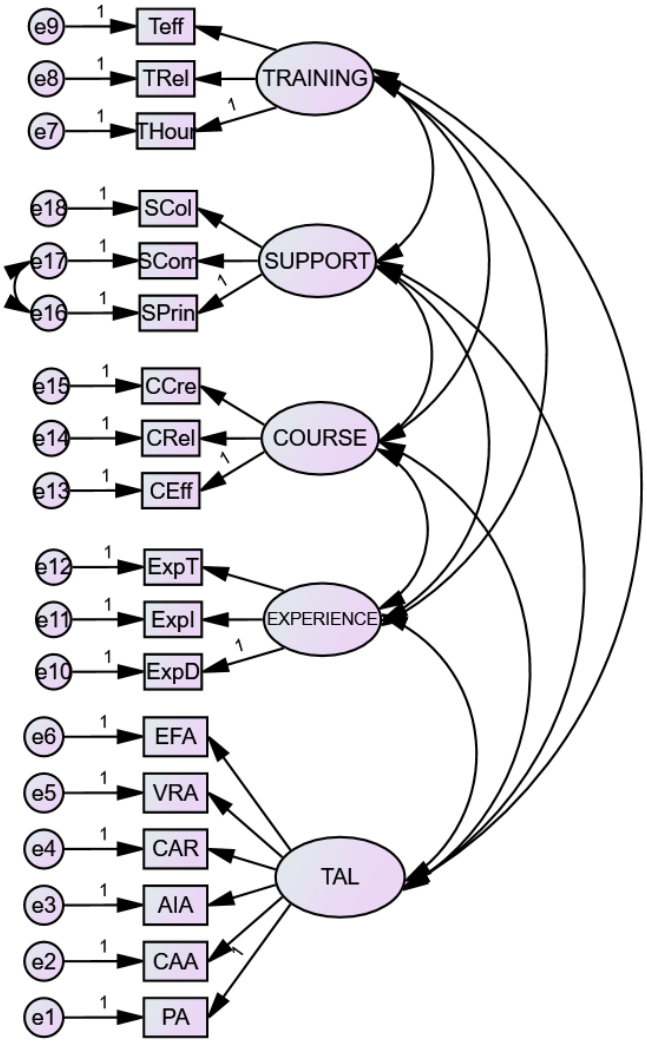


Figure 8 demonstrates refined TAL CFA model employing item parcelling, where the clustered items within each assessment literacy dimension were aggregated, creating composite indicators (i.e, PA, CAA, AIA, CAR, VRA and EFA) for assessment literacy construct. This parcelling approach significantly simplifies the model structure.

The TAL measurement (CFA) model fit was assessed against chi-square test statistics and several common model fit criteria, including CFI, TLI, RMSEA, and SRMR (Hu & Bentler, 1999) (see chapter three) to determine whether the proposed model fits with the data, meaning that the model is applicable in the population of the study. The combination of several criteria was used to provide thorough evidence of the proposed model fitness (Byrne, 2010).

All obtained values of TAL CFA model were within their respective acceptance level and thus yielded a good fit for the data as shown in table 18 below.

Table 18

TAL CFA Model Goodness of Fit Indices

Fit Indices	Required value	Obtained value
χ^2 /df	< 3.0	1.68
CFI	≥ 0.95	0.97
TLI	≥ 0.90	0.97
RMSEA	< 0.06	0.039
SRMR	< 0.08	0.046

It is necessary to note that some indicator variables in each factor/dimension, except for assessment-related support factor, have factor loadings less than 0.4 (shaded) as shown in the following table. Nevertheless, these items/indicators were kept in the model as they are critical in reflecting the constructs of interest and that by removing them might result in inadequate representation of the constructs (Cortina et al., 2020) or in other words reducing the content validity.

Table 19

TAL Item Factor Loadings

Indicator		Factor	Estimate
T-PA	<---	TAL	0.49
T-CAA	<---	TAL	0.58

T-AIA	<---	TAL	0.36
T-CAR	<---	TAL	0.44
T-VRA	<---	TAL	0.39
T-EFA	<---	TAL	0.65
TrainHour	<---	TRAINING	0.27
TrainRel	<---	TRAINING	0.98
TrainEff	<---	TRAINING	0.98
ExpDev	<---	EXPERIENCE	0.86
ExpInst	<---	EXPERIENCE	0.27
ExpTeach	<---	EXPERIENCE	0.43
CoursEff	<---	COURSE	0.98
CoursRel	<---	COURSE	0.97
CoursCre	<---	COURSE	0.26
SupPrin	<---	SUPPORT	0.78
SupCom	<---	SUPPORT	0.83
SupCol	<---	SUPPORT	0.98

The confirmatory factor analysis (CFA) of TAL model constructs shows that the measurement model of the proposed model achieves the model fit criteria despite of few low factor loadings (i.e., < 0.4) in the model. The reliability and validity of the measures are discussed in the following section.

3.2 Reliability and validity of social studies teacher assessment literacy measures

The reliability of the measures was evaluated by calculating the Composite Reliability (CR) values based on standardised regression weights of the Confirmatory Factor Analysis (CFA) of TAL model constructs and their predictors. The composite reliability is a more appropriate reliability measure for a SEM-based study, such as

this study, rather than Cronbach's Alpha since CR allows for unequal factor loadings across items of the constructs (Cheung et.al, 2024).

The composite reliability (CR) of TAL construct (using the six dimensions as the indicators) is 0.65 and the CR of the TAL predictors are as follows: Course (0.83), Training (0.83), Support (0.91), and Experience (0.55). Common recommended values as the benchmark for a satisfactory to good reliability for a measure of a construct in a SEM is equal or more than 0.70 (Hair Jr. et al., 2021). However, Hair Jr. et. al. also contends that in an exploratory investigation, values between 0.60 and 0.70 are still acceptable. In this study's newly developed data collection instrument, the Experience constructs have reliability level (0.55) approaching the acceptable value (0.60). Nevertheless, while acknowledging this as the limitation, the researcher retained the construct because first, it is crucial to the whole hypothesised model, second, the reliability value was only slightly below the acceptable level, and thirdly, two of the three items of the construct had good factor loadings and all of them were statistically significant. Table 20 (column 3) shows the CR of the constructs of the TAL model.

Table 20

CR, CV, and DV of TAL and Predictors' Scale

No.	Construct	Composite Reliability (CR)	Convergent Validity (AVE)	Discriminant Validity (Sq. root AVE)
1	2	3	4	5
1.	Teacher Assessment Literacy	0.65	0.89	0.94
2.	Training	0.83	0.66	0.81
3.	Course	0.83	0.66	0.81
4.	Support	0.91	0.78	0.88
5.	Experience	0.55	0.32	0.57

The construct validity of the TAL questionnaires was evaluated for both its convergent validity and discriminant/divergent validity. Convergent validity (CV) shows how close the indicator variables converge in measuring the underlying construct (Hair Jr. et al., 2021), which in a SEM-based study, the measurement (CFA) model must first achieve model fit criteria thresholds as the fundamental

requirement for calculating its convergent validity (Cheung et.al, 2023). The TAL measurement model in this study has fulfilled the model fit criteria as noted in the previous section.

Average Variance Extracted (AVE) score is used to indicate the instruments' convergent validity value where scores 0.5 or higher is expected (Hair et al., 2019). The convergent validity values of TAL and its predictor latent variables are presented in table 17 above in column 4.

Except for the Experience construct, which AVE is 0.32, all other constructs were above 0.5 and thus showing convergent validity. The assessment-related experiences latent construct in the TAL instrument was measured using three very clear items regarding the respondents' experience in student assessment using simple and short questions: "how long have you been a teacher", "how many times have you been an instructor in an assessment trainings/workshops", and "how many times have you served as national/provincial/local test developer". Therefore, substantively speaking, these three items specifically measured the teachers' assessment-related experience, and collectively capture the key aspects of the construct.

Discriminant validity (DV), on the other hand, is used to check whether the constructs in a model are statistically distinct from one another (Hair Jr. et al., 2021). Two methods of examining the discriminant validity of the constructs in the TAL model were performed: Fornell and Larcker criterion and Heterotrait-Monotrait (HTMT) ratio. First, using Fornell and Larcker (1981) criterion, the discriminant validity of the constructs in TAL model were tested by calculating the square root of AVE of the construct and then comparing the values with the correlation coefficients between the constructs with all other constructs in the model. Discriminant validity of a construct is evident when the square root of AVE value of the construct is higher than all other correlation coefficients between the construct and other constructs. Table 18 shows the square root AVE scores of all the constructs in the model (bold) and all correlation coefficients of the constructs.

Table 21

Examination of DV TAL Using Fornell and Larcker Criterion Matrix

Constructs	TAL	Training	Course	Support	Experience
TAL	**0.94**				
Training	0.02	**0.81**			
Course	-0.21	0.74	**0.81**		
Support	-0.01	0.66	0.53	**0.88**	
Experience	0.29	0.15	0.09	0.17	**0.57**

The table shows that the square root AVE of each construct in TAL model (bold) was higher than any other correlation coefficients of the constructs with all other constructs in the model, thus the discriminant validity of the TAL model is evident.

Secondly, the discriminant validity of the constructs was also checked by calculating the Heterotrait-Monotrait (HTMT) ratio between the constructs. It examines the ratio of the average correlations of the indicators between two constructs (heterotrait) to the average correlations of the indicators within each of the two constructs (monotrait) (Henseler et al., 2015). A discriminant validity is fulfilled when the HTMT ratio between two constructs is ≤ 0.90 (Hair et al., 2019). The table 22 below shows the HTMT ratio between all the constructs in TAL model.

Table 22

Examination of DV TAL Using HTMT Ratio

HTMT	TAL-Cours	TAL-Sup	TAL-Exp	TAL-Train	Cours-Sup	Cours-Exp	Cours-Trai	Sup-Exp	Sup-Trai	Exp-Trai
Ratio	-0.18	0.01	0.31	0.06	0.53	0.11	0.82	0.17	0.69	0.13

All the HTMT ratio between the constructs in the TAL model are below 0.90, thus providing another evidence that the model achieves the discriminant validity (DV) criteria.

3.3 Structural model of social studies teachers' assessment literacy

After the measurement (CFA) model of TAL hypothesised model as well as the reliability and validity were established, a structural model representing the impact

of the four predictors on social studies teacher assessment literacy was tested using Structural Equation Modelling (SEM) analysis.

First, the goodness of fit of the model needs to be examined. The fit indices criteria to determine model fitness include Chi-square/DF, as well as CFI, TLI, RMSEA and SRMR (Hu & Bentler, 1999). The fitness of the TAL structural model, which is highly similar with its' measurement model, is described in the following table.

Table 23

TAL Model Goodness of Fit Indices

Fit Indices	Rec. values	Obtained values
χ^2 /DF	< 3.0	1.67
CFI	≥ 0.95	0.97
TLI	≥ 0.95	0.97
RMSEA	< 0.06	0.042
SRMR	< 0.08	0.048

The table shows that the proposed structural model has good model fitness and proves that the model fits with the data. Thus, it can be concluded that the causal relationship represented in the model reflects the relationship between teacher assessment literacy and the predictors within the population of this study.

This study proposes that in TAL model the predictors have direct and/or indirect relationships with teacher assessment literacy. Assessment courses' and assessment support's impact on the assessment literacy were hypothesised to be both direct and indirect mediated through assessment training and assessment experience. Thus, besides as the direct predictors on teacher assessment literacy, assessment training and assessment experience were also hypothesised to act as mediating variables in the model.

The parameter estimates are the regression weights among the constructs in the model which indicate the impact of the predictors on the teacher assessment

literacy either directly or indirectly. The standardised estimates of the relationship between teacher assessment literacy and its predictors are presented in table 24 and in figure 9 as follows. The shaded rows in the table 24 indicate the insignificant relationship between constructs (i.e., p value > 0.05), thus are not shown in the figure 9.

Table 24

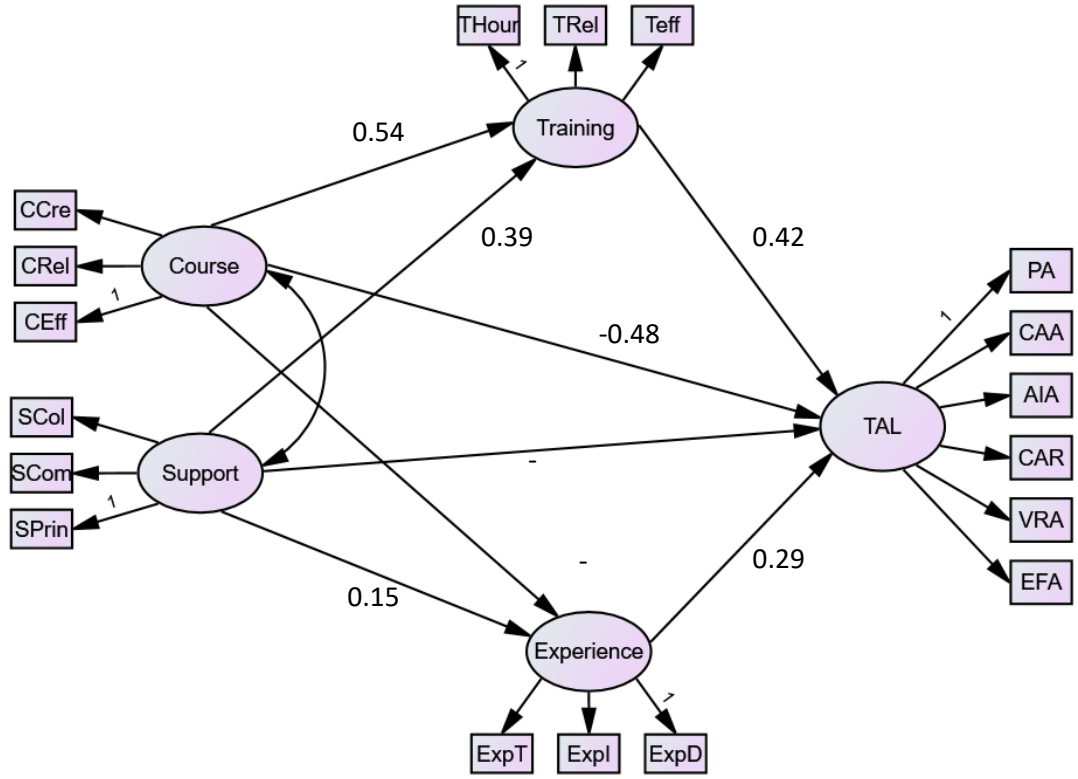
Standardised Estimates of TAL Model Structural Paths

Structural Path			Std.Est	S.E.	C.R.	*p* value
Training	<---	Course	0.54	0.029	5.06	0.000
Experience	<---	Course	0.017	0.077	0.249	0.803
Training	<---	Support	0.39	0.039	4.723	0.000
Experience	<---	Support	0.15	0.136	2.178	0.029
TAL	<---	Training	0.42	0.027	3.084	0.002
TAL	<---	Experience	0.29	0.005	2.868	0.004
TAL	<---	Course	-0.48	0.006	-4.65	0.000
TAL	<---	Support	-0.07	0.008	-0.88	0.38

Note. Std.Est = Standardised Estimates, S.E = Standard Error, C.R = Critical Ratio

Figure 9

Standardised Estimates of TAL Model's Structural Paths



The results indicate that most of the structural paths (the arrows pointing from one construct to another) in the TAL model were within the range of moderate strength (Cohen, 1988). Except for two parameter estimates having insignificant p values (shaded in the table), resulting in hypothesis (H7 and H12) rejection: first, from assessment course to assessment experience (p value = 0.80) and second, from assessment support to teacher assessment literacy (p value = 0.38), all other (six) hypothesised paths in the structural model were supported with regression weights ranging from $\beta = 0.15$ (from assessment support to assessment experience) to $\beta = 0.54$ (from assessment courses to assessment training). The complete results of the hypothesis testing of all the direct impact within TAL causal model is presented in the following table.

Table 25

Results of Hypothesis Testing of TAL Model

Hypothesis	Relationship	Std.Est	p value	Result
H1	Course - TAL	-0.48	0.000	Supported
H2	Course - Training	0.54	0.000	Supported
H3	Training – TAL	0.42	0.002	Supported
H4	Course – Experience	0.017	0.803	Not supported
H5	Experience – TAL	0.29	0.04	Supported
H6	Local support – TAL	-0.07	0.38	Not supported
H7	Local support – Training	0.39	0.000	Supported
H8	Local support - Experience	0.15	0.029	Supported

These results indicate that the model represent the approximate 'true' causal relationship among the constructs in the population as indicated by its fit indices (see **Table 23**

TAL Model Goodness of Fit Indices). In addition, the strengths of the impact of the predictors on social studies teacher assessment literacy either directly or indirectly were mostly moderate. However, the results also pointed out that two of the hypothesised relationships were not evident. First, the assessment supports that the teachers received from various sources (colleagues, teacher community, and principal) do not directly affect the teachers' assessment literacy. Secondly, the teachers' assessment-related courses in higher education do not affect their assessment-related experiences. The reason for this was elaborated in the second phase of the study through semi-structure interviews. In addition, the parameter estimate from assessment courses in higher education to teacher assessment literacy is negative (β = -0.48), which mean that the assessment topics that the teachers learned during their pre-service teacher education programmes emphasised different aspects compared to what they currently practice. How these findings contribute to the focus of the study are discussed in Chapter Six.

3.4 Mediation/indirect effect analysis of TAL structural model

As explained in the methodology chapter, SEM analysis allows for a complex causal relationship among the constructs in a model, which might include direct and indirect relationships (Kline, 2016). The indirect effect analysis in this study is important to test the hypothesis that the assessment courses and local support variables have both direct impact to assessment literacy as well as indirect effects through two mediating variables: in-service trainings in assessment and assessment-related experiences.

The indirect impacts of assessment courses and assessment support on teacher assessment literacy through assessment training and assessment experience were analysed using bootstrap technique in AMOS 28. Since there are two mediating variables in the TAL model: assessment training and assessment experience, estimand function in AMOS 28 was also performed to isolate and determine the individual indirect effect of the mediators in the model. The indirect impact analysis results of the TAL structural model are as follows.

Table 26

Indirect effect Analysis of TAL Model

Indirect effects	Std.Est	p values
Course→Training→TAL	0.012	0.000
Cours→Experience→TAL	0	0.739
Support→Training→TAL	0.015	0.000
Support→Experience→TAL	0.004	0.012

The mediation effect of assessment training on teacher assessment literacy for both assessment course (p value = 0.000) and assessment support (p value = 0.000) were statistically present. However, the mediation effect is very low, 0.012 and 0.015 respectively. As shown in the previous structural model analysis (see **Table 24**

Standardised Estimates of TAL Model Structural Paths and Figure 9

Standardised Estimates of TAL Model's Structural Paths), assessment course had significant direct negative impact on teacher assessment literacy (β = -0.48). It means the assessment training only partially mediates the effect of assessment course on teacher assessment literacy. The negative direct impact of assessment courses on teacher assessment literacy might cause the impractical (i.e very low, β = 0.012) mediation effect of assessment trainings. On the other hand, the previous structural model analysis also shows that assessment support had no direct impact on teacher assessment literacy. Therefore, the impact of assessment support on teacher assessment literacy was only indirect and was fully mediated through assessment trainings, even though the mediation effect was also very low (β = 0.015), and thus impractical.

Regarding the mediation effect of assessment-related experience, **Table 26**
Indirect effect Analysis of TAL Model shows that assessment experience only mediates the impact of assessment support (p value = 0.012) on teacher assessment literacy. Assessment experience did not mediate the impact of assessment courses on TAL as indicated by p value = 0.739. In addition, assessment experience fully mediated the impact of assessment support on teacher assessment literacy because assessment support did not have direct impact on teacher assessment literacy. However, the mediation effect of assessment-related experiences is extremely low (β = 0.004) and has no practical significance.

Statistically speaking, three of the four hypothesised indirect effects in the model were supported as shown in hypothesis testing results below. However, the magnitudes of the indirect effects were very or extremely small, thus do not have practical significance, necessitating further elaboration in phase two.

Table 27

Results of Hypothesis Testing of TAL Indirect Analysis

Hypothesis	Relationship	Std.Est	p value	Result
H9	Course - Training - TAL	0.012	0.000	Supported
H10	Course - Experience - TAL	0	0.739	Not supported
H11	Support – Training - TAL	0.015	0.000	Supported
H12	Support – Experience - TAL	0.004	0.012	Supported

The summary of the direct and indirect effect in TAL model is presented the table 28 below.

Table 28

Summary of Direct and Indirect Effect of TAL Model

Relationship	Direct effect	Indirect effect	p value	Conclusion
Course→TAL	-0.48		0.000	Negative direct impact
Course→Training→TAL		0.012	0.000	Partial mediation
Cours→Experience→TAL		0	0.739	No mediation
Support→TAL	0		0.38	No direct impact
Support→Training→TAL		0.015	0.000	Full mediation
Support→Experience→TAL		0.004	0.012	Full mediation

The following sections present the results of the quantitative analysis for the PAL model.

3.5 Measurement (CFA) model of principals' assessment literacy

As in the TAL measurement model, to examine the PAL measurement model, initially, second order CFA was conducted. This step identified 5 of the 22 items in

PAL inventory had non-significant p values ($p > 0.05$) which are: P-PA4 (p value = 0.17), P-SUP2 (p value = 0.73), P-SUP6 (p value = 0.43), P-EFA6 (p value = 0.28), P-POL2 (p value = 0.25). These five items were then removed in the subsequent analysis. Each other construct in the measurement model was measured by three items, as discussed in previous chapter, wherein all of them were statistically significant.

The remaining 17 statistically significant items measuring principal assessment literacy were parcelled into the five dimensions of PAL. The composite score (i.e. the average score) of the clustered items in each dimension was calculated and become the manifest variable (indicator) for PAL. The following two diagrams depict the measurement (CFA) model of PAL before and after using item parcelling (see page 125 in Chapter Three about item parcelling technique).

Figure 10

PAL Measurement (CFA) Model Before Item Parcelling

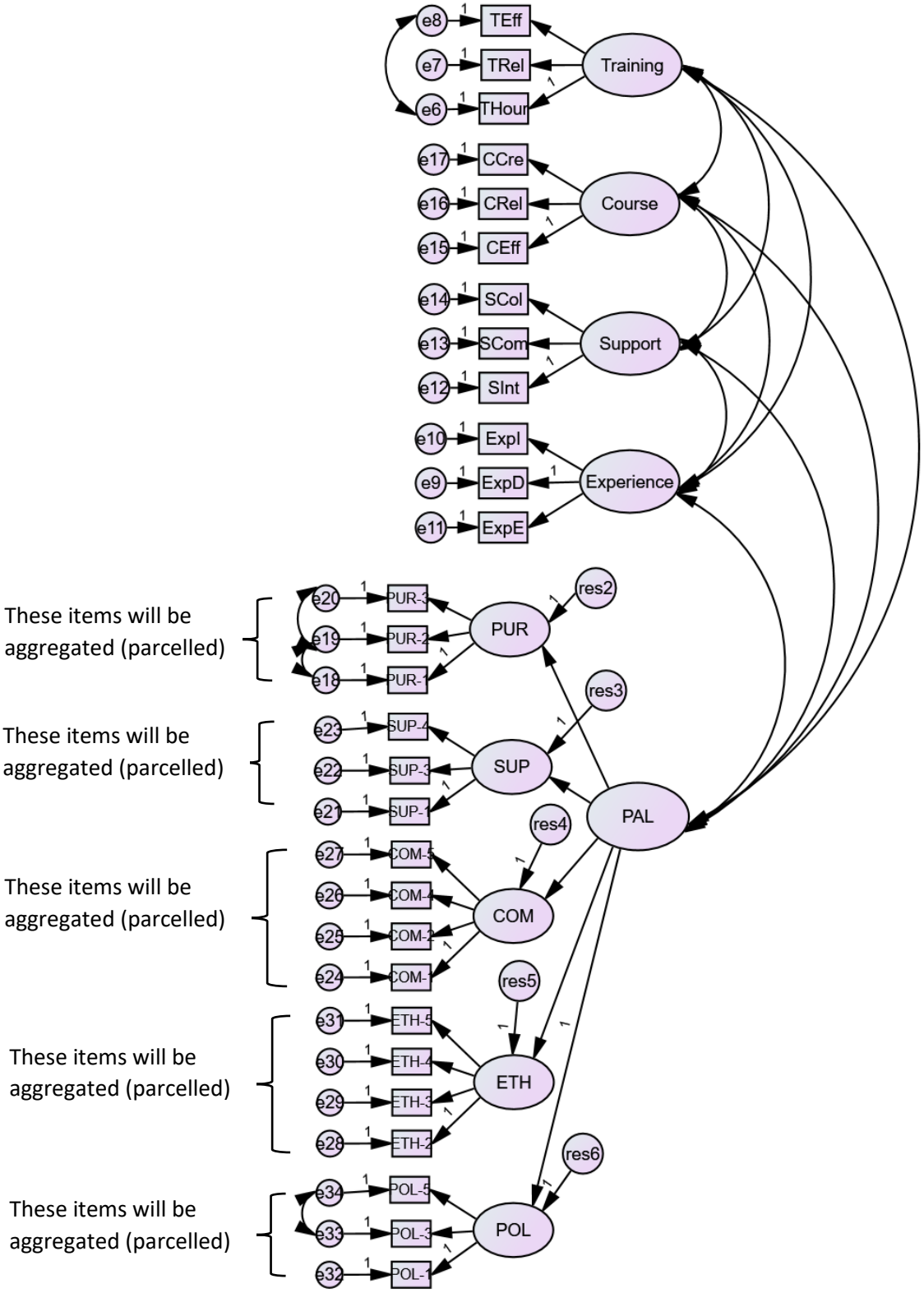


Figure 10 shows the PAL CFA model without parcelling technique. This unparcelled model incorporates all individual items within each assessment literacy dimension, for example, SUP1, SUP3, SUP4 for the Assessment Support (SUP) dimension. This makes the CFA model highly complex.

Figure 11

PAL Measurement (CFA) Model After Item Parcelling

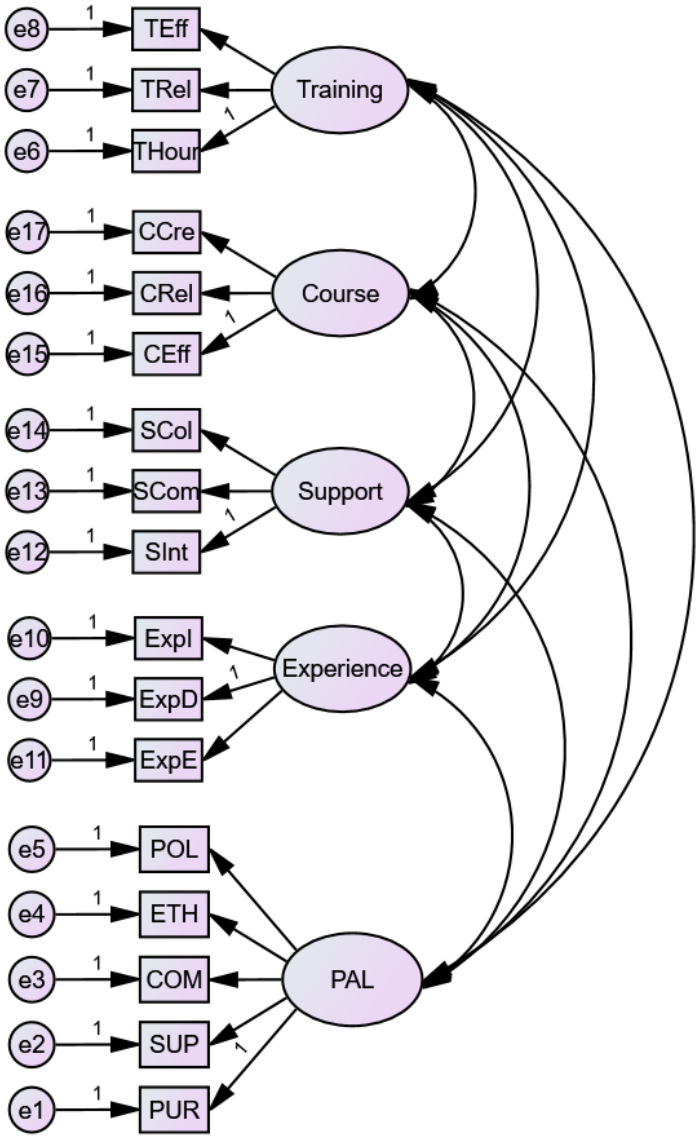


Figure 11 demonstrates refined PAL CFA model employing item parcelling, where the clustered items within each assessment literacy dimension were aggregated,

creating PAL composite indicators (i.e, PA, SUP, CAR, EFA and POL) for assessment literacy construct. This parcelling approach significantly simplifies the model structure.

As in the model for teachers, the PAL measurement (CFA) model fit was evaluated against five goodness-of-fit criteria: chi-square/df, CFI, TLI, RMSEA, and SRMR to determine whether the proposed model fits with the data, meaning that the model is applicable in the population of the study. The PAL measurement (CFA) model achieved all the model fit recommended criteria as presented in table below, thus yielded a good fit with the data.

Table 29

PAL CFA Model Goodness of Fit Indices

Fit Indices	Rec. values	Obtained values
χ^2 /DF	< 3.0	1.77
CFI	≥ 0.95	0.96
TLI	≥ 0.95	0.96
RMSEA	< 0.06	0.05
SRMR	< 0.08	0.06

Note. Rec. values = Recommended values

Table 30 shows the factor loadings of the indicators of all the constructs in PAL measurement model. Compared to TAL measurement model, PAL measurement model had fewer indicators in each factor/dimension with factor loadings less than 0.4 (shaded) as shown in the following table.

Table 30

Factor Loadings of PAL CFA Model

Indicator		Factor	Factor loading
P-PA	<---	PAL	0.52
P-SUP	<---	PAL	0.46

P-CAR	<---	PAL	0.61
P-EFA	<---	PAL	0.56
P-POL	<---	PAL	0.71
TrainHour	<---	Training	0.32
TrainRel	<---	Training	0.98
TrainEff	<---	Training	0.98
ExpDev	<---	Experience	0.73
ExpInst	<---	Experience	0.79
ExpEduc	<---	Experience	0.34
SupInt	<---	Support	0.91
SupCom	<---	Support	0.97
SupCol	<---	Support	0.82
CoursEff	<---	Course	0.96
CoursRel	<---	Course	0.94
CoursCre	<---	Course	0.29

The confirmatory factor analysis (CFA) of PAL constructs shows that the measurement model of the PAL model fulfils the model fit criteria and have only three items with low factor loadings (i.e < 0.4). As in the TAL measurement model noted earlier, these three items with low factor loading were retained to maintain the content validity of the related constructs. These results allow for subsequent reliability and validity analysis of the measures, which is discussed in the following section.

3.6 Reliability and validity of school principal assessment literacy measure

Likewise, the reliability of the measures in PAL model were evaluated by calculating the Composite Reliability (CR) values based on standardized regression weights of

Confirmatory Factor Analysis (CFA) of PAL and its predictors. The five indicators of PAL have Composite Reliability (CR) of 0.71. The Composite Reliability of the PAL predictors are as follows: Course (0.81), Support (0.93), Training (0.84), and Experience (0.67). All the constructs show either acceptable level (Experience = 0.67) or better reliability. The complete composite reliability of the constructs can be found in column 3 of table 31.

Table 31

CR, CV, and DV of PAL and Predictors Scales

No.	Construct	CR	Convergent Validity (AVE)	Discriminant Validity (Sq. root AVE)
1	2	3	4	5
1.	Principal Assessment Literacy (PAL)	0.71	0.98	0.99
2.	Training	0.84	0.68	0.82
3.	Course	0.81	0.63	0.81
4.	Support	0.93	0.83	0.90
5.	Experience	0.67	0.43	0.65

Similarly, the construct validity of the PAL questionnaires was evaluated of both its convergent validity and discriminant/divergent validity. The convergent validity (CV) values of PAL and its predictor variables range from 0.43 to 0.98 and are presented in table 31 column 4.

Except for the Experience construct, which AVE is 0.43, all other constructs are above 0.5 and thus showing convergent validity. The school principal assessment-related experiences construct in the PAL instrument was also measured using very clear items regarding their experience in student assessment using simple and short statements: “how long have you been a teacher”, “how many times have you been an instructor in an assessment-related trainings/workshops”, and “how many times have you been involved as national/provincial/local test developer”. Therefore,

even though the AVE score is slightly below 0.5, these three items clearly measured the principals' assessment-related experience.

Next, the discriminant validity of the constructs in PAL model was also evident as it fulfilled the Fornell and Larcker criterion and HTMT ratio. The square root AVE of each construct was higher than the correlation of the construct with all other constructs in the model (shown in table 32 below) and all the HTMT ratios between the constructs were below 0.90 (table 33).

Table 32

DV PAL Model Using Fornell and Larcker Criterion Matrix

Constructs	PAL	Training	Course	Support	Experience
PAL	**0.99**				
Training	0.28	**0.82**			
Course	0.11	0.79	**0.81**		
Support	0.29	0.74	0.57	**0.90**	
Experience	0.19	0.23	0.29	0.15	**0.65**

Table 33.

DV PAL Model Using HTMT Ratio

PAL-Cours	PAL-Sup	PAL-Exp	PAL-Train	Cours-Sup	Cours-Exp	Cours-Train	Sup-Exp	Sup-Train	Exp-Train
0.10	0.26	0.21	0.27	0.60	0.31	0.86	0.15	0.77	0.25

3.7 Structural model of school principals' assessment literacy

As with the teacher model, the structural model represents the impact of the four predictors on school principal assessment literacy. The first step is to examine the goodness of fit of PAL model. The PAL structural model fulfils all the model fit requirements as indicated in table 34 below.

Table 34.
PAL Structural Model Goodness of Fit Criteria

Fit Indices	Required values	Obtained values
χ^2 /DF	< 3.0	1.76
CFI	≥ 0.95	0.97
TLI	≥ 0.95	0.96
RMSEA	< 0.06	0.05
SRMR	< 0.08	0.06

It can be concluded that the causal relationship represented in the model reflects the relationship between the principal assessment literacy and the predictors in the population of this study.

As in TAL structural model, the relationship between the principal assessment literacy and the four predictors are direct and indirect. The impact of assessment courses and assessment support on the assessment literacy were hypothesised to be both direct and indirect mediated through assessment training and assessment experience. Assessment training and assessment experience in PAL structural model were also hypothesised to have direct impact on principal assessment literacy.

The standardised estimates of the relationship between principal assessment literacy and its predictors are presented in table 35 and in figure 12 below.

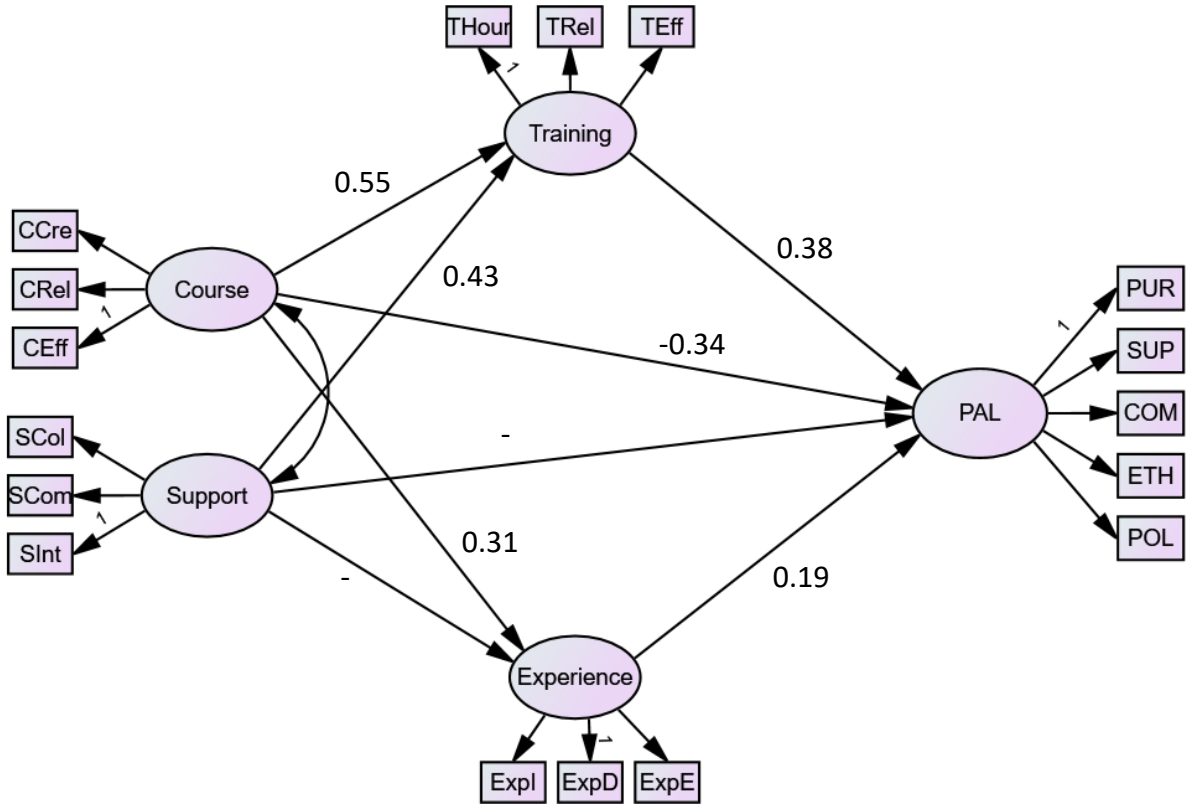
Table 35
Standardised Estimates of PAL Model Structural Paths

Structural Path			Std.Est	S.E.	C.R.	p value
Training	<---	Course	0.55	0.051	4.92	0.000
Training	<---	Support	0.43	0.052	4.68	0.000
Experience	<---	Course	0.31	0.08	3.29	0.001

Experience	<---	Support	-0.03	0.094	-0.37	0.715
PAL	<---	Training	0.38	0.028	2.22	0.027
PAL	<---	Experience	0.19	0.008	2.13	0.033
PAL	<---	Course	-0.34	0.01	-2.56	0.011
PAL	<---	Support	0.14	0.011	1.28	0.199

Figure 12

Standardised Estimates of PAL SEM Model



In the Principal Assessment Literacy (PAL) structural model, two parameter estimates showed insignificant p values (shaded in the table): assessment support to assessment experience (p value = 0.715) and assessment support to principal assessment literacy (p value = 0.199). These results show that assessment support does not directly affect the principal's assessment literacy, nor the principal assessment experience as hypothesised.

The rest (six) of the hypothesised direct impacts in the model were validated (having p value < 0.05) ranging from β = 0.19 (from assessment experience to principal assessment literacy) to β = 0.55 (from assessment courses to assessment training). As in the Teacher Assessment Literacy (TAL) model, the impact of assessment courses in higher education on PAL in this model is also negative (β = -0.34). While the magnitudes of the impact of the predictors on principal assessment literacy either directly or indirectly were mostly moderate, this model represents the approximate impact of the four predictors on principal assessment literacy in the population as indicated through the model fit values. This will be discussed further in the Discussion chapter. The complete results of the hypothesis testing of PAL model are as follows.

Table 36.
Results of Hypothesis Testing of PALI Model

Hypothesis	Relationship	Std.Est	p value	Result
H1	Course - PAL	-0.34	0.011	Supported
H2	Course - Training	0.55	0.000	Supported
H3	Training – PAL	0.38	0.027	Supported
H4	Course – Experience	0.31	0.001	Supported
H5	Experience – PAL	0.19	0.033	Supported
H6	Support – PAL	0.14	0.199	Not supported
H7	Support – Training	0.43	0.000	Supported
H8	Support - Experience	-0.03	0.715	Not supported

3.8 Mediation/indirect effect analysis in PAL structural model

As in the TAL model, in PAL model, assessment in-service training and assessment experience were also hypothesised to mediate the impact of assessment courses

and assessment support on principal assessment literacy respectively. The indirect impact analysis results of the PAL model are as follows.

Table 37

Indirect Effect Analysis of PAL Model

Indirect effects	Std.Est	p values
Course→Training→PAL	0.016	0.009
Course→Experience→PAL	0.005	0.038
Support→Training→PAL	-0.001	0.481
Support→Experience→PAL	0.015	0.006

Table 37 above shows that the principal's' participation in assessment in-service trainings mediates the impact of the principals' assessment course in higher education on their assessment literacy (p value = 0.009), but it did not mediate the impact of local support on assessment literacy (p value = 0.481). The table also shows that the principals' assessment-related experiences mediates both the impact of assessment course (p value = 0.038) and assessment support (p value = 0.006) on the principal assessment literacy.

Given the results of indirect analysis in AMOS 28.0, three of the four hypothesised indirect effects in PAL structural model were supported as shown in hypothesis testing results table below.

Table 38

Results of Hypothesis Testing of PAL Indirect Effect Analysis

Hypothesis	Relationship	Std.Est	p value	Result
H9	Course - Training - PAL	0.016	0.009	Supported
H10	Course - Experience - PAL	0.005	0.038	Supported
H11	Support – Training - PAL	-0.001	0.481	Not supported
H12	Support – Experience - PAL	0.015	0.006	Supported

Previous analyses (in **Table 35**
Standardised Estimates of PAL Model Structural Paths and Figure 12

Standardised Estimates of PAL SEM Model) have shown that assessment courses have negative direct impact on principal assessment literacy (p value = 0.011). Combining with the results of the indirect analysis shown in **Table 37**
Indirect Effect Analysis of PAL Model, it can be concluded that the effect of the principals' assessment courses on assessment literacy is partially mediated through in-service assessment training as well as through their assessment related experience. Previous analysis also shows that assessment support does not have any direct impact (p value = 0.199) on principal assessment literacy. Thus, the impact of assessment supports the principals received on their assessment literacy is fully mediated through their assessment experience but not through their assessment related in-service trainings. Nevertheless, like in the TAL model, all the indirect impacts in PAL model had very small magnitudes on the principal assessment literacy, which might not have practical significance, thus further elaboration in phase two is necessary.

The summary of the direct and indirect effect of the four predictors in PAL model is in the table 39 below.

Table 39.
Summary of Direct Effects, Indirect Effect, And Mediation of PAL Model

Relationship	Direct effect	Indirect effect	p value	Conclusion
Course→PAL	-0.34		0.011	Direct negative impact
Course→Training→PAL		0.016	0.009	Partial mediation
Cours→Experience→PAL		0.005	0.038	Partial mediation
Support→PAL	0		0.199	No impact
Support→Training→PAL		0	0.481	No mediation
Support→Experience→PAL		0.015	0.006	Full mediation

All the results of the direct and indirect effects in this quantitative phase were elaborated in more detail in the second phase of the study.

4. Summary of phase one findings

Phase one of this study addressed research questions one and two:

RQ1: *What is the level of Indonesian principals' and social studies teachers' assessment knowledge and skills*?

RQ2: *How do assessment-related experiences, assessment courses in higher education, in-service assessment professional development programmes, and local support contribute to Indonesian principals' and social studies teachers' assessment knowledge and skills*?

The analysis results revealed that the levels of the assessment knowledge and skills of the principals and teachers were 54 and 59 respectively in a 100-point scale, with variations across assessment competency dimensions in both groups. Principals demonstrated particular strengths in assessment purposes and ethics and fairness dimensions, both scoring 64, while teachers showed strongest competency in assessment construction and administration and assessment ethics and fairness dimensions, averaging 70 in both areas. However, principals showed their weakest performance in the assessment support for teacher dimensions (scored 48), while teachers struggle most in analysing and interpreting assessments results (scored 35).

Furthermore, in both groups, the same three factors influenced their assessment competencies: in-service training, assessment experience, and assessment-related coursework during higher education, all within moderate range. While trainings and experiences positively impacted both groups' assessment competencies, assessment coursework had a negative effect. This finding is explored further in phase two. Another factor, support from colleagues in their respective environments, showed no direct impact on their assessment knowledge and skills.

Statistical tests revealed that among principals:

- In-service trainings mediated the impact of assessment coursework on their assessment literacy.
- Assessment experience mediated the impact of both assessment coursework and local support on assessment literacy.

While for teachers:

- In-service trainings mediated the impact of both assessment coursework and local support on assessment literacy.
- Assessment experience only mediated the impact of local support.

It should be noted that all the statistical results concerning the mediating effects of trainings and experience variables in both models were very low and may have no practical implications. However, these findings warrant further investigation in phase two.

The following chapter presents the findings of the second (qualitative) phase of the study.

Chapter five: Qualitative findings

1. Introduction

In the qualitative (second) phase, six school principals and eight social studies teachers (total of fourteen participants, see Chapter Three), who took part in the online survey, voluntarily participated in the interview sessions, either face-to-face or online. From the fourteen interview participants, two principal-teacher pairs from two different schools were further explored to investigate the relationship between principals' assessment literacy and leadership and teachers' assessment literacy. The semi-structured individual interviews and paired interviews were conducted to address research questions three and four:

RQ-3: *How do school principals and social studies teachers navigate the enactment of mandated assessment reform against the challenges within their unique local context*?

RQ-4: *How do the Indonesian principals' assessment literacy and leadership in assessment affect social studies teachers' assessment literacy*?

The results of these interviews were used to enrich and provide deeper insight on the findings of phase one (quantitative) of the study through the participants' experiences and practices. Thus, the focus is to reveal the experiences and practices of school principals and social studies teachers in enacting the Emancipatory Curriculum of the *Emancipated Learning* policy especially concerning the assessment reform in their local contexts.

In line with theory of practice (Bourdieu, 1977) and multiplicity of enactment (Buxton et al., 2015), the experience of assessment reform enactment was seen through the interplay between the principals' and teachers' agency, underpinned by their assessment literacy, alongside the demands of the new assessment policy, and the unique characteristics of their school context. Hence, in unpacking the educators' experiences, this study considered the diverse school contexts including the students, the teachers and the local education authority, which may bring

challenges or even barriers and obstacles to fulfil the demands of the policy (Volante & Fazio, 2007; Xu & Brown, 2016). The difficult aspects inherent in the assessment reform, as identified by the principals and teachers during the interviews, were also explored.

As the results of the interaction between the educators' agency, the demand of the new policy, and their local contexts, these interviews uncovered the strategies, adjustments, modifications, and adaptations that the principals and social studies teachers applied to fulfil the mandate of the assessment reform, resulting in differential assessment practices. The differential applications of assessment policy may reflect the educators' assessment literacy more explicitly than merely the scores they got from the TALI and PALI instruments.

Building on the idea of assessment literacy as a multifaceted, negotiated, and context-dependent professional competence and social skills of educators, exploring the practices and experiences in applying student assessment of the school principals and teachers can provide a better understanding of the educators' assessment literacy as one of the critical aspects in the enactment of the *Emancipated Learning* policy. A deeper understanding of the principals' and teachers' assessment literacy emerged by exploring how they identify the challenges, barriers, difficult aspects of the reform, as well as the strengths of the new assessment policy, and how they develop and apply contextually appropriate strategies to adapt the policy within their schools.

Apart from providing deeper understanding of the principals' and social studies teachers' assessment literacy by investigating their assessment practices, the qualitative data also helped shed light on the influence of the four predictor variables examined in the first quantitative phase of the study namely assessment-related experiences, assessment-related courses in higher education, in-service assessment trainings, and local support in assessment practices. In addition, the two paired interviews, provided deeper understanding of the connection between the principals' assessment literacy and leadership and the teachers' assessment literacy.

The subsequent sections present the resulting analytical codes, categories, and themes and the details of each theme emerging from the qualitative data. The themes for the principals and teachers are presented in two separate sections, followed by a conclusion section. The last section, provide the findings from the paired interview analysis.

2. The emerging codes and themes

This study conducted three coding cycles involving several coding approaches such as structural coding, versus coding, and dramaturgical coding (Saldana, 2016) (see Chapter Three). Initially, in the first cycle coding, this study generated 23 main codes, 65 second level codes and 12 third level codes from all the fourteen interview transcripts (principals and teachers). Not all main codes have sub codes (see Appendix 10: Codes from first cycle coding process). An example of a main code, the related second and third level codes is as follows.

"*Challenges of enactment (main code) → More complex than previous assessment (2nd level code) → Differentiated learning (3rd level code)*".

In the second cycle coding, all the codes and the associated interview passages from the first cycle were carefully reviewed. Some of the interview transcripts were also revisited entirely to search for more relevant codes and clarify the existing codes. This action led to adding more new codes and removing some existing codes.

This process was followed by clustering the codes, which were based on the pre-determined issues as well as the ones arising during the interviews which were relevant to the focus of the study.

The second cycle coding resulted in eleven clusters of initial themes as follows:

1. Perceived level of the new assessment policy enactment.
2. Challenges of new policy enactment.
3. Potential barriers of new policy enactment.

4. Shared experience of new policy enactment.
5. Strategies for successful enactment.
6. Features of the new policy.
7. Support for successful enactment.
8. Assessment experiences.
9. Assessment courses in Higher Education.
10. Assessment in-service trainings and workshops.
11. Local support in assessment reform enactment.

In the subsequent analysis, guided by the research questions, the researcher then looked for emerging connections and patterns among the clusters of codes (Bryman, 2016; Creswell, 2014) and, as per the results, came up with five broader themes related to the principals' assessment literacy and teachers' assessment literacy in elaborating the findings from the quantitative phase. These themes apply for both groups, the principals and teachers, but were described separately in subsequent two sections in this chapter. The themes are as follows:

1. The features of the new assessment policy (theme 1)
2. Challenges & potential barriers at local contexts (theme 2)
3. Strategies for assessment reform enactment (theme 3)
4. Factors influencing the assessment literacy and the enactment of assessment policy (theme 4)
5. Self-perceived assessment literacy as the result of reform enactment (theme 5)

The following table shows the themes and the related codes.

Table 40

Themes and Codes from the Interviews

THEMES									
The features of the new assessment policy (Theme 1)		Challenges & potential barriers at local contexts (Theme 2)		Strategies for assessment reform enactment (Theme 3)		Factors influencing the assessment literacy and the enactment of assessment policy (Theme 4)		Self-perceived assessment literacy as the result of reform enactment (Theme 5)	
Principals	Teachers	Principals	Teachers	Principals	Teachers	Principals	Teachers	Principals	Teachers
Code 1: Advantages for students Sub-codes: • More just • Enjoy learning more	Code 1: Advantages for students Sub-codes: • More equitable • Enjoy learning more • Higher motivation	Code 1: Teacher readiness	Code 1: Conceptual aspects Sub-codes: • New concept clarity • Student behaviour assessment	Code 1: Supporting teachers Sub-codes: • Facilitating workshop, sharing session & collaboration • Providing more autonomy for teachers	Code 1: Focus on student distinct characteristics	Code 1: Local support Sub-codes: • Support from teachers • Support from principal community • Support from	Code 1: Local support Sub-codes: • Support from school colleagues • Support from principal • Support from local	Code 1: Improvement of assessment literacy	Code 1: Improvement of assessment literacy

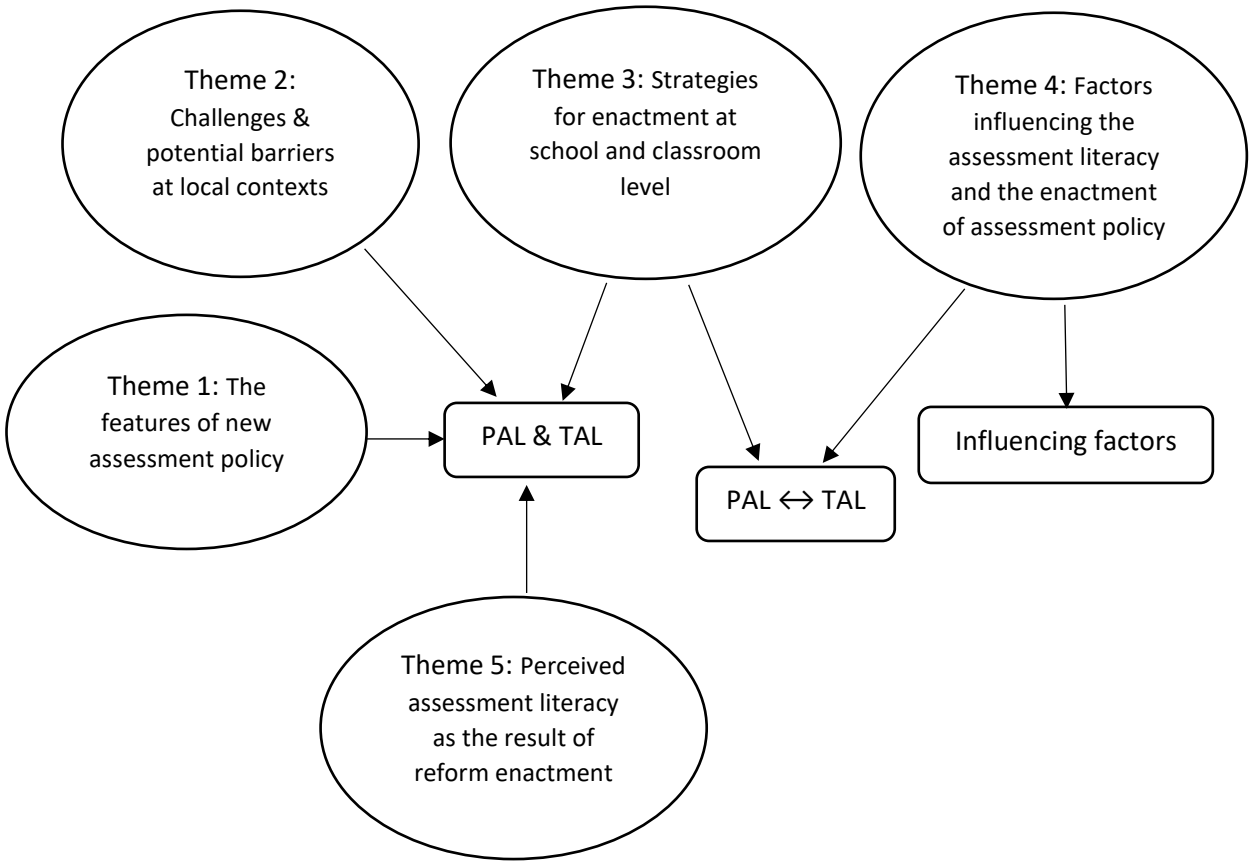
				• Close supervision		superintendent • Support from MoE	teacher community • Support from MoE		
Code 2: Advantages for Teachers Sub-codes: • Shared elements with previous policy • Enjoy the process more	Code 2: Advantages for Teachers Sub-codes: • More flexible • Enjoy the process more • Shared elements with previous policy	Code 2: Practical aspects of the reform Sub-codes: • Challenges from local administrators • More work • Diverse student characteristics	Code 2: Practical aspects Sub-codes: • Challenges from local administrators • More work • Student diversity	Code 2: School assessment policy Sub-codes: • School guidebook • Student final grade • Full assistance approach • Engaging parents	Code 2: Adjustment to local context			Code 2: Assessment leadership	Code 2: Opportunity to help students learn better
Code 3:		Code 3: Doubt with current	Code 3: Teacher readiness	Code 3: Selecting potential teachers at	Code 3:				

Core idea of mandated assessment	assessment policy	the initial stage	Flexible assessment formats
<u>Code 4:</u> Accommodating local culture			<u>Code 4:</u> Peer tutor

As highlighted earlier in previous chapters, based on the recent view of assessment literacy as a multifaceted, differentiated, and contextual abilities of educators, the themes that emerged from the interview data were expected to deepen the researcher's understanding of the principals' and social studies teachers' assessment knowledge and skill examined in the previous phase, the affecting factors, and the impact of principal assessment literacy and leadership towards teacher assessment literacy. The identifications, choices, decisions and actual practices concerning student assessment the educators made may reflect their assessment literacy more clearly.

The following diagram depicts how the themes emerging from the second phase of the study further elaborate the statistical results from the first phase and the relationship of principals' and teachers' assessment literacy.

Figure 13
Relationship of Themes and Statistical Results



In both groups of interview participants, themes one, two, three, and five provide further explanation regarding principals' and teachers' assessment literacy including the mastery level of the assessment concepts as well as its application. For example, Theme 1, *The features of new assessment policy*, presents the principals' knowledge of the current and the previous assessment concepts. This knowledge is directly related to the purpose of assessment as one element of principal assessment knowledge and skills measured in phase one (quantitative). Theme 2, *Challenges and potential barriers in local contexts*, revealed the educators' ability to recognise challenges and obstacles within their contexts that may hinder the assessment reform enactment, such as engaging parents into the school assessment programmes, which is related to communicating assessment results (CAR) dimension. Another example is the third theme, strategies for successful enactment, which present teachers' description of their actual application of constructing and administering assessments (CAA) as one dimension of the teacher assessment knowledge and skills measured in the quantitative phase.

Theme 4 provides deeper explanations concerning the four variables that influence principal and teacher assessment literacy analysed in the first phase, which are assessment-related experience, assessment courses in higher education, in-service assessment trainings, and assessment-related support. In addition, some parts in themes three and four, simultaneously shed light on the relationship between school principal assessment leadership and teacher assessment literacy. Theme 5 relates to the principals' and teachers' accounts of how the enactment of the current assessment policy has changed their assessment literacy and practices.

The emerging themes are described separately between the principals and the social studies teachers. This was done so that the data from the interviews can provide more focused explanations for each of the principals' and teachers' assessment literacy with the affecting factors, respectively. Even though the relationship model of the assessment literacy and the predictors examined in quantitative (first) phase were identical between the principal PAL model and the social studies teachers TAL model, the quantitative results pointed to some differences. These differences thus would require different elaboration from the

respective qualitative data. In addition, some of the dimensions of assessment literacy of principals and teachers measured through PALI and TALI (in quantitative phase) respectively are different. Therefore, the interview data of principals and teachers were described separately to provide deeper explanations of each group's assessment literacy level.

The following two sections present the details of all the themes arising from the second (qualitative) phase of the study for the principal and teacher respectively.

3. Themes concerning principal assessment literacy and the enactment of the new assessment policy

3.1 Theme one: The features of the new assessment policy

The principal interviewees pointed to strengths of the new assessment policy. The ability of the principals to describe the features of the new assessment policy indicates their understanding of the current assessment policy including its distinction from previous policy. The data regarding the principals' views of the features of the new assessment policy were divided into four main codes: Advantages for students, advantages for teachers, core idea of mandated assessment, and accommodating local cultures. The following table summarises the codes and sub-codes for this theme.

Table 41.
Codes of Theme 1 (Principal)

No.	Code	Sub-code	No. of participants (N=6)	Example of quote
1.	Advantages for Students	More just	5/6	*"This method ensures that all students' potential is recognized and nurtured" (P1-2L)*
		Enjoy learning more	1/6	*"The students find greater enjoyment and increased*

				happiness in their learning experiences…" (P2-1H)
2.	Advantages for Teachers	Shared elements with previous policy	3/6	*"The current policy is not totally different from the previous …" (P4-2H)*
		Enjoy the process more	1/6	*"… they enjoy more now as they have the flexibility to conduct assessments …" (P3-1M)*
3.	Core idea of mandated assessment		4/6	*"… they identified which students fall within high, moderate, or low abilities. This served as the foundation for differentiated learning" (P6-2M)*
4.	Accommodating local culture		1/6	"*The presence of school project and its assessment open wider opportunities for teachers to accommodate and introduce our local cultures*" (P3-1M)

a. Advantages for students

The current assessment policy is considered more just for students as recognized by most of the principal participants. It allows for a much wider range of students' abilities to be acknowledged and developed, for example:

> *Overall, this assessment is very good since it allows us to assess students holistically. This method ensures that all students' potential is recognized and nurtured. We acknowledge and accommodate a wider range of student strengths, unlike in the past when we used rankings* (P1-2L).

Further, for students, this new assessment policy enabled them to enjoy learning more because they have the opportunity to reveal their abilities in various ways. However, this benefit for students was only highlighted by one principal:

> *The students find greater enjoyment and increased happiness in their learning experiences because they can earn scores through assessments tailored to their preferences. Unlike traditional methods solely relying on written assessments, our approach allows students to showcase their understanding in diverse ways* (P2-1H).

According to the Ministry document (Centre of Assessment and Learning, 2022), there are more benefits for students expected from the implementation of the reform than those identified by the school leaders, such as increased motivation to learn and improved learning outcomes. This lack of awareness may be because the principals could not identify them or because the students have not experienced the benefits. It should be noted, however, that the principals might not have as much direct interaction with students compared to teachers, which could be the reason for the principals' apparent lack of awareness of the reform advantages for students.

b. Advantages for teachers

The principals highlighted that teachers benefit from some certain shared elements between the current and the previous assessment policy, which allowed the teachers to adapt to the new policy more quickly. This was acknowledged by half of the principal participants, for example "…*they are almost similar, they only use different terms, so teachers need to adjust the formats"* (P2-1H), or:

> *The current policy is not totally different from the previous, ……. previously, we did not do diagnostic assessments and differentiated learning. Now we need to analyse students' initial ability first using diagnostic assessments and then based on that we develop lesson plans* (P4-2H).

Interestingly, only one principal reported that teachers feel this new policy is more enjoyable due to the flexibility they have: "… *in general, they enjoy more now as they have the flexibility to conduct assessments* …" (P3-1M).

c. Core idea of the mandated assessment and accommodating local cultures

The interviews revealed that, except P3-1M, all other principals were aware of and able to describe the general process of assessment mandated by the reform. This is important because it reflects the principals general understanding of the assessment reform that they are required to implement. For example:

> *[What] the teachers are required to do is holistic assessment, in which the teachers need to gather authentic data of students' profile. So, like it or not they should do the initial assessment, because without which they cannot design lessons which meet students' characteristics and learning need* (P1-2L).

The school project programme along with its assessment is one of the important features of the Emancipatory Curriculum by which teachers can accommodate and introduce local cultures, emphasised under the current education policy. Nevertheless, this is only revealed by one principal: "*the presence of school project and its assessment open wider opportunities for teachers to accommodate and introduce our local cultures*" (P3-1M).

While the strengths of the assessment policy may help the schools implement the assessment reform, challenges and barriers on the other hand may prevent them from successfully applying the new assessment concepts as described in the following theme.

3.2 Theme two: Challenges and potential barriers at local contexts

The challenges and potential barriers of the new assessment policy within their local contexts described by the principals reflected their ability to identify and, later, may inform the strategies to deal with difficulties and obstacles they faced during the

implementation. The challenges include teacher readiness, practical aspects of the policy, and opinions against the mandated policy. The awareness of the challenges, barriers or obstacles are important in underpinning the principals' assessment literacy elements identified in the survey, such as assessment support for teachers (P-SUP), assessment communication (P-COM) and assessment local policy (P-POL). The following table presents the codes, the related sub-codes, and the examples of quotes.

Table 42

The Codes of Theme 2 (Principals)

No.	Code	Sub-code	No. of participants (N=6)	Example of quote
1.	Teacher readiness		6/6	*"Many teachers aren't that creative; they just wait to be told what to do" (P3-1M)*
2.	Practical aspects of the reform	Diverse student characteristics	5/6	*"… teachers with strong social mindset, hard for them to treat students differently, even if it's based on skill levels" (P6-2M)*
		More work	3/6	*"We're really short on time for analysing the diagnostic results" (P3-1M)*
		Challenges from local administrators	2/6	*"… however, compared to other cities, our local authority was less supportive" (P1-2L)*
3.	Doubt with current assessment policy		3/6	*"Honestly, I don't quite agree with eliminating national exam or standardised school tests and replaced them with teacher-made summative assessment" (P6-2M)*

a. Teacher readiness

All the school leaders reported that one of the challenges identified was dealing with teachers' readiness for assessment reform. For example, one principal revealed this concern:

> *We are [the teachers] still indoctrinated by past assessment approach where it mainly evaluates student academic achievement and intellectual abilities. But now we also need to focus on assessing student behaviour development* (P1-2L).

Similarly, another stated that:

> *Sorry to say but some of our teachers are … well…they have different levels of capacity in comprehending the new concepts, especially in relation to the paradigm of differentiated learning where teachers need to gather evidence [through assessment] before designing lesson plans* (P4-2H)

One principal raised the complexity issue of the new assessment policy which troubles her teachers: *"…they came to me and said, 'this is too complicated, too hard, we did wrong, …how we suppose to deal with this student characteristics...we are not sure how to do it'* (P5-2M). The teachers' feeling of unpreparedness in implementing the assessment reform might lead the principal to identify appropriate strategy to support the teachers.

b. Practical aspects of the reform

All principals reported that student diverse characteristics posed challenges in the enactment process, such as one principal who was concerned about the teachers' different treatment based on students' levels of competency: *"…. sometimes, students in the lower groups exhibit inferior behaviour since they are in the lower level"* (P5-2M).

Higher work load due to the teacher-student ratio is another practical aspect challenging the enactment of the policy mentioned by three of the school leaders,

as described by P6-2M: "*…to be honest, with teacher-student ratio of 1:160, the differentiated learning poses so many constraints when we must serve the high, medium and low groups of students properly*". These first two practical challenges, student diversity and teacher-student ratio, require principals' ability in creating an effective school assessment policy to address them.

The response from local education authorities became a practical challenge for some principals. P1-2L disappointedly commented that his local education authority was not adequately supporting the programme: "*there are 6 pilot schools in this city, all are private schools, however compared to other cities, our local authority was less supportive*". Another was concerned about local education policy that is not in line with the basic idea of the assessment reform:

> *Our local authority think it will be difficult to determine the regency's education quality if we only apply the new assessment policy, thus they require the schools to keep administering standardised tests as the indicator of the regency education quality* (P3-1M).

The ability of the principals to communicate effectively with their local education authorities might have an effect on the extent they are able to enact the assessment reform.

c. Doubt with the current assessment policy

There are potential barriers that might hinder the success of the policy enactment since it comes from the principals as the direct executors of the policy. Some principals expressed their doubts about the new assessment policy.

> *My feeling…. but this has nothing to do with theory though…. I don't think this [assessment policy] is perfectly appropriate because using multiple methods to achieve one [learning] goal at some points will give us troubles* (P2-1H).

…this is the fact that some principals and school superintendents are not sure about the assessment policy and thus they think that collective standardised tests or even national exams are still necessary (P3-1M).

The challenges that come from their teachers and local education authorities and/or the potential barriers that originated from the principals themselves represent the complexity of the assessment reform enactment. The responses to the challenges and barriers may reveal the principals' assessment literacy more apparent through how they develop their school assessment policy, how they support the teachers, and how they communicate with teachers and local authorities. The following theme regarding the strategies the principals use will reveal them.

3.3 Theme three: Strategies for assessment reform enactment

Various strategies were identified during the interviews with the principals. These include establishing school assessment policy, supporting teachers in the reform implementation, engaging parents, and selecting potential teachers at the initial stage of the enactment.

Table 43.
The Codes of Theme 3 (Principal)

No.	Code	Sub-code	No. of participants (N=6)	Example of quote
1.	Supporting teachers	Facilitating workshop, sharing session & collaboration	6/6	*"Every Monday, we organise learning community session, where we share good practices" (P6-2M)*
		Providing more autonomy for teachers	5/6	*"I'm not here to tell what to do, they can figure things [the assessment] out themselves using available resources" (P1-2L)*
		Close supervision	6/6	*"I encourage teachers to regularly check the PMM [online learning*

				platform] to learn" (P4-2H)
2.	School assessment policy	School guidebook	2/6	"*To oversee the implementation, we agree to create a guidebook" (P2-1H)*
		Student final grade	2/6	*"Final grades incorporate both summative test results and formative assignments" (P5-2M)*
		Full assistance approach	2/6	*"… all ten subject teachers of year seven students follow their students until year nine" (P4-2H)*
		Engaging parents	2/6	*"We invite parents to a session explaining our online assessment system" (P2-2M)*
3.	Selecting potential teachers at the initial stage		3/6	*"I'm counting on the fast learner to guide those who need extra help" (P1-2L)*

a. Supporting teachers

Several strategies of supporting teachers in enacting the reform were reported by all the school leaders. One strategy emphasised by all the participants was facilitating trainings, workshops or other professional learning activities for teachers. This action directly reflects the practical aspect of principal assessment literacy especially regarding the dimensions of assessment-related support for teachers (SUP). They were described those as follows.

> …. *the learning speed of teachers is not equal, so I asked the fast learner teachers to assist their colleagues. I organised best practice sharing sessions in which the teachers can share their good practices to other teachers in teacher room for 15 minutes* (P1-2L).

> *In first year [of the implementation], I conducted in house training basically to equip teachers who have not participated in similar trainings. We follow this by organizing workshops* (P4-2H).

Slightly differently, P5-2M reported that she initiated what she termed as "open class" (i.e., class observation), where some teachers were invited to come in, observe and later give feedback to a colleague on her/his pedagogical and assessment practices in a teaching session, including the lesson plan.

Secondly, except P5-2M, all principals reported supporting teachers by providing more autonomy in their assessment practices. For example, when responding to the question: to what extent do teachers need to follow the school assessment guidebook? A principal answered:

> *… oh, it is flexible, this [the guidebook] only provides general guidelines, they can improvise, in this era we should not confine them to certain resources, so they may adjust their assessment practices according to the characteristics of their subjects* (P2-1H).

Another support for teachers consistently mentioned by all the principals was close supervision. The principals mentioned a number of supervision practices such as observing classroom assessment practices (P1-2L, P3-1M), examining assessment instruments (P1-2L), ensuring teachers' participation in regular sharing sessions (P2-1H), organising regular discussions on MoE online resources (P4-2H), encouraging and maintaining teachers' motivation (P5-2M), providing specific guidance for teachers (P6-2M).

b. School assessment policy

The introduction of a number of school assessment policy actions were reported as the principals' initiative in implementing the reform, such as a guidebook on policy and looping approach. Two of the principal interviewees mentioned that they established a guidebook for teachers which explains the procedures the school should follow in carrying out student assessment. This was expressed by P1-2L and P2-1H:

> *…. we need to establish an operational procedure related to assessment which the teachers can refer to, to avoid haphazard practices. So, generally we refer to guidance from MoE and also to school assessment procedures* (P1-2L).

> *To oversee the implementation, we agree to create a guidebook. We must refer to this guidebook when doing our tasks both for the learning process and the assessment* (P2-1H).

Two principals (P4-2H and P6-2M) reported looping approach strategy, where all subject teachers move vertically with their student cohorts since they commence their study in year seven through subsequent years eight and nine. One of the principals described his policy:

> *In our current system, all ten subject teachers of year seven students follow their students until year nine. This strategy is intended to facilitate the teachers to have better student data administration and thus will enable them to monitor student development* (P4-2H).

This school policy likely facilitates the realisation of one of the main ideas of the Emancipatory Curriculum in *Emancipated Learning* policy which is emphasising the development of students' full potential.

However, there was also school assessment policy that a principal initiated that might violate the basic assessment principle of validity. P6-2M emphasised to teachers of all subjects that they should acknowledge and accommodate students' talent or achievements in non-academic fields into the final score:

> *I told teachers that they need to appreciate students' talent and achievements in extra-curricular programmes by incorporating it into the final score of the subjects. We need to assess students on both aspects, not only academic but also non-academic such as sports et cetera* (P6-2M).

While this might seem like the principal's intention to facilitate student potential and realise holistic assessment, this action raises questions concerning the validity of assessment made.

c. Selecting potential teachers at the initial stage

The careful selection of teachers to be involved aimed to ensure that the initial implementation of the reform would be successful. Three principals reported performing this strategy: P1-2L, P2-1H, and P4-2H. For example, P4-2H stated: *"I deliberately selected teacher colleagues whom I know to be creative and innovative, and I believe would be committed to supporting the policy implementation in the first year for year seven students"*.

The above three strategies developed and applied by the principals, especially to deal with various challenges and obstacles in enacting the assessment reform show the principal assessment literacy more clearly than merely the scores of assessment knowledge and skills. Their practices and experiences disclose the parts of their assessment literacy that were not captured through the scores from the PAL inventory in phase one.

3.4 Theme four: Factors affecting PAL and the enactment of assessment reform

Theme four is devoted to elaborating four variables that were hypothesised to influence principal assessment literacy namely: assessment courses in higher education, assessment-related support, in-service assessment trainings, and assessment-related experience. The following table presents the codes, the sub-codes, and the example of quotes.

Table 44

The Codes of Theme 4 (Principals)

No.	Code	Sub-code	No. of participants (N=6)	Example of quote
1.	Local support	Support from teachers	3/6	*"Go ahead and sign-up sir. If you're selected, we'll support you. They've fully supportive this whole time" (P6-2M)*
		Support from principal community	5/6	*"Regarding assessment, I often share with [principal] colleagues" (P4-2H)*
		Support from superintendent	5/6	*"They're focused on collecting data about Merdeka Curriculum progress" (P5-2M)*
		Support from MoE	6/6	*"… PMM from the MoE, it provides teachers with opportunities to increase their competency in assessment*" (P3-1M)
2.	Assessment courses in HE	Emphasising theory	2/6	"*In universities, topics about assessment emphasised the theoretical aspects*" (P3-1M)
		Focusing on cognitive aspects	2/6	"*…… the assessment [topics] was dominantly about assessing student knowledge, …*" (P2-1H)
		Important as a foundation	1/6	"*The theory is useful, like linking between a student competency to evaluate with the appropriate instruments …" (P3-1M)*
3.	Assessment-related experience	Quickly adapting to changes	1/6	"*Due to my experience, … I can adapt quickly*" (P2-1H)
		Identifying teachers' improvement	2/6	"*Based on my experience, I can see that teachers' understanding has changed*" (P4-2H)

		Enabling to support teachers	1/6	*"I can walk them through the current assessment policy, both formative and summative" (P3-1M)*
4.	In-service training in Assessment	Increasing understanding	2/6	"*The workshops that I have participated, …, were very good and very helpful*" (P6-2M)
		Enabling to support teacher more	1/6	*" I am able to provide feedback to teachers based on my [assessment] knowledge [gained through in-service trainings], …"* (P3-1M).

a. Local support in assessment

The assessment-related local support variable in the quantitative phase was hypothesised to directly affect principal assessment literacy. This variable was deliberately limited to the local support for the principals to implement the assessment reform from three groups: teacher colleagues, local principal community, and school superintendent. The restriction was made because the contextual aspect is one of the important elements in this investigation as it might lead to differences to the enactment of the national policy as well as affect the principal assessment literacy according to the concept of assessment literacy employed in this study.

The assessment-related support was analysed in the quantitative phase and elaborated more in the qualitative phase. However, during the interviews, the principals also mentioned support from the Ministry of Education, especially for the pilot schools. This data was also analysed because it was deemed important to find out more about the support from the government for pilot and non-pilot schools.

Some of the principals acknowledged that there are a group of teachers in their schools who support his/her decision to apply as a pilot school in the *Emancipated Learning* programme:

> *Before I decided to apply as a pilot school, I asked my teachers their opinion and they said they were ready to back me up. So, I said, okay then we need to learn together* (P4-2H).

P2-1H emphasised that teacher capacity was one of the significant aspects of the assessment enactment: "*The supporting factor, clearly it is the teacher competencies because with which we can achieve the policy expectations*". However, the teacher support is more toward the implementation of assessment policy but not directly related to the principal assessment literacy.

There was not any direct effect of local principal community (the Principal Discussion Group/PDG translated from Indonesian language '*Musyawarah Kerja Kepala Sekolah*/MKKS) to the principals' assessment literacy nor the assessment enactment because, first, the meetings were mainly discussing school management aspects as described by P3-1M: "*The MKKS is facilitating, which is important to support [the school], such as organising trainings for school administrative staffs*". Secondly, the principals whose schools have been implementing the new policy were often requested to share their knowledge in the community meetings as reported by P2-1H, P4-2H and P6-2M.

In addition, concerning the assessment reform, the role of the school superintendents was deemed minimum by most principals due to their assessment competency (P4-2H), their doubt of the new assessment policy (P3-1M), or their focus of the assistance which is mostly about administrative and general issues (P2-1H, P5-2M and P6-2M). For example, P4-2H stated that: "*They do their roles, but I must say that they do not help much regarding student assessment, they motivate us to implement the emancipated curriculum though*", and P6-2M said: "*When they come to our school to do coaching, usually it is about general stuffs*".

All the principal interviewees admitted that the support from MoE was significant including the coaching programmes with the facilitators: "…*the facilitators provide constant assistance including the assessment topics*" (P1-2L), the Project Management Officer (PMO) regular meetings or workshops: "…*all the progress or*

the changes in the programmes, including when we face difficulties, we discussed it in PMO meetings" (P4-2H), and the digital education platform (PMM) sponsored by the government: "*The teaching platform, the PMM from the MoE, it provides teachers with opportunities to increase their competency in assessment*" (P3-1M).

P3-1M, who was from a non-pilot school, only mentioned about the digital platform (*Platform Merdeka Mengajar*/PMM in Indonesian) from the government that support his teachers, since his school did not accept other support as pilot schools did.

b. Assessment courses in higher education

Regarding the assessment courses in higher education, some of the principals recalled that the courses were more theoretically driven and lacked real world application as described by P2-1H: "*In my Diploma programme, I didn't really get it how to apply this [assessment for learning] since it was mainly theoretical*", and P3-1M: "*In universities, topics about assessment emphasised the theoretical aspects*".

P2-1H further explained that the focus of the assessment topics taught in university was related to measurement of student's knowledge but put less attention to the evaluation of student behavioural aspects and school climate which are given equal weight with the knowledge aspects in the current assessment policy: "*…… the assessment [topics] was dominantly about assessing student knowledge, while now there are other competencies that are equally important, student character and school environment*" (P2-1H). P4-2H also mentioned: "*it was about test item analysis. But it was not directed toward improving teaching and learning*".

However, P3-1M acknowledged that the assessment theories taught in university are important as the foundation of the application of student assessment: "*the theory is useful, like linking between a student competency to evaluate with the appropriate instruments, especially in evaluating student's higher order thinking skills*".

The qualitative data reveals that despite laying some foundation for student assessment practices, the assessment courses in higher education did not adequately cover important elements within the current assessment concepts.

c. Assessment-related experiences

Some principals reported that assessment-related experiences improve assessment literacy in various ways. For example, P2-1H said that it enables her to quickly adapt the change of assessment policy: "*Due to my experience, I think my assessment concept mastery is getting better, so when the curriculum changed, I can adapt quickly*".

Particularly, the experience affects the principal assessment literacy related to the dimension of Supporting Teachers in Student Assessment (SUP). For instance, P3-1M confidently commented that his experience made him more able to give suggestions to teachers: "*I can give directions to teachers as expected by the new assessment policy including formative and summative assessments*". With his experience, P4-2H could identify teachers improvement concerning student assessment: "*Based on my experience, I can see that teachers' understanding have changed*".

These findings documented evidence that the principal experience in assessment increases their assessment knowledge and skills and their assessment leadership practices.

d. In-service trainings in assessment

The principals' participation in various in-service professional learning in assessment, including those before the current policy change, have contributed to the increase of the principal assessment literacy, as stated by P4-2H: "*Participating in trainings, including those provided in PMM, improved my understanding and I can share it to my teachers*", and P6-2M: "*The workshops that I have participated, including the ones in previous curriculum, were very good and very helpful*".

Additionally, it also increases confidence in providing feedback and suggestion on teachers' assessment practices:*" I am able to provide feedback to teachers based on my [assessment] knowledge [gained through in-service trainings], which increased the creativity in their assessment practices"* (P3-1M).

As with the experience, the findings reveal that in-service trainings also enable the principals to play their role as assessment leadership better, besides improving their own assessment literacy.

3.5 Theme five: Self-perceived assessment literacy

The last theme that emerged from the principal interviews is self-perceived assessment literacy. There are two clusters of code under this theme: improvement of assessment literacy and assessment leadership.

Table 45
The Codes of Theme 5 (Principals)

No.	Code	No. of participants (N=6)	Example of quote
1.	Improvement of assessment literacy	4/6	"*I think I learned a lot, and now I understand better" (P5-2M)*
2.	Assessment leadership	3/6	*"As a leader, teachers ask me when they have problems, …. So, it motivates me to keep increasing my assessment literacy"* (P4-2H)

The first code concerns the principals' statements that the current assessment reform enactment has enhanced their assessment literacy. These were reported by those from pilot and non-pilot schools. For instance, P3-1M from a non-pilot school stated that the monitoring system of the reform has led the principal and teacher assessment literacy to enhance: *"… I believe so, I think it is the encouragement …. the monitoring system, from superintendent [representing the local authority] to school leader level, by which the learning process was guided"* (P3-1M).

Unsurprisingly, the principals of the pilot schools admitted that the coaching and mentoring system from the Ministry of Education during the piloting period have increased their assessment literacy, as P5-2M stated: "*I think I learned a lot, and now I understand better, that it* [the assessment reform] *is actually not as difficult as what other people thought, I think the real reason is because they just don't want to step out from their comfort zone".*

The interview data also revealed the principals' awareness of the necessity to have assessment literacy as part of their role as school leaders. They thought it is their responsibility to be assessment literate so they can assist their teachers in carrying out student assessment tasks. These principals expressed those thoughts:

> *As a leader, teachers ask me when they have problems, they say it is the responsibility of a leader to provide solution. So, it motivates me to keep increasing my assessment literacy* (P4-2H).
>
> *Essentially, I need to take an initiative to learn the new regulation first because not all teachers are willing to take time to learn that after I distributed it. So, it is my duty to summarize it and then share it with them* (P6-2M).

This theme is important in demonstrating that the principals developed an enhanced assessment literacy, allowing them to subjectively evaluate their assessment competencies and relate them to their daily responsibilities. Their accounts complemented the researcher's interpretation of their experiences regarding the assessment reform enactment.

4. Themes concerning social studies teacher assessment literacy and the enactment of the new assessment policy

4.1 Theme one: The features of the new assessment policy

The findings from social studies teachers show that there were three features of the new assessment policy of interest to them, which are the advantages for students,

the advantages for teachers, and the shared elements with previous assessments policy.

Table 46.
The Codes of Theme 1 (Teacher)

No.	Code	Sub-code	No. of participants (N=8)	Example of quote
1.	Advantages for students	More equitable	7/8	*"[In the past] we don't differentiate [the assessment], everyone gets the same tests" (T3-2H)*
		Enjoy learning more	4/8	*"I noticed students tend to perform better with assessments of the current curriculum" (T6-2H)*
		Higher motivation	1/8	"*….it is the various assessment methods and techniques that increase student motivation" (T6-2H)*
2.	Advantages for teachers	More flexible	6/8	*"The assessment in Emancipatory Curriculum gives us more flexibility and focus to students" (T2-2H)*
		Enjoy the process more	7/8	*"Although the administrative work is more demanding, I find the practical implementation is rewarding" (T8-1L)*
		Shared element with previous policy	6/8	*"In the implementation, I don't find it too difficult since I have had similar experience" (T1-1M)*

a. Advantages for students

Most teachers acknowledged that the current assessment policy benefits students in several ways. First, it is fairer for students:

> *In the past, we treat students without considering their different capacity, whether it is low or high. In emancipated curriculum, we focus*

more on student unique abilities which we capture through diagnostic assessments (T2-2H).

We are asked to create assessment that more relevant to students, so we need to understand them first (T5-2M).

Secondly, the assessment approach makes students enjoy learning more, as reported by two teachers:

In my observations, when I applied this assessment, students enjoyed learning more since they are not forced to achieve a pre-determined level as were in the past (T2-2H).

I see students enjoy more, even though it is more complicated for the teacher, but they enjoy more… (T4-2M).

In addition, it also increases student motivation to learn, leading to better learning outcomes as described by T6-2H: "….*it is the various assessment methods and techniques that increase student motivation, which in the end improve student learning outcomes*".

The findings highlighted that teachers from different social and cultural contexts, across provinces and rural/urban areas, were able to grasp the basic idea of assessment for learning and trying to apply it in their classrooms.

b. Advantages for teachers

The assessment reform also benefits teachers in a number of ways such as it is more flexible and more interesting. It is more flexible because teachers have more space to innovate as expressed by six out of eight of the participants, for example: "*It allows us to innovate more, it is more flexible because we do not have to strictly follow what is written in the lesson plan*" (T1-1M). T5-2M described the flexibility as follows: "…*oh I might try oral test, next time I can try using pictures, or I can ask the student to make presentations, so it is more open for me*".

For the teachers, the new policy is more interesting despite it is more demanding. This perception was conveyed by most teachers, such as T3-2H: "*Even though it means more work, but it is more enjoyable for me…more fun*", and T4-2M: "*Generally, I am happy because although it is more complicated and we must learn, but it is not monotonous because teachers have more freedom*".

These findings confirm those from the principal reports that the reform offer more flexibility and was more engaging for teachers, although it is also more challenging. These challenges are elaborated in the next section.

Another feature of the assessment policy that benefit teachers is that it has some similarities with the previous one, which for some teachers helped them in the transition process. The similarities were recognised by most of the teacher participants, for example: "*Basically, in my opinion they are similar*" (T6-2H), or "*For me, I just need to make some adjustment because I have similar experience from previous curriculum*" (T4-2M), and "*My school has just received training on the emancipatory curriculum, and it is not far different from our previous curriculum*" (T8-1L).

The teacher identification of the key features of the assessment reform provides additional evidence of their assessment literacy beyond the scores of their assessment knowledge and skills.

4.2 Theme two: Challenges and potential barriers at local contexts

During the interviews, the teachers were asked to identify and describe the challenges and potential barriers they experienced during the implementation of the assessment reform. Generally, the teachers' accounts regarding the challenges can be divided into two codes namely conceptual aspects and practical aspects.

Table 47
The Codes of Theme 2 (Teachers)

No.	Code	Sub-code	No. of participants (N=8)	Example of quote
1.	Conceptual aspects	New concept clarity	5/8	*"Honestly, I am still getting familiar with diagnostic assessments" (T8-1L)*
		Student behaviour assessment	6/8	*"I addressed behavioural issues with speaking directly with students involved" (T5-2M)*
2.	Practical aspects	More work	5/8	"*The cognitive-diagnostics assessment was hard because we need more time to analyse students' answers…*" (T3-2H).
		Student diversity	5/8	*"We need to prepare multiple [assessment] instruments for different capability levels" (T4-2M)*
		Challenges from local administrators	3/8	*"We remind them [local authority] about the new assessment policy, but they insist…" (T3-2H)*
3.	Teacher readiness		4/8	*"Some colleagues only enthusiastic at the beginning, after a while they go back to the previous methods" (T7-1M)*

a. Conceptual aspect

Most teachers reported that clarity of the assessment concept became one of the challenges. For instance, T8-1L expressed his confusion: "*I was not sure whether what I have done was as expected or not, there are many ambiguities*". In addition, T6-2H specifically mentioned diagnostic assessment as the challenge for her and her colleagues: "*It was not only us who have an issue with diagnostic assessment, I believe other schools have the same problem. We did not do it in the first year because we were not sure yet*".

Further analysis of the data shows that another challenge related to conceptual aspects identified during the interview was combining final score from different

types of competencies. In the previous curriculum, assessment of student's knowledge, skills, and attitude/behaviour were separated and had devoted spaces within student report card, in which the results of behaviour assessment were reported in the form of notes from the teachers. However, current assessment policy requires teachers to produce a single score representing the three domains. This shift poses challenges for teachers. T7-1M expressed this when being asked about the challenges of the new policy: "*In this new regulation, we need to decide one final score do we …? while in the past we did separately between knowledge and behaviour*".

An important issue emerged from the interview data was student attitude/behaviour assessment. This issue has arisen, to some extent, as the result of the policy of combining scores from three domains. The teachers reported approaching attitude/behaviour assessment differently. Some teachers tended to incorporate their observations on students' behaviour into subject scores directly even though they were aware that this practice may be biased, such as T4-2M: "*I did it* [behaviour assessment] *through observation, though it may be biased, but students who lack of engagement during discussions will get lower scores compared to those showing enthusiasm and more active*".

Other teachers chose to first try to improve students' (negative) attitudes/behaviour and only when it was not showing a positive improvement for a longer period of time than it would affect the final score. For example, T6-2H stated: "*When the negative behaviour was persistent for long period, even after I gave some advice, then it would have consequences on the final score*".

On the other end, a few teachers, such as T7-1M, preferred to exclude students' behaviour assessment from the final score, even when a student's undesired behaviour remained unchanged after receiving the teacher's advice. In such cases, she chose to engage the parents to discuss the issue. Conversely, when students exhibited outstanding positive behaviour, the teacher would reward them with some gifts.

b. Practical aspects

A comparison of the findings between the practical and conceptual aspects shows that the challenges related to practical aspects were echoed more than those of the conceptual aspects. Higher workload and complexity of the new assessment policy was among the main challenges for teachers due to the various issues such as time and administrative demand, for examples:

"*The cognitive-diagnostics assessment was hard because we need more time to analyse students' answers to develop lesson plans*" (T3-2H).

"*I have to say that the administrative work is more demanding, we need to prepare more documents concerning assessment*" (T5-2M).

"*In my opinion time is the issue since I have to split my time to do not only teaching but also other additional tasks*" (T6-2H).

Another practical aspect mentioned by many teachers, which require more work was dealing with diverse students' characteristics. This was expressed by, for example, T3-2H: "*In doing formative assessment I use several techniques including written tests, observation, performance tests, etc, which is not easy, because I need to consider students' diverse characteristics*". Dealing with slow learning students, in particular, was also a practical issue, as described by T2-2H: "*The challenge is when I have to deal with slow learners especially when the topics require deeper analysis. So, I need to think about different questions to ask*".

The interviews with teachers revealed that practical challenges also came from local education authorities. A teacher (T3-2H) reported her local education office continued to administer local standardised test, despite some teachers' reminders about the new assessment policy. Two teachers from the eastern part of Indonesia expressed frustration over the lack of support from their local education authorities when applying the new regulations. For example: "… *it's been more than two years, but there is hardly any support from the local authority*" (T6-2H).

These teachers' reports are consistent with the principals' accounts that local authorities posed practical challenges through policy mismatches and insufficient support.

c. Teacher readiness

The interviews with half of the teachers aligned with the findings from the principal group regarding teacher readiness. For example, a teacher expressed her concern regarding teacher's resistance to change as one of the challenges related to practical aspects of the new policy:

> *I can say that those who oppose are bigger. They say what they have done so far was fine, why they need to make the assessment more complicated* (T6-2H).

Teacher awareness of the local challenges they encounter might lead to differential strategies. The following section presents findings about the strategies the teachers did in enacting the assessment reform within their individual contexts.

4.3 Theme three: Strategies for assessment reform enactment

While the principal participants' account about the strategies on the enactment of assessment reform was related to school level, the social studies teachers reported their experiences specifically relating to the teaching of the subject and classroom assessment strategies. The social studies teacher participants reported multiple classroom assessment strategies during the interviews which included, among others, a focus on students' distinct characteristics; adjustment and adaptation to local context; flexible assessment formats; peer tutoring in student groups.

Table 48

The Codes of Theme 3 (Teachers)

No.	Code	No. of participants (N=8)	Example of quote
1.	Focus on student distinct characteristics	7/8	*"In performance assessment, I divided the students according to their learning style, which ones are visual person, which ones are audio learner" (T7-1M).*
2.	Adjustment to local context	6/8	*"For example, for local culture topics, I adjusted the formative assessments according to their interests and what they're good at" (T6-2H)*
3.	Flexible assessment formats	7/8	"*In the end of semester summative assessment, I use several formats, not only multiple-choice types, …*" (T5-2M)
4.	Peer tutoring	3/8	*"I appointed high-ability students as team leaders and help organise their classmates" (T4-2M)*

a. Focus on student distinct characteristics.

This category was the main strategy expressed by most of the respondents, which reflected their awareness and attention to the core focus of the policy: student wellbeing and differentiated learning approach. For instance, T2-2H, in response to her student differences, gave them options to respond to an assessment task to their preferences: "*I gave them opportunity to answer the questions the way they prefer, either orally, using Canva application* [graphic design platform to create poster/presentation]*, or in writing*". Similarly, T4-2M said: "……. *they can describe their answer through a concept map, slide presentation, performing drama, or creating a short video*". While T7-1M and T8-1L clearly linked their assessment designs to student's learning style:

"*In performance assessment, I divided the students according to their learning style, which ones are visual person, which ones are audio learner*" (T7-1M).

"*I realised that some of them are visual learners and others are audio, so I varied the assessment*". (T8-1L).

Another teacher, T3-2H, though, also considered student characteristics (i.e., ability level), applied stricter approach. For high achieving student group, she allowed them to choose the types of assessment. However, the students needed her approval of those assessments first before they proceeded. On the other hand, for students in slower learning groups, she decided what kind of assessment they would get.

b. Adjustment and adaptation to local context

Most teachers also made some adjustments when assessing their students' learning progress according to student local characteristics or cultures. One of them adapted test items appropriate to the level of her students:

> "*In assessment, I adapt it according to my students' profile. For example, in writing test items, I realise that my students' level is behind those in Java Island, so I make them simpler*" (T1-1M).

Another teacher considered local cultures when administering assessments:

> "*When it comes to local cultures, I can't do that. For instance, in topics about food tradition, if we use standards of western regions like eating etiquettes or tradition, students from eastern regions are different*" (T6-2H).

Both teachers are from the eastern part of Indonesia. These findings confirm previous studies (Azzizah, 2015; Dharmawan & Suryadarma, 2021) concerning the disparity in education quality between the eastern and the western part, which might require different approaches and techniques in instructional and assessment practices.

c. Flexible assessment formats.

As the findings from teachers focus on students' distinct characteristics and their awareness of student capacity and local cultures, most participants employed various methods of assessment whether it was for diagnostics, formative, or summative assessments. For example, T6-2H in conducting diagnostic assessment, she allowed students to answer questions orally when it suited them better: "……*sometimes I noticed that some students find it difficult to provide the answers in writing, so I welcomed oral answers*" (T6-2H).

T8-1L used different techniques in formative assessment: "*In formative assessment, I do not only use paper and pencil test, sometimes I do quizzes using the Kahoot application, other times I administered debate such as when discussing sociology topics*". Although the format of summative assessment were mostly paper and pencil tests, some teachers tried to use multiple techniques, as described by T5-2M: "*In the end of semester summative assessment, I use several formats, not only multiple-choice types, but 3 or 4 other types*".

d. Peer tutoring in student groups

Another strategy mentioned by some of the participants was distributing high ability students to all student groups to provide peer tutoring or to motivate other students. The following teachers describe the strategy as follows:

> *Student participation in group discussion is low, they mainly only read what is in the textbook, no further discussion…… so, I often use the bright kids as the motivator. They ask questions or help their friends to answer by giving some clues* (T1-1M).
>
> *Based on the results of diagnostics assessments, I appointed high achieving students as the leader of the groups* (T7-1M).

These findings reveal that the teachers applied multiple strategies in response to the unique characteristics of their classrooms and local cultures. Additionally, they also indicated that the assessment reform provided more space for teachers to exert

their agency to tailor their assessment practices to be more contextually appropriate.

4.4 Theme four: Factors affecting TAL and the enactment of assessment reform

This theme provides explanation to the findings of four variables which were hypothesised to affect social studies teacher assessment literacy as well as the enactment of assessment policy. The four variables, as with the principals, are assessment courses in higher education, assessment-related local support, in-service assessment trainings, and assessment-related experience.

Table 49.
The Codes of Theme 4 (Teacher)

No.	Code	Sub-code	No. of participants (N=8)	Example of quote
1.	Assessment-related local support	Support from school colleagues	3/8	*"The teachers in my school gather every month. We discussed, not only student assessment, but also other topics …" (T3-2H)*
		Support from principal	6/8	"*Teachers are encouraged [by the principal] to attend workshops regarding assessment*" (T4-2M)
		Support from local teacher community	6/8	"*In our MGMP, classroom assessment is a rare topic to discuss*" (T6-2H).
		Support from MoE	2/8	"*This year I intensively participated in mentoring programmes for pilot schools, which increase my understanding of diagnostic assessment …" (T6-2H)*
2.	Assessment courses in HE	Basic principles	3/8	"*The evaluation courses in my undergraduate study, I*

			think, was not significant. It does not match" (T1-7L)
3.	In-service trainings	4/8	*"Yes, it was helpful, through the learning community meetings, some colleagues shared their best practices" (T3-2H)*
4.	Assessment experiences	3/8	"*With my past experience, I found it not too difficult to do this assessment, although there are some differences between current assessment and the past*" (T1-1M)

a. Assessment-related local support

The local support concerning student assessment includes support from colleague teachers, the social studies teacher community in the city/town, and school principal. Supports from these three groups were investigated in the quantitative questionnaire in phase one. The researcher also included support from the Ministry of Education since the respondents mentioned that besides those groups.

The support from colleague teachers varied either in terms of the degree of support or how it occurs. Some teachers felt that their colleagues were supportive such as the experience of T4-2M: "*Yes, we support each other. Sometimes, we discussed about particular assessment tasks….is it appropriate? Is this test item too difficult* [describing the conversation] …?". However, other teachers felt that their colleagues did not help at all like T6-2H reporting: "*In my school, the teachers, especially the senior ones, they tend to stay in their comfort zone. That is why I mostly learn by myself*". T3-2H explained that her school organised a monthly meeting for all teachers discussing various topics: "*The teachers in my school gather every month, we discussed, not only student assessment, but also other topics such as teaching modules and media*".

Regarding support from principals, most of the interviewees recognised that they received support from their school leaders. The types of support included facilitating trainings/workshops, monitoring or motivation. T3-2H informed that her principal was very attentive to teachers: "*He was very active, from checking our teaching modules to organising regular reflection and sharing sessions every Friday*". T5-2M enthusiastically described her principal as highly motivating: "*I am glad that my principal was pro-active and always wanted the best. His motivation gave us new energy*".

Many of the teacher interviewees reported that their local social studies teacher community (in Indonesian *Musyawarah Guru Mata Pelajaran/MGMP*) organised regular meetings in their cities/towns. The meetings discussed various teaching aspects including student assessment. However, topics about assessment did not get enough attention as acknowledged by the following teachers:

"…. *sharing about student assessment is rarely done*" (T3-2H).

"*In our MGMP, classroom assessment is a rare topic to discuss*" (T6-2H).

In relation with government support, the researcher found that some teachers talked about the support from the Ministry of Education more enthusiastically than those from the three groups (colleagues, leaders, and teacher community). This was especially coming from teachers from the pilot schools who received various support from the MoE. T6-2H highlighted the mentoring programme that helped her with the assessment implementation: "*This year I intensively participated in mentoring programmes for pilot schools, which increase my understanding of diagnostic assessment that I need to apply in the classroom*". T4-2M emphasised the teaching online platform developed by MoE (the PMM) that facilitates teacher learning more effectively: "*All teachers can access the PMM and independently participate in moderated online learning, including assessment topics, and get certificates*".

The interview results were in line with the quantitative findings that local support did not directly affect teachers' assessment literacy. However, the interviews findings also suggested that local support (especially from principals and teacher colleagues) encouraged teachers' participation in various professional learning activities in assessment, which in turn could improve teachers' assessment literacy.

b. Assessment courses in higher education

The assessment courses here mean any courses related to educational assessment, evaluation, and measurement. Those courses may be taken during teachers' study in undergraduate or post-graduate studies. Indonesian teachers may complete assessment coursework pre-service through two pathways: as part of an undergraduate pre-service teacher education programme or, for those without an undergraduate teaching degree, a compulsory 6-month teaching certification. The teachers reported that the topics provided them with basic principles of assessment and assessment theory. They said:

> *I took learning evaluation and learning models, which also covered student assessment. But it was only the basic principles* (T4-2M).
>
> *I would say that there are some correlations between the courses I took in my undergraduate and current curriculum. It gives me foundation on basic assessment principles and some assessment techniques* (T6-2H).

However, T7-1M thought that the courses were not significant in relation to the current assessment policy, as she said: "*The evaluation courses in my undergraduate study, I think, was not significant. It does not match*".

The emphasis on the theoretical understanding of the assessment courses in university, as reported by the three teachers interviewed, explains the quantitative finding as to why the assessment courses variable had a negative impact on teachers' assessment literacy. The teacher survey respondents in phase one might think that what they learned in higher education was not what they need within the current assessment policy.

c. In-service assessment trainings

Many teachers acknowledged that the in-service trainings concerning student/classroom assessment significantly helped them in applying the new assessment policy. In general, the trainings or workshops improved teacher's assessment literacy skills as described by T4-2M: "*It is very significant, because through which we can learn how to conduct assessments better*". Other teachers highlighted the impact of in-service trainings on specific aspects of assessment such as test item development: "*Through the trainings and workshops, we get new understanding and new experiences in writing test items*" (T6-2H), and development and assessment of student literacy and numeracy skills:

> *I participated in assessment training in relation to student literacy and numeracy skills. This makes the assessment more meaningful, not only assessing student memorisation skills, but also their critical thinking skills* (T7-1M).

These findings confirm the impact of assessment in-service trainings on teacher assessment literacy shown in phase one.

d. Assessment-related experiences

In the quantitative phase, teachers' assessment-related experience's impact on teacher assessment literacy was examined. The experiences refer to: teaching experiences, which obviously include student assessment practices; experiences as educational assessment instructors either at school-level, regional level or national level; and experiences as test item/assessment developer team member, at any level.

Most of the teachers reported that their past experience really helpde them in applying the current assessment policy. T1-1M confirmed that her experience enables her to easily adapt to the new assessment approach: "*With my past experience, I found it not too difficult to do this assessment, although there are some differences between current assessments and the past*". T4-2M also emphasised the importance of her experience as teacher and test developer: "*My experience both*

as a teacher and as a test item writer is really helpful. Even though there are changes, it is really helpful".

The findings from the interviews (theme four) provide further information concerning the impact of the four predictors on teacher assessment literacy, examined in the phase one of the study. Through the rich narratives regarding local support, to some extent this theme also indicates the relationship between principals' and teachers' assessment literacy, particularly regarding the teachers' accounts about the support they received from their school leaders. In particular, the findings from the paired interview presented later will demonstrate the relationship between principals' assessment literacy and leadership and teachers' assessment literacy.

4.5 Theme five: Self-perceived Assessment Literacy

The last theme from interviews with teachers captured teachers' perceptions regarding their assessment literacy, especially since the enactment of the new assessment policy. There are two codes related to this theme: improved assessment literacy and the opportunity to help students learn better.

Table 50.
The Codes of Theme 5 (Teacher)

No.	Code	No. of participants (N=8)	Example of quote
1.	Improvement of assessment literacy	7/8	"*It [the policy] requires us to be better at creating assessment that actually fits our students' need*" (T5-2M)
2.	Opportunity to help students learn better	7/8	*"… from students' baseline knowledge [gained through diagnostic assessment], we design lesson plans tailored to students' need"* (T3-2H)

In general, the majority of the teachers both from the pilot and non-pilot schools reported that the assessment reform enactment has increased their assessment literacy:

"*I think this whole process make me have better understanding about assessment, increase my creativity and innovation in student learning as well as in assessment practices*" (T1-1M).

"*As a teacher, my assessment literacy increases, since I need to upgrade my skill*" (T8-1L).

However, some teachers also conveyed that some local issues might hinder them from achieving better assessment literacy, as revealed by a teacher: "

> *Yes, it increases, but not significantly. From 1 to 10, I would say it is 6. The reason is because during two years of implementation, I need to take care of many other matters*" (T6-2H)

Some teachers perceived what they are doing now as an opportunity to help students learn better:

"I think it [new assessment policy] *is not that hard, it just requires us to consider students more when we design assessments, and we need to understand them more" (T5-2M).*

"This differentiated approach requires us to learn more to serve students' learning need better" (T7-1M).

The findings on teachers' self-perceived assessment literacy provides further evidence concerning the relationship between their assessment literacy and the enactment of assessment reform. On one hand, assessment literacy influences their practices, on the other hand, the reform implementation improves teacher assessment literacy.

The next section provides the findings from the analysis of two paired interviews in phase two of the study.

5. Paired interview analysis results

A paired interview method was employed specifically to address research question four: How do the Indonesian principal assessment literacy and leadership in assessment affect social studies teacher assessment literacy? This involved exploring principals' assessment-related decision making and actions and the potential influence on teachers' assessment competence and practices.

The pairs in this study were identified when principal-teacher from the same schools participated voluntarily in interview sessions, after previously completing an online survey. Two pairs emerged in this study, labelled as School A and School B respectively, with their respective principals and teachers referred to as principal A/B and teacher A/B. So, in presenting the findings from the paired interview, the previous labelling system for the principals and teachers from the interview is no longer applied.

The relationship between the principal's and teacher's assessment literacy in each pair were explored through identifying and analysing relevant codes alongside associated interview excerpts expressed by either the principal or teacher from the same school. These codes from both pairs have been clustered into three themes developed earlier from the previous interview data analysis and presented above. These themes are Theme 2 (challenges & potential barriers at local contexts), Theme 3 (strategies for successful enactment at school and classroom level), and Theme 4 (factors affecting TAL and the enactment of assessment reform).

Before presenting the findings for each pair, the contexts of each of the pairs are provided, followed by the scores of assessment knowledge and skills from phase one.

5.1 Paired interview one: School A

School A is a private school located in an urban area and a provincial capital city in eastern region of Indonesia. The principal A scored 50 in assessment knowledge and

skills, placing him in the low category, while teacher A scored 88, making her one of the two highest-scoring teachers among all teachers participating in the interviews.

During the interview, principal A highlighted teacher readiness as one of the challenges since many teachers at the school were still entrenched by the old assessment paradigm. The strategies implemented by principal A appear to stem partly from this teacher readiness issue. In leading the enactment of assessment reform in his school, principal A reported providing support to teachers, first, through building teacher assessment capacity by facilitating workshops, sharing sessions and teacher collaborations. Second, through enabling autonomy for teachers in their assessment practices. For example, principal A said: "*I organised best practice sharing sessions in which the teachers can share their good practices to other teachers in teacher room for 15 minutes*". The principal's accounts were corroborated by teacher A, who was interviewed separately. Teacher A said: "*I have to admit the principal support was substantial and impactful*".

Another strategy employed by principal A, which highlights the connection between principal and teacher assessment literacy, was selecting highly capable teachers, including teacher A, in the initial phase of enactment. This step was taken to ensure the smooth rollout of the new curriculum and assessment reform. It also demonstrates the principal's awareness of the competencies of his teaching staff. In addition, principal A also reported that he developed a school assessment guide, which provided direction and guidance for teachers in enacting the assessment reform in their classrooms.

Despite principal A's apparent low score on the assessment knowledge and skills, both his and his teacher's accounts revealed that principal A played a vital role in implementing assessment reform in school A through various supports and school-level assessment policies. This underscores his significant contribution to enhancing the assessment literacy of the school's teachers.

5.2 Paired interview two: School B

School B is a public school situated at a sub-urban area in the central part of Indonesia. In contrast to school A, in school B, the principal achieved a higher score than the school's teacher in the assessment knowledge and skills measurement, with scores of 77 and 68, respectively. However, it is important to note that while the principal and teacher at school A both had been involved in the implementation of the assessment reform for two years, in school B, the principal had participated for two years, whereas the teacher had only involved in the reform implementation for less than one year, specifically during the second year of new curriculum and assessment reform.

Teacher readiness also became one of the issues in school B, and for which reason, principal B initiated intensive teacher assessment capacity building efforts through various workshops and sharing sessions for teachers. Interestingly, even though principal B also provided autonomy for teachers, he played an active role in the professional learning activities by serving as the resource person in these activities. These strategies might have positively influenced the assessment competencies of teachers at the school. Arguably, one indicator of this influence was the score (i.e., 68) achieved by teacher B, which was only slightly below the middle category, despite her limited, less than one-year exposure to the new assessment concepts.

Another important initiative by principal B that bridges his assessment literacy to that of the teachers was his proactive approaches to learning the new assessment reform concepts and sharing his knowledge with the teaching staffs. Beyond merely communicating policy updates and new assessment concepts, principal B established a school-level assessment policy based on his interpretation of the national guidelines. Principal B reinforced this local policy by repeatedly emphasising its details to the teachers on multiple occasions. For example, teacher B reported: "*Our principal regularly reminds us that students were allowed to take assessment ahead of schedule if they completed their learning material faster*". This local policy was well understood and articulated by teacher B during her interview, conducted separately.

The initiatives taken by school B principal demonstrated a strong contribution to both the implementation of assessment reform at the school and the enhancement of teachers' assessment literacy.

6. Summary of findings of phase two

This chapter presents the findings of the qualitative (second) phase of the study, encompassing description of the resulting codes, categories, and themes as well as the details of each of the themes. The qualitative findings were used to address the third research question: *How do school principals and social studies teachers navigate the enactment of mandated assessment reform against the challenges within their unique local context?* and fourth research question: *How do the Indonesian principals' assessment literacy and leadership affect social studies teachers' assessment literacy?* Findings relating to the third and fourth research questions also shed light on the previous findings in the quantitative phase.

Five themes were developed in relation to the principal and teacher respectively concerning their assessment literacy and the enactment of the assessment new policy in their schools:

1. The features of the new assessment policy.
2. Challenges & potential barriers at local contexts.
3. Strategies for successful assessment reform enactment.
4. Factors affecting assessment literacy and the enactment of assessment reform.
5. Self-perceived assessment literacy.

These themes display the leaders' and educators' thoughts, ideas, and experience in relation to the enactment of the reform. In particular, these findings show the multiplicities of the enactment of assessment reform as the result of the interplay between the principals' and teachers' agency underpinned by their assessment

literacy, the demand of the new assessment policy, and the unique characteristics of their school context.

The findings from the two paired interviews in this research provide a more detailed glimpse of the relationship between principals' leadership in assessment and teacher assessment literacy. The description of how the principals and teachers interact in relation to student assessment within the framework of the current assessment reform helped reveal deeper insights into each group's assessment literacy.

Based on the conceptual framework of assessment literacy as a multifaceted, negotiated, and context-dependent professional and social competence and practice of educators, the differential assessment policy implementations and the dynamic interaction between principals and teachers in enacting the assessment reform may reflect the assessment literacy of both groups more thoroughly than merely the assessment literacy scores they achieved from the TALI and PALI instruments in phase one. Therefore, in the next chapter, these themes and the paired interview findings are discussed in more detail to explain the principal and social studies teacher assessment literacy more comprehensively.

Chapter six: Discussion

1. Introduction

This chapter discusses the study's findings to provide a comprehensive interpretation of the research outcomes and to address the overarching research question: *What is the current state of Indonesian principals' and social studies teachers' assessment literacy in enacting the assessment reform in diverse school contexts?*

In particular, the focus of the discussion is to explore the current state of Indonesian school principals' and teachers' assessment literacy, as well as the factors influencing their understanding and enactment of the assessment reform. While the quantitative strand of the research focused on measuring the leaders' and educators' levels of assessment knowledge and skills and on examining its complex relationships with four key factors, the qualitative strand provided deeper insights into the practical and contextual dimensions of assessment literacy.

Specifically, the study sought to address the research questions:

(1) *What is the level of Indonesian principals' and social studies teachers' assessment knowledge and skills*?

(2) *How do assessment-related experiences, assessment courses in higher education, in-service assessment professional development programmes, and local support contribute to Indonesian principals' and social studies teachers' assessment knowledge and skills*?

(3) *How do school principals and social studies teachers navigate the enactment of mandated assessment reform against the challenges within their unique local context*?

(4) *How do the Indonesian principals' assessment literacy and leadership in assessment affect social studies teachers' assessment literacy*?

Collectively, the findings related to the four research questions address the overarching research question. By synthesising the quantitative and qualitative findings, the discussion in this chapter seeks to provide a more thorough understanding of school principals' and social studies teachers' assessment literacy, situating them within broader theoretical and empirical research, and practical landscapes in relation to principal and teacher assessment literacy.

2. Principals' and teachers' assessment knowledge and skills levels.

2.1 Principals' assessment knowledge and skills.

As demonstrated in Chapter Four, the present study found that the school principals had an assessment knowledge and skills level of 54. This result was slightly below the average score of the teachers, which will be discussed in the next section. Previous studies have yielded inconsistent findings regarding the relative assessment knowledge and skills of school principals compared to teachers. Similar with the current study, Lee-Perry (2013) reported that high school principals in Montana USA had slightly lower average score (59) of assessment knowledge and skills than teachers (62.6). In contrast, in her doctoral research, Rosas (2014) found that elementary school principals in California had an assessment knowledge and skills level ten percent higher than that of the teachers (66.1 vs 55.1).

It is worth noting that while the two USA-based studies used the same assessment literacy instruments to compare the assessment knowledge and skills of school leaders and teachers (i.e., treating the two as the same), this current study adopted different approaches. This study argues that the assessment literacy required for a principal should encompass dimensions beyond classroom assessment practices, which are the primary focus of teachers. Stiggins and Duke (2008) emphasised that a principal must be assessment literate at classroom-level, programme-level, and institutional-level to effectively fulfil their role as an instructional leader. Therefore, among the five dimensions of the principal assessment knowledge and skills framework used in this study, just three were intended to overlap with those

assessed for teachers, while the other two were specifically related to principal's leadership role in school-level assessment.

The dimension with the lowest average score of the principals was Assessment Support for Teachers (P-SUP). While the other dimensions had average scores of 60 or higher, the principals scored only 48 on average for that dimension. This finding is somewhat surprising because the qualitative data revealed that supporting teachers through facilitating a range of professional learning activities was one of the main strategies the principals performed to ensure the successful enactment of the assessment reform. A closer examination of the five PALI items measuring this dimension shows that one particular item, asking the principals to identify teachers' weaknesses in classroom assessment practices, received a very low score (24). This brought down the overall score of this dimension.

The score of the item explains the contradiction between the quantitative and qualitative findings. Given the focus of principals' assessment responsibilities extends beyond classroom assessment, it is reasonable that their knowledge and skills in this area are lower than those of teachers. In addition, from Bourdieu's theory (1977), this fact stems from the Indonesian principals' habitus. Their long professional experience and leadership training programmes have instilled an embodied disposition that led them to predominantly focus on managerial and administrative functions, while tend to neglect the pedagogical and assessment practices of teachers (Lee and Hallinger, 2012).

However, it remains essential that the principals also possess adequate competency in this aspect. According to Stiggins and Duke (2008) the ability to evaluate teacher's classroom assessment competencies and help teachers learn to assess accurately and use the results productively is one the skills principals should embody to be effective leaders in school assessment practices.

Interestingly, the principals seem aware of the gap and implemented strategies to address their lack of skills in this area. P1-2L and P5-2M, for example, used a peer coaching strategy through classroom observation, dialogue among colleagues and

sharing sessions. The principals' approach which prioritised teacher collaborations aligned with Wiese and Nortvedt (2023) that a collective and shared whole-school practice is effective in the development of teacher assessment competencies. A different approach to address similar issue among elementary principals in California USA was reported (Rosas, 2014). The principals sought support from assessment specialists to improve teachers' classroom assessment for learning practices.

Two dimensions received the highest average scores among the principals: Ethics and Fairness in Assessment and Purposes of Assessment, both scoring 64. Regarding Ethics and Fairness dimension, the principals shared the same view with teachers, viewing the current assessment policy as more equitable for students and highlighting this as one of the strengths of the assessment reform. This view was expressed, for example, by P1-2L and P2-1H. Both the quantitative and qualitative data show the principals' grasp of the vertical discourse of the assessment reform (Willis et al., 2013) through this dimension. Possessing adequate capacity in this dimension, provides the leaders with necessary foundation to address one of the fundamental demands of the assessment reform - ensuring fair treatment for and prioritising student diversity.

The interview data also shed light on why the principals achieved high scores in the Purposes of Assessment dimension. For instance, P3-1M attributed his good understanding of assessment purposes to his extensive experience in teaching, which enabled him to provide guidance to teachers. P6-2M expressed that he took the initiative to familiarise himself with the new policy and subsequently shared the main points of the new regulations with teachers. This proactive approach may explain the principal's strong understanding in assessment purposes mandated in the new policy, allowed him to be a knowledgeable authority in assessment practices (Smith & Engelsen, 2013).

These integrated quantitative and qualitative findings about the Purpose of Assessment dimension of assessment literacy indicate the principals' use of their cultural capital (i.e., knowledge and understanding of assessment purposes) (Harker

et al., 1990) to fulfil their assessment leadership role. This capital appeared to have been developed in different ways, for example through extensive teaching experience (P3-1M) and through conscious effort to learn the new assessment concepts (P6-2M).

These findings likely reflect the principals' solid understanding of the main points of the assessment reform, especially the emphasis on the application of assessment for learning, aiming to serve students' diverse learning needs and promote student wellbeing (MoE, 2022b).

2.2 Teachers' assessment knowledge and skills.

The results of the current study show that the assessment knowledge and skills' level of social studies teachers were at 59 (on the scale of 100). This result can be considered moderate based on comparison with several previous studies such as Alkharusi et al. (2011), Umam and Indah (2020) and Prizovskaya (2018) reviewed in Chapter Two and discussed further below. The result of this study aligns with studies conducted over the past decade in various Asian countries such as Alkharusi et al. (2011), Yamtim and Wongwanich (2014), Dunlai and You (2015), and Ara and Saeed (2020), as well as in other parts of the world such as Kanjee and Mthembu (2015), Pfeiffer-Hoens (2017), and Prizovskaya (2018). These studies, which utilised instruments with similar or overlapping assessment dimensions, have consistently reported low to moderate levels of teacher assessment literacy.

Among the six-assessment knowledge and skills dimensions, the strongest dimension of assessment literacy of the social studies teachers was Ethics and Fairness in Assessment. The average score of this dimension was slightly above 70. The interview data validate the quantitative finding. Some teachers (e.g., T2-2H, T4-2M and T5-2M) reported creating more equitable assessment practices by designing multiple assessment methods and instruments, tailored to students' diverse capabilities, which they identified in advance through diagnostic assessments. Notably, T2-2H and T4-2M reported a few times allowing students to choose assessment methods that they prepared in advance that may best reflect

their students' learning progress. This practice aligns with trends observed in Austria, Ireland, Norway, and Türkiye where secondary school teachers have incorporated alternative formats of assessment to accommodate student cultural diversity in their classrooms (Herzog-Punzenberger et al., 2020).

In addition, the practice of fairness in assessment among Indonesian teachers aligns with the differentiated teaching methods that most schools currently apply, as confirmed during the interviews with the teachers and school principals. Interestingly, Heng, Song, and Tan (2021) found that although differentiated teaching has been promoted in Singapore due to the increasing student diversity, the practice of grade-oriented summative tests, which may not accommodate student diverse learning needs, remain prevalent in their schools. This persistence was driven by the high value placed on examination results by various education stakeholders including students, educators, parents, and the wider community.

Both the assessment practices in Singapore and Indonesia reflect the contemporary conception of socio-contextual dependent teacher assessment literacy, which Willis et al. (2013) defined as a contextually situated social practice where teachers collaboratively articulate and negotiate classroom and cultural knowledge with colleagues and students throughout the assessment process to help students achieve learning goals. Whilst most Singaporean and Indonesian (Puad and Ashton, 2020; Arrafii, 2021b, Azis, 2015) teachers share similar assessment-related habitus (Bourdieu, 1977), which prioritise grading practices and test results, Indonesian educators' disposition is also strongly formed by the national philosophy of Pancasila. This philosophy emphasises tolerance and fairness within a highly diverse society. It is this national philosophy, alongside the assessment reform mandates, that likely facilitated the paradigm shift in assessment practices among Indonesian teachers, a change that is clearly reflected in the quantitative and qualitative data.

Another notable finding is that the teachers demonstrated the lowest proficiency in Analysing and Interpreting Assessment Results, with a proficiency level of only 35. A similar deficiency was reported among Thai primary teachers (Yamtim and Wongawanich, 2014). Primary teachers in charter schools in Montana America,

however, were different. They demonstrated the lowest proficiency in Assessment Ethics (Lee-Perry, 2013). This fact among many Indonesian teachers may be attributed to the introduction of the assessment practices mandated by the recent policy change, for which the teachers have limited or no previous experience or knowledge. The qualitative findings corroborated the results. Many of the teachers interviewed admitted that they experienced difficulties in interpreting assessment results, particularly related to two aspects: diagnostic assessment and the integration of conceptual, practical and behavioural learning domains into a single grade.

First, teachers expressed difficulties in interpreting the results of diagnostic assessment (T1-1M, T3-2H, and T6-2H). Even though two teachers (T1-1M and T2-2H) mentioned that in the past they performed this type of test, the current diagnostic assessment is different in terms of the method, the analysis and the use of the results. In addition, for many of the teachers, the practice of diagnostic assessment is completely new. The Learning and Assessment Guidebook published by MoE states that diagnostic assessment, conducted at the beginning of a semester or a topic/unit, aims to identify students' strengths and weaknesses to provide teachers with a basis for planning learning activities tailored to students' needs (Centre of Assessment and Learning, 2022). In doing so, teachers need to gather previous relevant assessment results, design and conduct diagnostic assessment, and then analyse and interpret data from both sources to determine students' current competencies. Clark et al. (2022) highlighted that while diagnostic assessment offers detailed insights into student status toward learning objectives, effectively utilising this information requires solid understanding of assessment concepts.

Secondly, teachers encountered difficulties in interpreting and integrating the assessment results of conceptual, practical, and behavioural domains into a single grade as highlighted by, for example, T3-2H, T4-2M, and T7-1M. Under the previous curriculum, knowledge or conceptual, practical and behavioural domains were assessed and graded separately (MoE, 2016). Conceptual aspects were evaluated through written and oral tests, while practical competencies were assessed using

performance assessments such as task-based activities, projects and/or student portfolios (Arrafii, 2021b). Behavioural domain was generally assessed using observation techniques and anecdotal records. However, the current assessment framework requires teachers to employ multiple assessment approaches to evaluate the three domains simultaneously and synthesise the results into single grade that reflects the three aspects (Centre of Assessment and Learning, 2022). The shift has proven challenging since the teachers have been accustomed to assessing the three aspects separately for more than a decade.

The adoption of diagnostic assessment in the initial phase of learning and the recommendation for multiple assessment methods to evaluate students integrated and holistic competencies reflect the government emphasis on the assessment for learning approach in classroom assessment practices. Black and Wiliam (2018) argued that assessment should be intertwined with teaching practices to guide instruction and support student learning. However, this national initiative presents new challenges for teachers due to the limited knowledge of the newly recommended concepts, which was partially due to lack of assessment trainings in pre-service education.

The teachers' weaknesses in the Analysing and Interpreting Assessment results dimension require serious attention from the government given the crucial role of this dimension in the implementation of assessment for learning approach – a key emphasis in the new assessment policy. Through the ability of analysing and interpreting assessment results, the teachers can effectively utilise assessment results to modify and adapt their instructional designs and practices to better meet students' diverse needs and characteristics.

3. Factors affecting the principals' and teachers' assessment knowledge and skills

As hypothesised, the findings indicate that assessment-related experiences, assessment courses in higher education, in-service trainings, and local support played varying roles and degrees in shaping the principals' and teachers' assessment knowledge and skills. While the quantitative findings provided valuable

insights into the influence of these factors on assessment knowledge and skills, a deeper understanding of their impact emerged from qualitative data, which offer nuanced perspectives of how the educators perceived and experienced these factors in their professional contexts. The following sections discuss the effects of each of these variables in detail, comparing and integrating quantitative results with qualitative analyses to provide a more comprehensive understanding of their effects. The discussions are structured separately for principals and teachers, situating them in relation to the literature reviewed and the theoretical frameworks of this study.

3.1 Factors affecting principals' assessment knowledge and skills

Overall, the direct and indirect impacts of the four predictors on principals' assessment knowledge and skills with varying degrees, represented in the hypothesised model, were evident as indicated by the model fit statistics in Chapter Four (quantitative findings). While the other three predictors had moderate direct effects, local support showed no direct impact on principals' assessment knowledge and skills. In addition, even though the statistical tests detected indirect impact of assessment courses and local support through assessment trainings and/or experience, the magnitudes were very small, and thus further investigation in phase two became very important. The impact of each predictor is discussed in detail in the following sections.

a. Assessment-related experiences

The SEM analysis detected a significant but relatively low effect of principal assessment experiences on their assessment competencies ($\beta = 0.19$, p value < 0.05). The qualitative data confirmed the presence of the effect, revealing that principals' experiences allowed them to adapt more quickly to the changes of assessment policy (P2-1H), give suggestions to teachers concerning the new policy (P3-1M), and identify teachers' development in assessment skills (P4-2H). These align with the leadership competencies in assessment recommended by Stiggins

and Duke (2008), which emphasise the importance of principals playing a role in guiding effective school-wide assessment practices.

Concerning the limited effect of principal experiences on their assessment knowledge and skills, Popham (2011) emphasised that experience alone may not develop sufficient assessment literacy without adequate trainings and hands-on learning in assessment. This is particularly relevant in Indonesia where Lee and Hallinger (2012) reported that principals often prioritise management and administrative tasks over academic matters in the day-to-day activities. The small impact of the principals' experience may be explained by the significant time they dedicated primarily to managerial and school administrative responsibilities (Lee & Hallinger, 2012) as it can lead to less familiarity with academic matters or teacher classroom practices. Issue about the relationship between administrative tasks and academic responsibilities should be addressed in government-initiated preparation and in-service programmes for school principals since both themes are of equal importance for school leadership responsibility.

b. Assessment courses in higher education

The principals' assessment courses in higher education showed negative impact on their assessment competencies, with standardised estimate of β = -0.34 (p value < 0.05). This finding suggests that the assessment courses that the principals have undertaken were not aligned with the current practice of student assessment outlined in the curriculum policy. This is not a highly unexpected result since while some studies revealed that assessment courses in universities increased knowledge and confidence level of assessment (Alkharusi et al., 2011; DeLuca et al., 2013; Deneen et al., 2016; A. P. Rogers et al., 2020), other studies highlighted that the courses often lack crucial components that prepare educators effectively with essential assessment knowledge, skills, beliefs or practices. For example, they may fail to cover key areas such as assessment for learning (DeLuca & Klinger, 2010), how to incorporate assessment into instruction (Schneider & Bodensohn, 2017), and assessment beliefs (Looney et al., 2018).

In addition, there seems to be a mismatch between the focus of what the principals have learned from the assessment courses and the current assessment practices they are required to perform. The average age of the principal respondents was slightly below 49 years. It is highly probable that the topics of student assessment they studied in higher education more than 25 years ago were significantly different from the priorities of the new assessment policy. This historical gap in focus may explain the negative impact observed.

The interview data validate this interpretation. The principals signalled the mismatch in their responses, citing several concerns: the courses were overly theoretical and lack of practical component (P2-1H and P3-1M); they focused solely on evaluating students' cognitive skill rather than the current requirements of assessing students' holistic competencies (P2-1H); they did not include guidance of how to use the assessment results as feedback to improve instruction and student learning (P4-2H). These problems were also highlighted in Xu and Brown's (2016) systematic review study as the contributing factors to the ineffectiveness issues of assessment education. Popham (2011) further noted that assessment topics often receive very little proportion or are even entirely absent from many pre-service education programmes.

In light of these findings, it is essential for the government to evaluate and make necessary changes in the assessment education curriculum both in the pre-service undergraduate level and graduate level, so it adequately equips educators with appropriate knowledge and skills as mandated by professional standards and assessment policy. Considering that education policies, including those concerning student assessment, will continually evolve, the government could prepare anticipatory measures within the teacher education curriculum.

The curriculum for teacher education in the field of student assessment should comprehensively and equitably cover essential elements that will remain relevant with the educators' responsibilities, both at the school and classroom level, in the long term. It should not solely emphasise one or two aspects, such as measurement theory or statistics (Xu and Brown, 2016). This is particularly critical since teacher

education programme, especially at pre-service or undergraduate stage, could equip prospective educators with a solid foundation of necessary skills.

As a result, when education reforms are introduced, these educators would already have robust base of knowledge, which can be further enhanced through in-service trainings, informal professional learning activities and professional community meetings. Popham (2011) argues that in-service trainings are crucial in addressing the gap in pre-service education programmes, as they provide teachers with practical experience and techniques for utilising assessment to enhance instructional practices.

Nevertheless, a consequence that must be anticipated from such educator pre-service curriculum policy is the possibility to expand the total credit/unit allocation for assessment-related courses. Given that historically, assessment topics have often been underrepresented in pre-service programmes (Popham, 2011; Xu & Brown, 2016) and considering, on the other hand, the critical role of student assessment in enhancing the quality of teaching and learning, this approach is entirely reasonable.

The SEM indirect analysis detected indirect effects of assessment courses on the principals' assessment knowledge and skills mediated through both in-service trainings and experience. However, the effect sizes were very low and close to zero with standardised estimates of $\beta = 0.016$ (p value < 0.01) and $\beta = 0.005$ (p value < 0.05), respectively. Whilst the indirect effects of assessment courses in higher education had no practical significance, its direct effects on both in-service trainings and assessment experiences were much stronger, accounting for 0.55 and 0.31 respectively, which subsequently affect principal assessment literacy.

The findings may be interpreted that the principals with prior assessment education in their undergraduate and/or, especially, postgraduate degrees were prioritised to attend any in-service professional development trainings in assessment. This is because they were considered to have pre-requisite knowledge for the trainings. Additionally, the courses they completed may have provided the principals priority

and credentials for appointments as exam item writers or facilitators in assessment workshops or trainings (i.e., enhancing their assessment-related experiences). From the perspective of theory of practice (Bourdieu, 1977b), the principals' education background in assessment serves as a form of cultural capital, granting them a higher status and greater opportunities to assume roles in the field of assessment compared to their colleagues. These experiences thus contribute to the principal competencies in assessment.

c. In-service professional development in assessment

The impact of in-service professional development programmes on the principals' assessment knowledge and skills was present with standardised estimates of β = 0.39 (p value < 0.05). In agreement with the quantitative analysis, the qualitative responses revealed that the assessment in-service trainings the principals have participated in throughout their career have helped them navigate the current assessment reform. For instance, it enabled them to comprehend and evaluate assessment decisions a teacher made (P3-1M). In addition, one of the principals (P4-2H) reported that he participated in the same assessment trainings with his teachers, which allowed him to transfer new insights with other teachers as well with other school principals through local principal community meetings.

This finding further supports the study of Brookhart and Moss (2013) of the importance of school principal involvement in professional development programmes. They found that principals who actively engage in a professional development programme on formative assessment were better prepared to offer teachers with more specific feedback and practical guidance and foster a more effective and collaborative conversation with their teaching staff. However, when the trainings lack effectiveness, the effect might be limited as reported by Morris (2017) who discovered in her doctoral research that in-service trainings in assessment had a limited effect on principal assessment leadership due to the shortcomings in the delivery methods.

The findings indicate that the leaders' participation in a range of in-service professional development programmes strengthen their cultural capital (Bourdieu, 1977), leading them to go beyond merely a formal leader holding an institutional capital to a respected leader for the staffs and colleagues, whose directives and advice as an assessment leader are harder to challenge.

This combination of findings provides some support for the conceptual premise that hands-on and up to date continuous professional development programmes equip educators with necessary expertise, inform them with updated knowledge, allowing them to contribute to improvement initiatives, when delivered effectively. Xu and Brown (2016) suggested that for in-service trainings to be effective, they should be situated within specific requirements of diverse educational contexts. This means that the focus of such programmes will vary depending on time, place and local requirement, necessitating a supportive system and collaborative efforts from multiple stakeholders to ensure their success.

d. Local support in student assessment

Contrary to expectations, local support for principals from teachers, principal professional community in the region/city, and school superintendent had no effect on their assessment literacy (p value = 0.199). The qualitative findings further confirm the SEM analysis results. Some of the interviewed principals acknowledged that they sought favour from teachers, particularly their commitment to exert their effort to ensure the implementation of new policy (e.g., P4-2H). P3-1M expressed that his teachers' assessment expertise was central in helping him led the reform enactment in the school. However, these data demonstrate that the teachers' support for the principals was more about the implementation of assessment policy rather than directly enhancing the principal assessment literacy.

Likewise, the qualitative data also revealed the interaction between the principals and other principals in the local principal professional community or with school superintendents did not improve their assessment literacy. For example, P3-1M

described that even though the principal community meetings were useful, they mostly focused on school management issues.

P4-2H reported that from his judgement, the superintendent lacked sufficient knowledge about current assessment concepts, limiting their role to general encouragement for implementing the new policy. In addition, P4-2H commented that the school superintendent only talked about 'general stuffs' during the coaching sessions, providing little guidance on assessment. These findings highlight the gap in meaningful support for the principals' assessment literacy from their professional network.

Whilst the local support had no direct impact on principals' assessment knowledge and skills, the SEM results also revealed a small indirect impact of local support, mediated through principal assessment-related experiences, with standardised estimate of $\beta = 0.015$ (p value < 0.01). Some of the interviewed principals, such as P2-1H, P4-2H, and P6-2M, reported that they were frequently requested by other principals to share their knowledge and experience about the enactment of the new assessment policy, especially ever since their schools were designated as the pilot schools. The opportunities to act as the speaker or facilitator concerning the new assessment concepts ultimately deepened their own assessment knowledge. This explains the indirect impact of local support (i.e., through the principal community) on their assessment knowledge, which was mediated by their experience as assessment facilitator in workshops sessions.

The situation described in the above paragraph can be interpreted as the manifestation of the existence of symbolic power among the school principals, according to theory of practice (Harker et al., 1990). Principals from pilot schools possess symbolic capital which earns them recognition and legitimacy as the individuals who are deemed superior in the eyes of other principals. Furthermore, this symbolic capital closely linked to their cultural capital (Bourdieu, 1977), exemplified by the intensive trainings they received from the Ministry of Education. These capitals grant them perceived social advantage, leading them to the appointment as the resource persons in the principal workshops.

3.2 Factors affecting teachers' assessment knowledge and skills

Identical with the SEM model of the principals, the SEM model of the teachers, which include the same four direct and indirect predictors, also fits the data well as the model fulfils all the recommended fit indices criteria. The analysis results of the direct effects of the four predictors in the teacher model mirror those in the principal model. The only difference between the two models lies in the indirect effects of assessment courses and local support, which vary slightly in their impact.

a. Assessment-related experiences

Assessment-related experiences was found to have positive influence on assessment competence with magnitude stronger than that in the principal model ($\beta = 0.29$, p value < 0.01), where more experienced teachers demonstrating a higher grasp in assessment knowledge. However, there are mixed findings in previous studies regarding the relationship between teachers' experience and their assessment competencies. This study's findings align with studies showing that years of teaching experience had significant positive association with assessment literacy (DeLuca et al., 2016a; Mertler, 2005; Zolfaghari & Ashraf, 2015). As other literature also demonstrated, other investigations have found no such connection (Alkharusi et al. 2011 and Dunlai and You, 2015). Whilst the quantitative finding of this study aligns with the first group of the studies, indicating that experience does have an impact on assessment competencies, the effect size was moderate.

The data from the interviews support the quantitative results. Most teachers reported that their past experiences, whether from daily classroom activities, serving as a test item writer, or as a facilitator in assessment workshops/trainings, have helped them adapt to and implement the assessment reform. These experiences enabled them to make decisions that bridge the differences between the previous and the current assessment approaches. Similarly, Yan, Zhang, and Fan's (2018) concluded that regular engagement in assessment activities over the course of teaching career enhances teacher's assessment expertise.

The inconsistent results among studies regarding the impact of experience on assessment literacy may indicate that its effect is intertwined simultaneously with other factors as "there is no single answer to building teachers' capacity in assessment" (Livingston & Hutchinson, 2017, p. 303).

b. Assessment courses in higher education

Similar to the principal SEM model, the teacher SEM analysis results indicated that assessment courses taken in higher education, either during initial teacher education or postgraduate studies, had a negative impact on the teacher assessment literacy, with regression standardised estimate of $\beta = -0.48$ (p value < 0.01).

This finding likely stemmed from the same issue observed in principal model, pointing to the misalignment between the focus of the course content in their past assessment coursework with the emphasis of current assessment themes. Given the average age of the teacher respondents was 43.3 years, which means it has been approximately more than twenty years ago since most of them completed their pre-service teacher training, what they learned in the university was likely no longer relevant with the current assessment policy.

For more than two decades, there has been a clear shift of focus in student assessment landscape from largely centred on psychometric principles, measurement theory, and standardised testing toward using assessment for improving student learning and social aspect of assessment (DeLuca et al., 2019; Pastore & Andrade, 2019). DeLuca and Bellara (2013) noted that the misalignment between content or syllabi of university assessment courses with assessment-related policy and standards is a key reason as to why some assessment courses fail to prepare teachers effectively. This misalignment and the gap in preparation likely contribute to the negative impact observed in this study.

The qualitative findings further support the SEM result. Some interviewed teachers who had taken assessment courses during their pre-service education admitted that the assessment topics covered were only basic, not significant or did not match with

their current practices. These findings confirm Xu and Brown (2016), highlighting several criticisms regarding assessment education in higher education, pointing issues such as insufficient coverage limited to introductory content or even disconnected from real-world classroom practices. A recent systematic review (Pastore et al., 2025) reported that mismatch issues between initial teacher education assessment coursework and teachers' required assessment practices were found in many countries. This insight aligns with the quantitative findings, highlighting a gap between what teachers learned in their assessment coursework with what they need in their classroom. Viewing it through the lens of theory of practice (Bourdieu, 1977), the reform in the student assessment policy has changed "the rule of the game" of the field. The shift in the focus of assessment has devalued the existence of one of the sources of teachers' cultural capital, making it less relevant with the current demand of the field.

Different to that in the principal model, the indirect effect of assessment courses on assessment knowledge and skills was only mediated through in-service trainings. Likewise, the effect size was also very small, with standardised estimates of only $\beta = 0.012$ (p value < 0.01), thus has no practical significance. On the other hand, despite having very small indirect impact on assessment literacy, assessment courses showed much stronger (0.54) direct impact on in-service training, which subsequently affected teachers' assessment knowledge and skills, discussed in the next section. A similar reason to that in the principal model is appropriate to explain the mediating effect of in-service professional trainings in the teacher model.

The findings indicate that teachers who undertake more assessment courses in higher education tend to participate more in and benefit more from assessment-related in-service trainings. In this regard, the in-service professional development programmes in assessment, which are more recent and aligns more closely with recent assessment policy, may help offset the misalignment between outdated past pre-service assessment courses and the ongoing assessment reform. This finding corroborates the idea of Popham (2011) and Volante and Fazio (2007) emphasising that, because many pre-service education programmes do not adequately prepare

teachers for real-world assessment challenges, in-service trainings are essential in bridging this gap by equipping teachers with hands-on experience and strategies for using assessment to inform instruction.

c. In-service professional development in assessment

Prior studies have noted the importance of in-service professional development on teacher assessment knowledge (Christoforidou & Kyriakides, 2021; Poskitt, 2014; Randel et al., 2016) as well as on teacher assessment beliefs, motivation, and practices (Bostrom & Palm, 2020; Saputra et al., 2020). In accordance with previous studies, the SEM analysis of this current research confirmed that in-service trainings directly affected teacher assessment knowledge and skills with standardised estimates of $\beta = 0.42$ (p value < 0.01), which is slightly stronger than that in principal model.

The qualitative data showed that most of the interviewed teachers perceived the trainings and workshops they attended were effective in helping them enact the new assessment concepts and approaches. Some of the teachers particularly highlighted that the trainings or workshops increased specific skills such as writing test items (T6-2H) and assessing critical thinking skills in literacy and numeracy (T7-1M).

It is important to note that most of the teachers interviewed were from pilot schools, who received regular trainings and workshops from the Ministry of Education during the first year of implementation. As introduced in Chapter One, the MoE organised monthly workshops for pilot schools in the same region and quarterly meetings involving teachers, school leaders, and local education authorities from several regions. These training sessions were designed to help the teachers understand and apply the new concepts of the reform, ensuring its successful rollout. This finding confirmed previous studies by Poskitt (2014), Lysaght and O'Leary (2017), and Christoforidou and Kyriakides (2021), which suggested that teacher professional development programmes are more effective when they are customised to meet the specific needs of the participants.

The regular meetings and workshops organised by the Indonesian Ministry of Education for the pilot schools followed an approach similar to Design-Based Implementation Research (DBIR), a collaborative small-scale site-based professional development programme developed for primary school teachers in the Republic of Ireland (Lysaght and O'Leary, 2017). The workshop sessions and meetings brought together teachers and principals from the pilot schools, representatives from local education authorities, and facilitators from Ministry of Education. They introduced new concepts, discussed the enactment strategy of the new curriculum and assessment, and addressed any issues that arose throughout the process.

However, designing in-service professional development programmes which meet teachers' diverse needs, especially on a nation-wide scale, requires particular strategies and effort. This is partly because what teachers actually need may differ from what they expressed (Xu & Brown, 2016) or they may list broad needs rather than being specific (DeLuca & Klinger, 2010). One effective strategy could be applying collaborative site-based professional development programmes tailored to specific needs of local schools, involving national and local education stakeholders. In addition, a tool designed to identify teachers' areas of growth such as Assessment for Learning Audit Instrument (AfLAI) (Lysaght and O'Leary, 2017) may be valuable to ensure that the programmes are targeted and effective.

d. Local support in student assessment

Similar to the SEM model for principals, the quantitative results did not show any significant direct effect (p value = 0.38) of local support variable (from colleagues, subject teacher community, and principals) on teacher assessment knowledge and skills. Pastore (2023) argues that local support in the form of collaboration with school colleagues is a contextual aspect which affects teachers' continuous capacity growth. However, in this study, it was not evident.

The qualitative data, however, showed varying teacher experiences. Two teachers (T3-2H and T4-2M) conveyed that the collaboration among teachers in their schools was positive. The teachers supported each other in the implementation of new

assessment policy and regularly met to discuss various aspects of student learning and assessment-related issues. Similarly, Rezagah (2022) found that support from and collaboration with school colleagues, through discussion, strategy exchange, feedback sessions with colleagues, enhanced teachers' assessment conceptual and practical knowledge. This reflects Vygotsky's social constructivism theory that posits that knowledge is constructed through interaction with others in forms such as collaboration, dialogue, and shared activities (Adams, 2006). In particular, the findings illustrate what Vygotsky called 'horizontal peer-to-peer interactions' (i.e., among colleagues) which can foster elaboration and metacognitive awareness (Pritchard & Woollard, 2013).

Nevertheless, there are other teachers (e.g., T6-2H and T7-1M) who admitted lack of support from colleagues during the assessment reform implementation. They reported encountering pessimistic attitudes, resistance to the changes, and preference for the previous assessment practices, which made collaboration and adaptation more challenging. These reluctant attitudes primarily came from some highly experienced teachers who were comfortable with the previous assessment practices. To some extent, this group of teachers leveraged their symbolic capital (Bourdieu, 1977) as "senior" teachers, derived from their vast experience and skills in assessment, to justify their behaviour.

These mixed findings of local support from teacher colleagues may contribute to the reason why the statistical analysis did not detect any direct effect of local support variable toward assessment knowledge and skills of teachers. The absence of collegial support was also reported by Aria et al. (2021). Aria et al.'s study revealed that EFL teachers in Riau province Indonesia preferred not to share their assessment findings or practices with colleagues because they wished to maintain their advantages in terms of expertise and knowledge. These findings suggest that collaborative sharing of assessment practices and challenges with other teachers does not always occur, even within a school setting, which could hinder the development of assessment literacy.

Most of the teachers interviewed acknowledged that they received various support from their school leaders, primarily in the forms of facilitating trainings or workshops, monitoring the process or motivating the teachers. However, as discussed in the previous section (i.e., section 2.2), this support often lacked technical aspects of classroom assessment practices, which likely contributed to the insignificant result of the impact of local support on teacher assessment literacy in the quantitative analysis.

Additionally, even though most teachers reported that social studies teacher community meetings were conducted regularly in their regions, topics about assessment, especially the new concepts introduced in the assessment reform, were rarely discussed. One of the teachers (i.e., T6-2H) reported that differing interpretations regarding the new assessment concepts among teachers sometimes led to heated debate. In that situation, they tended to avoid discussion on this topic altogether. The teachers' account on their experience with the local professional community add to the explanation why local support showed no significant impact on teacher assessment literacy. Umam and Indah (2020) documented similar evidence from a study regarding EFL teachers in West Java of Indonesia. They reported that teachers had limited improvement in their assessment literacy when the EFL teacher community meetings in their city were poorly organised and lacked clear objectives.

While the local support had no direct impact on teacher assessment knowledge and skills, statistically the SEM results detected very small indirect effects, mediated through both in-service trainings (β = 0.015, *p* value < 0.01) and experience (β = 0.004, *p* value < 0.05), thus showing no practical significance. However, further statistical results also show that local support affects teacher participation in in-service trainings (β = 0.39, *p* value < 0.01) and assessment-related experiences (standardised estimates = 0.15, *p* value < 0.05). Subsequently, as highlighted earlier (see section 3.1.a and 3.1.c) the two latter variables directly affect teacher assessment literacy.

As noted earlier, most teachers interviewed expressed that their school leaders facilitated various workshops and trainings to support the successful enactment of assessment reform. Some teachers also conveyed support from their colleagues where they collaboratively organised regular and informal or casual workshops and sharing sessions. The teachers' participation in multiple forms of in-service professional learning activities in turn enhanced their assessment literacy as shown in previous sections (sections 3.2a and 3.2c). These qualitative data reinforce the statistical findings demonstrating the direct impact of local support on in-service trainings and the indirect impact of local support on assessment literacy through professional development activities.

Further, interview responses align with the statistical results concerning the impact of local support on assessment experiences as the teachers explained that support from leaders and colleagues often came in the form of trust and opportunities to take on assessment-related roles such as being a test item writer (T3-2H) and/or a facilitator in assessment workshops (T6-2H). These qualitative data also showed the direct impact of local support on assessment-related experiences as well as the indirect effect of local support on assessment literacy through experience.

The qualitative data displays a dynamic interaction between teachers and colleagues as well as the unique contextual circumstances that each school encountered in enacting the assessment reform, which influenced the teachers' assessment literacy. The Indonesian teachers' awareness of their local context, including the interaction with their colleagues and school leaders, reflects 'horizontal discourses', which refers to common-sense knowledge that locally grounded, context dependent and specific to particular situations (Bernstein, 1999). Willis et al. (2013) argued that the horizontal assessment discourses shape the socio-cultural perspective of assessment literacy, emphasising the importance of context and collaboration in shaping teachers' understanding and practices.

4. The enactment of assessment reform

The third research question of this study is 'how do school principals and social studies teachers navigate the enactment of mandated assessment reform against the challenges within their unique local context?'. Building on the contemporary perspective that assessment literacy is a dynamic, multifaceted, and context-dependant professional competence and social practice (DeLuca et al., 2019; Willis et al., 2013), this study expanded its focus from focusing only on examining the assessment knowledge and skills of principals and social studies teachers, as measured in phase one (quantitative) to exploring how they implemented the assessment reform within their schools. The school environment where educators interact with students, colleagues, leaders, parents and wider community is believed to shape their assessment literacy (Rogers et al., 2020).

This study employed theory of practice (Bourdieu, 1977), particularly the concepts of habitus, capital and field, together with the theory of multiplicity of enactment (Buxton et al., 2015) as the analytical frameworks to interpret and comprehend how principals and teachers enact the assessment reform within their schools and how these enactments reflect their assessment literacy. The details of how this study employs both theories have been described in Chapter Two. Broadly speaking, these theories help interpret various decisions and actions taken by the leaders and educators, which arise from the interplay between their individual agency, the demands of the national reform, and the structural features that shape their school environments.

Findings from the qualitative (second) phase of this study, where principals and teachers expressed their thoughts and understandings, and reported their assessment practices through semi-structured interviews, provided insights into the application of their assessment literacy. Interview data shed light on their comprehension of the core features of the new assessment policy and the distinctions from the previous policy; the challenges and obstacles that could hinder its enactment; their judgments, decision making processes and actions concerning student assessment; as well as the principals' and teachers' understanding of their

school environments, which may shape the enactment of the assessment reform. In addition, the interviews revealed the educators' self-perceived assessment literacy in the context of the assessment reform implementation. Along with the findings in the quantitative phase, these findings provided a deeper understanding of the principals' and teachers' assessment literacy.

The discussion in this section is organised around the themes derived from all the interviews conducted and each will be explored as follows.

4.1 Features of the new assessment policy

The findings from the qualitative phase indicate that the principals and teachers, regardless of whether they scored high, moderate, or low, were able to identify the key features of the assessment reform, albeit to varying degrees. First, they identified that the current assessment policy brings greater benefit for students. The teachers (T2-2H, T4-2M, T6-2H, T7-1M) discovered that the current assessment policy was more engaging and enjoyable for students, which led to higher motivation to learn (T6-2H). However, only one principal (P3-1M) explicitly recognised this benefit, which is likely because most principals have limited direct interaction with students in classroom activities.

Another benefit for students was that it is more equitable. This feature was recognised by most of the teachers and principals interviewed, who perceived that, through the mandated assessment approaches, students' potential is valued and cultivated and multiple ways to exhibit their grasp of subject content are approved. These findings resonated with Black and Wiliam (2018), whose research on learning-oriented teacher assessment practices found that students demonstrated increased satisfaction with their achievement and higher motivation under a similar assessment framework.

The findings further revealed that both teachers and principals acknowledged that the assessment reform has introduced positive changes to teachers. Most teachers interviewed stated that the current assessment system offers greater flexibility,

allowing them to innovate their assessment methods, which make the process more enjoyable, despite presenting new challenges. Thus, while the assessment reform is more adaptable and engaging, teachers also acknowledged that it comes with increased challenges. Furthermore, most teachers and principals pointed out that the new assessment policy shares some similarities with the assessment framework in the previous curriculum, facilitating the teachers to implement the new concepts. One principal, P3-1M, added that the strength of the current assessment system is that it allows teachers to incorporate local cultures trough school-based projects.

Upon reviewing teachers' responses on this theme in Table 46.

The Codes of Theme 1 (Teacher) in Chapter Five (i.e., qualitative findings), it can be revealed that three teachers, T1-1M, T7-1M and T8-1L, articulated fewer assessment reform core features compared to other teachers. In addition, their responses during the interviews suggested uncertainty about particular key concepts such as diagnostic assessment and the integration of the assessment results of three domains into a single score.

These responses might signal a weak grasp of the core concepts, which may be attributable to their limited experience, as they only have been involved in enacting the reform for one year. In addition, the assessment knowledge and skills scores of T8-1L (scored low: 32) and T7-1M (scored moderate: 68 i.e., below 70), further reinforced the association between assessment literacy and teachers' ability to articulate key features of the reform.

A similar pattern emerged from the findings among the principals. P3-1M, who was in his first year of implementing the policy and scored 64 in assessment knowledge and skills (placing him in the middle position within the moderate category), identified fewer key features of the reform compared to other principals. Interestingly, P3-1M also perceived that current assessment practices are simpler, which is at odds with the view of many teachers. Additionally, despite his moderate level of assessment knowledge and skills (64), during the interview, P3-1M expressed high confidence in advising teachers about the new assessment policy,

as noted in the previous section (3.1.a) regarding the impact of experience on assessment knowledge and skills.

These expressions indicated his belief that the new assessment concepts are not difficult. The view of simplicity of the reform could be the reflection of his habitus, a concept from Bourdieu (1977), which describes an individual's set of dispositions, deeply ingrained habits, perceptions and ways of thinking, that individuals develop overtime through their experiences, social and family background, education, and cultural environment, shaping how individuals view the world, interpret events, and make decisions. P3-1M's high confidence and perspective on the assessment reform, which is contrary to the views of many teachers, might stem from his personal interpretation, rather than from a deep engagement with the new policy's complexities. His responses confirmed Leong et al. (2018) study about conceptualisation and practices of AfL in South East Asian countries, which suggests that people's beliefs affect how assessment concepts are perceived and enacted.

By contrast, even though P1-2L and P5-2M also scored low (50) and mid-range within the moderate category (68), respectively, on assessment knowledge and skills, they were able to articulate the important characteristics of the new assessment policy more clearly than P3-1L. The cause of the differences from P3-1M may be attributed to their longer exposure, as both had been implementing the reform for two years. In contrast, despite having only one year of implementation experience, P2-1H scored high (81) on assessment knowledge and skills and was also able to identify most of the core features of the assessment reform. These three principals (P1-2L, P5-2M, and P2-1H) were able to articulate the key characteristics of the new assessment policy more accurately than P3-1M, a pattern that, to some extent, reflects a differential view of assessment literacy which is context-dependent and shaped by various background experiences (Willis et al., 2013, Herppich et al., 2018, and DeLuca et al., 2019).

The findings suggest that the ability of both groups (principals and teachers) to identify the key features of the assessment reform may depend on either the length of exposure to the new concepts through experiences and relevant trainings, or the

proficiency level of the assessment knowledge and skills, or a combination of those factors. Interestingly, even though P3-1M did not score that high (64) and identified fewer attributes of the assessment reform, he was the only principal to highlight that the current assessment system facilitated teachers to integrate local cultures more easily, an important characteristic of the reform. P3-1M may reflect what Charteris and Smardon (2021) identified as a sociocultural assessment leader, which emphasises the contextual and situated nature of principal assessment literacy.

The insights shared by the teachers and principals demonstrate the extent to which they were familiar with the core aspects of the reform. Viewed from Bernstein's framework of vertical-horizontal discourses (Willis et al., 2013), their familiarity with the assessment policy reflects their grasp of the vertical discourses in assessment, which meets the criteria of knowledge with "…specialised criteria" (Bernstein, 1999, p. 159). The leaders' and educators' solid grasp of the core concepts of the reform is crucial since it provides the foundation for the enactment effort. Drawing from Buxton et al.'s (2015) work, the core concepts of the assessment reform can be classified as part of political structure (i.e., policy), which is influential and affect practices. However, the findings reported here suggests that despite receiving uniform policy socialisation from the national government, the educators exhibited varying degrees of understandings and some differing perspectives, likely due to their individual experiences and contexts.

The following section discusses a theme related to the challenges the principals and teachers encountered during the enactment of the assessment reform.

4.2 Challenges of new policy enactment

Drawing on the multiplicity of enactment theory (Buxton et al., 2015), both the structural features of the educational context in which educators work, and their individual agency, determine the extent to which and how educators choose educational practices. The structural forces within a school setting may manifest in the form of social, political, or cultural demands that affect the enactment of reform. Recognising the challenges, barriers, or obstacles, which exist within their

school and community contexts is central for principals and teachers, as it enables them to strategically navigate their actions and exert their agency in the implementation of the assessment reform. The ability to identify and address multiple contextual challenges, especially those related to the mandated assessment practices, becomes part of the educators' assessment literacy, as a context-dependent professional and social practice (Willis et al., 2013).

The results from the qualitative analyses highlighted that there are five sets of overlapping challenges identified by both principals and teachers. Responses from the teachers were clustered into two main categories: conceptual aspects and practical aspects. On the other hand, the principals pointed out that there are three types of challenges and/or potential barriers namely: teacher readiness, practical aspects of the reform, and doubts toward current assessment policy.

a. Challenges from conceptual aspects of the reform

During the interviews, more than half of the teachers reported that a lack of clarity of certain new concepts becomes one of the challenges. For example, T6-2H, from Papua Island (eastern region), who scored high in the quantitative phase, admitted that initially she was uncertain about how she should do the diagnostic assessment. As a result, she postponed performing it until the second year of her participation. She also claimed that many other teachers had the same issue as her. In addition, T8-1L, who just participated in the implementation of the new policy for one year at the time of the interview, admitted to having a weak grasp of the core concepts of the reform, likely contributed to his lower score compared to other interviewees.

The assessment policy of generating a single score from three learning domains: i.e., knowledge, skill, and behaviour domains, were also mentioned as one of the conceptual challenges by some teachers. The lack of clarity in this new assessment concept led to different implementations. The teachers reported different approaches in incorporating behaviour assessment into a single score of social studies subject. Some of them considered student behaviour as a significant element of grading practice where student positive/negative behaviour directly

affect the grade. This practice was also found in Puad and Ashton's (2020) study in West Nusa Tenggara province Indonesia. Meanwhile, some other teachers preferred to provide guidance and monitor student's behaviour development over a longer period, especially in the case of negative behaviour, before deciding whether to incorporate behaviour assessment into the final grade. A few teachers even completely excluded behaviour assessment from the grading system. Instead they applied it solely in the form of feedback and advice or giving non-grade reward such as recognition or small gifts.

These findings suggest that the challenges concerning conceptual aspects (i.e., lack of clarity) have led to multiplicity of enactment (Buxton et al., 2015) of the assessment reform at classroom level among teachers. The teachers used their existing knowledge (i.e., cultural capital) about the new concepts and personal judgement (i.e., agency) to make decisions on how and to what extent they would apply the concepts into practice.

These findings highlight the need for the Ministry of Education (MoE) to address the challenges related to the clarity of some conceptual aspects of the reform expressed by the teachers. As Oo et al. (2024) in their systematic review study and Hasim et al. (2018) in their research concerning formative assessment policy in Malaysia suggested that teachers' dispositions and literacy in a reform is one critical factor of the success of a reform enactment. Therefore, immediate action is required to mitigate potential negative consequences, especially where wider implementation of the new curriculum and the assessment reform expands in the coming years.

b. Challenges from practical aspects of the reform

Both teachers and principals also raised concern about the practical aspects of the assessment reform, which they felt could hinder its successful enactment. These included higher workload due to administrative demand (T3-2H), complexity of the new assessment policy (T5-2M), and class size (P6-2M), and time allocation (P3-1M). The issue of class size or teacher-student ratio has long been acknowledged as a persistent challenge in Indonesian classrooms (Puad & Ashton, 2020; Zuhdi, 2015)

and has also been reported in neighbouring countries such as the Philippines (Cagasan Jr. et al., 2016) and Malaysia (Sardareh, 2016). The large class sizes especially affect the application of student-centred approaches such as differentiated teaching and assessment for learning. These findings confirm the work of Black and Wiliam (2018) who suggested that time constraints and large class sizes are the practical difficulties teachers have when applying student learning oriented assessment practices.

Diverse student characteristics was also associated with increased workload as highlighted by many principals and teachers such as P6-2M and T2-2H. Aligned with what Sardareh (2016) found in Malaysian ESL classrooms, this challenge also arose as the consequence of the differentiated learning and assessment for learning policy, which require schools to adopt instructional strategies and assessment methods tailored to students' ability levels and learning needs. The challenge of addressing student diversity becomes even more daunting when combined with the issue of large class sizes, a common characteristic of many classrooms in Indonesia, especially for the remote/rural areas schools.

Some teachers and all the principals also raised concerns about the teacher readiness issue as a part of the practical challenges. The readiness issue refers to teachers' attitude (i.e., reluctance) toward the reform rather than to the conceptual element. For instance, T6-2H from Papua Island stated that many of her colleague exhibit reluctance toward the reform because they were already comfortable with the previous assessment practices. Similar findings regarding teacher unwillingness toward changes in assessment practices were also reported by Arrafii (2021a) in another province in Indonesia.

Two principals (P3-1M and P4-2H) attributed teachers' resistance to the deeply ingrained assessment paradigm (i.e., habitus; Bourdieu, 1977), which was predominantly summative and performance-oriented, with a strong emphasis on scoring and grading practices. They also admitted that changing teachers' assessment paradigm to adopt the new policy was not easy. This assessment paradigm remained prevalent among teachers in other parts of Indonesia such as in

Gowa region of South Sulawesi province as reported by (Azis, 2015) and in Mataram of West Nusa Tenggara province (Arrafii, 2021a; Puad and Ashton, 2020). The resistance to change can pose a threat to the success of assessment reform, as it may result in the reform being implemented superficially. This phenomenon was evident in South Korea, where Choi (2017) found that elementary teachers showing a reluctant response to performance assessment mandate tended to comply only with the surface level of the assessment reform, without meaningful engagement in the new practice.

Several principals and teachers also pointed out practical challenges related to the attitudes of local education authorities, which are grouped into two categories. The first category concerns a lack of support from the local education authorities for schools adopting the new curriculum, including its assessment policy. This attitude was found in three regions in eastern parts of Indonesia: North Sulawesi province (P1-2L) and two provinces in Papua Island (T5-2M and T6-2H).

The second category involves local education authorities establishing local assessment policies that did not align with the principles and objectives of the national assessment reform. This concern was raised by a principal (P3-1M) and a teacher (T3-2H), both from schools on Java Island. A similar discrepancy in educational policy, particularly regarding student assessment, between the central government (i.e., the MoE) and a local education authority on Java Island was also reported by Zulaiha et al. (2020) in their research on the assessment literacy of EFL secondary teacher.

The different attitudes of the local education authorities that teachers from different regions encountered is another manifestation of the concept of field in educational policy (Bourdieu, 1977). In this field, the teachers were experiencing power dynamics and struggles when enacting the national reform within a structured environment, shaped by some sets of norms, values and rules.

It has been long acknowledged that there is a clear disparity between the education quality in Java Island (i.e., part of western region) and non-Java Island, and

particularly, eastern parts of Indonesia lagging behind the other parts of Indonesia (Azzizah, 2015; Dharmawan & Suryadarma, 2021). The issue of disparity in education quality may be associated to the differences in attitudes observed between the local education authorities in eastern regions of Indonesia and Java Island in terms of the enactment of the assessment reform. However, regardless of the type of challenges posed by the local education authorities, the underlying cause partly lies in the government system and structure of Indonesia where local education authorities operate under regional government instead of under the Ministry of Education, with the regional governments falling under the Ministry of Home Office.

It is important to note that the diverse practical challenges faced by the schools require them to apply different approaches and strategies to address them effectively, which may be considered as the multiplicity of enactment (Buxton et al., 2015) in the implementation of the assessment policy. Therefore, when evaluating the success of a school in carrying out national education policies, such as the assessment reform, it is essential to consider the degree of support or challenges from local education authorities. Furthermore, the ability of the principals and teachers to navigate the challenges posed by local education authorities, particularly in executing the assessments reform, should be viewed as one of the indicators of their assessment literacy.

c. Comparison of challenges between urban and rural areas

The findings of this study have documented evidence of a range of challenges, issues, and potential obstacles that principals and teachers faced in enacting the assessment reform, many of which are shaped by the characteristics of the schools' contexts. According to Buxton et al. (2015) although the educators operate under the same policy and may even share a similar school context, their capacity to implement the policy may vary depending on the extent to which they can tolerate (i.e., exercise their agency) the various challenges present in their individual environment.

It is also important to note that while challenges related to conceptual aspects and to some practical aspects such as diverse student characteristics and the responses of local education authorities, were experienced by teachers and principals from both urban and rural areas, challenges in the form of resistance, stemming from paradigm shift in assessment (i.e., part of practical challenges) primarily were found among teachers from rural areas, as noted by principals and teachers of this study (T1-1M, T6-2H, P4-2H, and P5-2M). Similar resistance to new assessment policy in the previous curriculum among teachers from rural areas in Indonesia has also been documented by Arrafii (2021a) and Azis (2015).

This attitude may be attributed to limited access to information and educational resources among the teachers from this area, fewer opportunities to participate in government-organised training programmes, and relatively weaker quality of teacher professional community and network in these rural regions. From the lens of theory of practice (Bourdieu, 1977), such reluctance of the teachers of the rural areas may stem from their habitus and existing cultural capital (e.g., educational resources, trainings, and guidance). The differences in the habitus and possession of cultural capital is believed to perpetuate social hierarchy and inequality in an education system (Harker et al., 1990). Further, the above situation illustrates a curriculum policy field, where the government, leveraging its structural power, imposes a national policy, which lead to resistance from certain teachers, particularly those with limited cultural capital. This underscores the importance of the Indonesian government paying more serious attention to rural areas to ensure broader and faster acceptance of newly launched policy and enhance the likelihood of successful implementation.

In addition, some other practical challenges are closely tied to the context of the schools with all the unique characteristics inherent to the contexts, which is highly relevant with the concept of field in the theory of practice (Harker et al., 1990). Some schools, particularly those located in rural and/or eastern regions of Indonesia faced challenges such as very low teacher to student ratio and heavy administrative burdens due to the low number of staffs. On the other hand, many urban schools did not face these issues as their rural counterparts. The varying conditions of the

fields where the reform is implemented play a key role in determining how the principals and teachers make decisions and act when putting the assessment reform into practice.

4.3 Strategies for successful enactment

The current study found that the principals and teachers employed a variety of strategies to implement the mandate of the new assessment policy. Three key strategies were adopted by principals: establishing school-based assessment policy, providing support to teachers, and selecting potential and enthusiastic teachers at the initial stages of the implementation.

According to Fulmer et al. (2015), school-level policy, along with school climate and access to professional learning, are the meso-level factors that influence classroom assessment practices. This has been demonstrated by Adie et al. (2021) in a two-year project study in Queensland Australia, which focused on shifting assessment culture and practices from summative assessment and high-stakes testing to feedback-only assessment.

In the current study, many of the interviewed principals mentioned that they developed a school-level policy related to student assessment. For instance, two principals from urban area schools (P1-2L, P2-1H), took the initiative to create a school handbook to guide teachers in implementing the new policies. Meanwhile, P4-2H and P62M, from rural and urban areas respectively, established a policy where teachers who taught year seven students would continue with the same cohort as they progressed to year eight and nine. This approach was intended to help the teachers have more thorough and accurate data regarding their students, thus enabling them to tailor teaching strategies and assessments to best meet their students' needs. This later policy aligns with the national mandates emphasising the importance of focusing on student characteristics and abilities in classroom learning (MoE, 2022).

Another school-level assessment policy implemented by the principals related to determining final grades at the end of a semester or academic year. For example, P5-2M stipulated that final grade should incorporate both summative and formative assessment results, while P6-2M emphasised that student non-academic achievements or extracurricular accomplishments should be considered when determining the final grades of any subjects.

By developing school-level assessment policies, these principals have led teachers in their schools in translating the assessment reform into their own context. This seems to be consistent with Renihan and Noonan's (2018) findings, showing that the principals believed that the provision of clear vision and direction for instruction and assessment, and promoting discussion on assessment practices among colleagues, are key aspects of their leadership role in assessment.

However, the school policy established by P6-2M, that teachers of any subjects should accommodate students' non-academic or extracurricular achievements, such as in sports, into a subject's final grades might pose a validity threat. Modern validity theory defines it as the extent to which the interpretations and uses of assessment results are justified rather than the property of the test itself (Kane, 2013). It is more about the claim or inferences drawn from a score or grade. Thus, if a student receives a score, for example 60 in social studies assessment, but the teacher awards a final grade of 75 by factoring in the student's accomplishment in a national-level sport event, the grade of 75 cannot be claimed as an accurate measure of the student's proficiency in social studies.

This approach illustrates how incorporating non-academic achievements into academic grading can obscure the true assessment of student subject-specific knowledge and skills. It also applies to integrating behaviour assessment into subject final grades, discussed earlier in section about conceptual challenges. This finding suggests the importance of possessing sufficient knowledge of assessment to avoid practices that deviate from the fundamental principles of assessment.

Even though previous literature such as Oo et al. (2024), Adie et al. (2021), and Roger et al. (2020) suggested that engaging parents was an important school policy in the enactment of reform initiatives, this study found that only two principals in this study, P1-2L and P6-2M, actively pursued this approach. The two schools are located in urban areas. The primary reason for the limited prioritisation of this approach among the interviewed principals appears to be the challenges they face in involving the parents in their children educational process. These challenges stem from factors such as family's financial hardships and the prevailing belief that education is primarily school's responsibility, which is a prevalent feature in many rural areas of Indonesia. The issue of low parental involvement in children schooling has been supported by previous studies in another rural area of Indonesia such as in Arrafii (2021a).

The selection of which relevant aspects to include or exclude in school-based assessment policies aimed at realising the assessment reform is the result of dialectics of the leaders' agency and the characteristics of the schools' contexts, revealing the multifaceted enactment of the reform across schools and regions (Buxton et al., 2015).

The next main strategy employed by the principals was providing support for teachers. The most prominent aspect of this strategy is that the principals organised various forms of capacity building for their teachers, including workshops, sharing sessions, teacher collaborations, and encouraging participation in training programmes. All the interviewed principals reported that they implemented most of these activities, either regularly or on ad hoc basis. For example, some principals (P1-2L, P2-1H, P4-2H) organised peer-tutoring sharing sessions, held monthly or weekly, where more experienced or skilled colleagues were invited to share their best practices. Another support provided by the principals was facilitating workshops, where teachers collaborate to design student assessment instruments (P3-1M, P4-2H, P6-2M). These findings are similar to those of Adie et al. (2021) who also found that the school leaders in their study organised teacher capacity building through professional development activities in the forms of workshops, training sessions, discussions and collaborative teacher meetings where they discuss their

experiences, share resources or collaborate on best practices to implement feedback-only assessment strategies. The leaders' initiatives to promote collaboration among teacher colleagues potentially improve teachers' assessment literacy and reflect Vygotsky's social constructivism theory of how knowledge is constructed and developed (Adams, 2006).

Moreover, the majority of the principals stated that they provide teachers with the freedom to explore innovative assessment ideas and implement student assessment creatively. The principals also mentioned that another form of support for teachers is conducting intensive supervision to ensure the assessment of students comply with the new policy guidelines.

Another interesting strategy employed by P1-2L, P2-2M and P4-2H was selecting and collaborating with potential teachers at the initial stage of the assessment reform enactment. They did this because they need to make sure that they have a strong and reliable team to ensure a smooth transition from the previous to the current prescribed assessment practices. This last strategy, to some extent, corroborates the idea of Charteris and Smardon (2021) of sociocultural assessment leadership, which emerges and is mediated through the social and cultural contexts of schools and collaboration between leaders, teachers, and students.

The various strategies employed by the principals were the implementation of the school principals' institutional capital (formal leadership role). However, they also applied other forms of capital, such as cultural, social and symbolic capital, to make the assessment leadership practices effective (Harker et al., 1990), without which the leaders' formal functions might be less effective and may even face resistance from staffs. For example, P6-2M used his knowledge mastery of the new assessment concepts, acquired through his own initiative to learn the new concepts (i.e., a form of cultural capital), to enhance teachers' understanding and guide the implementation of the school's assessment policy. Another example, the principals' strategy for forming a pioneer team of the assessment reform enactment demonstrates the deployment of either symbolic or social capital. They may leverage their symbolic capital, where their authority as the final decision maker

legitimises their selection of the team members or they might draw upon their social capital, where they mobilise their trusted network and alliances with teachers who already support the reform.

With the exception of parents' engagement strategy, which was employed only by two principals from urban areas in this study, the other strategies were carried out by principals both from urban and rural regions. While these other assessment-related strategies appear to originate primarily from the school leaders' agency, the principals' decision whether to engage parents or not reflects a clearer manifestation of the structure-agency dialectic proposed by Buxton et al. (2015). This highlights the interplay between social structural factors, such as community dynamics and parental involvement challenges, and the principals' capacity to make decisions and act, shaping their strategies and approaches across different contexts.

As discussed in the previous section, the teachers acknowledged that students' diverse characteristics, such as general competence, learning style, and cultural background, became one of the challenges in enacting the assessment reform. To address this challenge and ensure effective application of the new assessment policy, teachers adopted several strategies. The qualitative findings reveal four main strategies employed by teachers in this study to implement the assessment reform in their classrooms: focus on students' distinct characteristics, adjustment to local context, flexible assessment formats, and peer tutoring. In comparison to principals, teachers employed narrower range of strategies. This is likely because teachers concentrate on classroom-level implementation, whereas principals must craft implementation at school level, which entails greater complexity.

The first strategy, focusing on student distinct characteristics, aligns directly with the core of the *Emancipated Learning* policy, which emphasise tailoring instruction and assessment practices to students' individual conditions and learning needs. Therefore, it is not highly surprising that this approach was adopted by most of the interviewed teachers, regardless of their schools located in urban or rural areas, reflecting the teachers' understanding and awareness of the focus of the new national policy. This strategy considers both the ability levels of the students as well

as their learning styles. For examples, teachers like T4-2M, T7-1M, dan T8-1L, prepared assessment instruments tailored to student different learning styles such as visual, auditory, or kinaesthetic learners. According to Fulmer et al. (2015), this strategy is influenced by micro-level factors, which include immediate classroom context, such as student and teacher dynamics.

Secondly, at meso-level (Fulmer et al., 2015), most of the interviewed teachers also reported adapting their assessment practices to the specific conditions of their schools, considering both the general academic abilities and cultural aspects of the students. For some teachers, this practice apparently constituted a response to the disparities in educational quality between the eastern and western parts of Indonesia, particularly Java Island, and to the cultural diversity across Indonesia, since most teaching, learning, and assessment materials provided by the central government (MoE), either in digital forms through the online platform or in print, were predominantly based on the more populous western part of the country. For example, T1-1M from eastern part of Indonesia (North Sulawesi), noted that her students' mastery of learning materials was lower compared to those in Java Island, so she adjusted the difficulty level of the assessments accordingly. Similarly, T6-2H adapted her assessments to align with the cultural context of her students. For example, when teaching about a social studies topic of basic needs, such as food, she incorporated local Papua tradition and practice related to food security and cultural customs. These strategies illustrate the teachers' adaptive practices to the field of their schools' rural conditions (Harker et a., 1990), in which the students' actual circumstances are often overlooked by the central government programmes and services, and reflect the teachers' assessment literacy as a situated and multifaceted professional competence and social practice (Willis et al., 2013).

As a result of the teachers' effort in accommodating student distinct characteristics and diverse capacity, as well as the local cultures, most teachers reported that they used flexible assessment formats when evaluating their students' learning progress. For example, T2-2H dan T4-2M, prepared multiple assessment instruments, allowing their students to choose the format that best demonstrates their abilities. Additionally, T6-2H replaced a written test with an oral test after observing that her

students could better articulate their answers verbally. These results align with the study by Herzog-Punzenberger et al (2020) about culturally responsive assessment practices which showed that secondary school teachers in Austria, Ireland, Norway and Türkiye adapted their grading practices to accommodate students' diverse characteristics.

Some teachers also implemented peer-tutoring as the strategy for applying the assessment reform in their classrooms. In the higher education teaching and learning setting, a systematic review study revealed that feedback from students' peer is one of the most effective formative assessment strategies for improving students' learning outcome (Morris et al., 2021). During the interview sessions, many of the principals and teachers highlighted that a common practice in the current assessment method is grouping students in a classroom based on the results of diagnostic assessments. Typically, students are divided into low, medium and high ability groups. Teaching and assessment activities are then differentiated for each group. However, some teachers chose not to cluster students based on their abilities. Instead, they formed mixed-ability groups, ensuring that each group include high-achieving students who could assist and motivate their peers. These teachers aimed at avoiding negative effects of ability grouping approach, such as fostering superiority among high-achieving students and causing stress and low self-esteem among low-achieving students.

Beyond demonstrating the degree of their context-dependent assessment literacy, those strategies also highlight teachers' empathy toward their students, which explain their highest average score in Ethics and Fairness dimension of assessment literacy discussed previously. In addition, this study also shows that the assessment practices of teachers and principals in Indonesia, under the current educational policy, are moving toward joining the global trend of prioritising non-discriminatory pedagogical and assessment practices that support student learning through responsive assessment practices within classrooms (Pastore, 2023).

One of the prominent ideas of theory of practice (Bourdieu, 1977) is the reproduction of inequality and social hierarchy through the interplay of an

individual's habitus, the amount and type of capital, and a social field with its own rules and characteristics. From this perspective, the variation or differences in how an educational policy is implemented across different regions, as the results of the dynamic interaction of the habitus and capital of the local education personnel and the specific context of each school, can be seen as a form of inequality.

According to Bourdieu's theory (Harker et al. 1990), national education policies often disproportionately advantage urban schools, as the principals and teachers there typically possess habitus and amount and types of capital that align more closely with the policy requirements. Conversely, schools in rural and less advantage areas are often marginalised and inevitably labelled as "underperforming", thus perpetuating social inequality the policy makers might aim to reduce.

The idea of inequality arguably is most relevant to an education system that apply national standardisation, including in student assessment through the tool of national examination as the standard benchmark of success. In such standardisation system, the success of a school, its administrators, teachers, and students are primarily determined by the exam results. However, this idea is less appropriate in an education system which prioritise and accommodate regional diversity.

In this study, whilst Bourdieu's framework of habitus, capital, and field effectively explain the reasons of the varied enactment of the assessment reform across regions and schools, its emphasis on inequality is less relevant to Indonesia's current educational landscape. The ongoing assessment reform, guided by the Emancipated Learning policy, explicitly values regional diversity. It grants the schools autonomy and flexibility to adapt their educational programmes, including the instructional and assessment methods, that fit their students' needs and local socio-cultural contexts. Therefore, under this policy, variations and differences should not be seen as a disparity of success, rather as necessary and positive adjustments to local conditions, which is more relevant for quality improvement of each school.

For this reason, this study employed multiplicity of enactment theory to interpret the variations in the enactment of assessment reform (Buxton et al., 2015).

According to this theory, differences in how the principals and teachers enact the reform are conscious responses to their schools' contexts. Rather than judging the varied implementations as compliance or non-compliance with the policy, the theory frames them as the manifestations of the principals' and teachers' assessment literacy and professional judgment.

On the other hand, this study also asserts that under the current educational policy, that grants significant autonomy and flexibility to principals and teachers in conducting student assessment, their assessment knowledge and skills is pivotal to ensure that the education quality improves across all regions of Indonesia.

4.4 Self-perceived assessment literacy

This section explores the perspectives of the school principals and teachers regarding their assessment literacy, particularly in relation to the ongoing implementation of assessment reform. According to Looney et al. (2018) and Xu and Brown (2016), teachers' and principals' perceptions of their assessment literacy play a crucial role in shaping their professional identity as assessment practitioners, which in turn influences their assessment works. In their conceptual research papers, Looney et al. (2018) introduced the concept of teacher assessment identity, whereas Xu and Brown (2016) used slightly different term: teacher identity as assessor. Looney et al. (2018) argue that teacher assessment identity is constituted by their beliefs, feelings, roles, knowledge and skills, related to assessment.

The qualitative findings from the interviews with the teachers revealed two categories of codes within this theme: Improvement of teacher assessment literacy and opportunities to better support student learning. Meanwhile, the analysis of interview data with leaders also yielded two categories of codes in this theme: Improvement of assessment literacy and the role of principals as assessment leaders in their schools.

This study found that most teachers interviewed expressed positive attitude toward the new assessment policy, despite it pushing them to deepen their understanding

of various assessment principles and concepts. Some teachers (T1-2M dan T3-2H) admitted that while certain assessment concepts had been introduced and even practiced before, they had not been thoroughly understood. The enactment of the assessment reform provided an opportunity to enhance their comprehension of these concepts. The teachers acknowledged that the entire reform process has improved their assessment literacy significantly. However, T6-2H noted that for the reform to be effectively executed and for teachers to advance their assessment competencies, they need to anticipate and address various challenges and obstacles that may arise within their individual context.

The national education policy mandates that schools provide educational services tailored to students' characteristics, abilities and learning needs by granting greater autonomy to principals and teachers in designing and implementing assessment practices at both school and classroom levels. Interestingly, this increased autonomy appears to have fostered a stronger sense of responsibility among educators to strengthen their understanding of assessment and to exercise greater creativity, as they are now required to be more independent in determining appropriate assessment methods and techniques for their students.

This approach marks a departure from the previous curriculum, where assessment policy was more prescriptive, with the Ministry of Education dictating specific techniques and methods for measuring different types of student competencies. This study confirms Xu and Brown (2016) argument that teachers' perceptions of their assessment capabilities and roles continuously evolves throughout their professional career through various forms of professional learning, reflection, and collaboration with colleagues, particularly in response to policy innovation in the field of assessment.

As noted in the literature, assessment should be embedded in the teaching practice to guide instruction and support student learning (Black and William, 2018). This study found that teachers share this perspective, acknowledging that the new assessment practices allow them to better support student in maximising their potential. The assessment for learning approach has enabled teachers to design and

implement learning experiences tailored to student characteristics and learning needs, leading to more optimal outcomes. Specifically, the teachers reported that the initial diagnostic assessment as well as various forms of formative assessments during the lesson facilitated them to support students.

Like teachers, principals also reported a better understanding of assessment since engaging with the assessment reform process. Several factors contributed to this improvement including: a formal monitoring system, which motivated the principals to familiarise themselves with the key assessment concepts (P3-1M); intensive guidance programmes provided by the MoE, though this was primarily available to pilot schools; and personal initiatives of the principals to gain a deeper understanding of the assessment concepts and policies (P4-2H and P6-2M). The principals' understanding of the assessment reform policies and the key concepts plays a crucial role in the successful implementation of the new policy. This findings align with the work of Renihan and Noonan (2018), who emphasised that an assessment reform initiative place new demand on the assessment leadership competencies required of school principals.

Another important finding of this study is that several principals (P4-2H, P5-2M, and P6-2M) stated that their efforts to familiarise themselves with the new assessment policy and concepts were part of their responsibility as a school leader to provide direction and guidance to teachers in implementing the assessment reform. P6-2M particularly linked his initiative of studying and sharing the insights of the new policy with teachers to his awareness of the high workload of teachers. This finding manifests the principals' role as assessment leaders.

The present findings accord with a previous study that demonstrated some threshold levels of success a principal need to apply to be an effective assessment leader such as: bringing irreversible change by shifting teachers' paradigm toward assessment; being coherent and integrative by encouraging teachers to develop a coherent assessment framework that benefit students' learning; being a catalyst for transforming the direction and value of education through holistic assessment practices (Tay & Tan, 2019).

5. The Relationship between Principal Assessment Literacy and Teacher Assessment Literacy.

This section discusses the fourth research question, which examined the relationship between principals' assessment literacy and teachers' assessment literacy. The analysis of the relationship between the two groups was conducted using a paired interview method involving two schools, whose principals and teachers participated and contributed to both phases of this study. The paired interview approach allowed for more accurate and detailed analysis of the relationship between principals' and teachers' assessment literacy as they came from the same school environments. Other schools participating in this study were not included in the analysis since either only the principals or the teachers were involved in the interview sessions.

The two principals revealed effective assessment leadership strategies for supporting teachers enacting the mandated assessment reform in their classrooms, while also improving teachers' assessment literacy, despite having some differences in the leadership approaches.

When comparing the individual assessment knowledge and skills scores between the principals and teachers in each school, there was contrasting patterns between school A and B. In school A, the principal's score was much lower than the teacher's, while the opposite pattern was observed in school B. When interpreting these individual scores however, it is important to note that, every observed test score reflects two components: the true ability of the test takers and some degree of measurement error. Multiple variables can contribute to measurement error and influence the observed score, including human error, environmental/external factors, respondent bias, and/or instrument limitations (Kline, 2005).

Thus, based on the scores no consistent pattern can be identified highlighting the relationship between the principals' and teachers' assessment literacy across the two schools. In addition, these scores alone do not fully represent the leaders' and educators' assessment literacy, especially when viewed from the lens of assessment

literacy as a dynamic, context-dependent, multifaceted professional and social practice (Willis et al., 2013). According to Willis and colleagues, evaluating educators' assessment literacy requires considering both vertical discourse (specialised theoretical knowledge) and horizontal discourse (contextual and practical knowledge) as proposed by Bernstein (1999). Consequently, when examining the association between the principals' and teachers' assessment literacy, it is crucial to also account for the decisions and actions they take in assessment practices, in response to the specific contextual aspects of their school environments.

5.1 Similarities between school A and B

The principals in both schools encountered teacher readiness issues when enacting the assessment reform within their schools. They noticed that the readiness problems were caused by old assessment paradigm hold by many of the teachers in their schools. Heng et al. (2021) suggested the paradigm on assessment have significant effects on teacher's understanding and practice of assessment.

To address this, both principals employed similar prominent supporting strategies by facilitating teacher assessment capacity building through workshops, sharing sessions and teacher collaboration. Both principals also provided autonomy for teachers in their assessment practices. Supporting teachers in enhancing assessment literacy is also one of the dimensions of principal assessment knowledge and skills, as measured in the quantitative phase of this study (see Chapter Three).

By facilitating workshops, sharing and collaboration sessions, they aimed to enhance teachers' assessment literacy, ensuring the successful implementation of the assessment reform. This approach aligns with what Choi (2017) called "scaffolding", which refers to facilitating teachers' professional learning to adopt the principles of reform. Choi emphasised that mandating reform alone is insufficient and cannot directly instil understanding and teacher willingness necessary for practice the transformation. It must be accompanied by the role of

principals to ensure that the teachers have the capacity and commitment to implement the reform. This is further supported by Tay and Tan (2019) suggesting that one of the characteristics of a successful assessment leader is bringing irreversible change by shifting teachers' paradigm toward assessment.

Another similar strategy found in both schools was establishing school-level assessment policy based on the principals' interpretation of the national policy. This finding supports an earlier study in Indonesia by Maisyaroh et al. (2021) focusing on principals' efforts to enhance community participation in the era of *Emancipated Learning* policy. Maisyaroh and colleagues emphasised that the effectiveness of principals as leaders and managers partly evaluated by their ability to appropriately comprehend and implement the policy reform to the specific context of their schools. The finding also aligns with Ysenbaert et al. (2020) study that in the context where schools are granted greater autonomy and flexibility in instructional and assessment practices, as in the current Flemish education system, the principal's role is central in establishing school-level assessment policies and culture. This approach helps create and develop a shared vision and understanding of student assessment practices among educators.

5.2 Differences between school A and B

The researcher identified some differences between the two schools, which despite these differences, highlight the connection between principal and teacher assessment literacy. First, principal A selected some highly capable teachers, including teacher A, in the initial phase of the reform enactment. This step demonstrates the principal's awareness of the competencies of his teaching staffs which can be capitalised for the success of school programmes.

Secondly, whilst both principals initiated a range of in-school professional learning activities, principal B took a more active role by serving as the resource person in these activities. Furthermore, whilst both principals established school assessment policy, principal B went further by continuously learning the policy updates, translating them into the local context, and repeatedly emphasising the policy in

multiple staff meetings. This approach reflects what Smith and Engelsen (2013) called “knowledgeable authority in assessment practice”, through which he can earn teachers’ trust as the assessment leader.

From social-constructivist perspective (Pritchard & Woollard, 2013), principal A tended to adopt horizontal peer-to-peer approaches, emphasising collaboration among teachers, whereas principal B typically engaged in vertical expert-novice interactions through more directive approaches as an assessment expert driving the reform implementation. Nevertheless, the paired interviews revealed strong commitment, maximum effort as well as sound strategies by both principals in fulfilling their role as assessment leaders.

5.3 Summary of paired interviews regarding the relationship between principal and teacher assessment literacy

The assessment literacy scores of the principals and teachers from these two schools could not indicate the existence of any relationship between the principal’s and teacher’s assessment literacy. However, when examining the implementation of the assessment reform in these schools, a connection between the two groups in terms of their assessment literacy becomes apparent. This connection is particularly visible through the supports provided by the principals to enhance the teachers’ assessment literacy. In addition, both principals established school-level assessment policy which indirectly stimulated the development of teachers’ capacity.

The findings suggest both principals hold threshold levels of effective assessment leadership, characterised by being coherent and integrative as they guide teachers to build an aligned assessment system that enhances student learning and serve as catalysts for transforming assessment practices at the school (Tay & Tan, 2019).

Therefore, it can be concluded that the relationship between the principals’ and teachers’ assessment literacy is rooted in the principals’ support and the school-level assessment policy that they developed. It is worth noting that “support to

enhance teacher assessment literacy" and "local assessment policy" are two of the five dimensions of principal assessment literacy measured in phase one of this study. Moreover, the relationship is reciprocal. Teachers' assessment capacity can in turn, support principals in executing their assessment leadership roles. In this study, this reciprocal connection was evident in the principal's strategic decision to select and involve a group of teachers with good assessment skills in the initial implementation period to ensure successful roll out of the reform.

6. Summary of discussion

The quantitative findings reveal that the assessment literacy of school principals and teachers fall within the moderate category when compared to prior relevant studies. Among the various dimensions of assessment literacy, analysing and interpreting assessment results (for teachers) and supporting teachers in assessment practice (for principals) emerged as the weakest areas for each group respectively. This requires serious attention from MoE given the crucial roles of those assessment literacy dimensions. This finding represents one of the practical implications of this study.

Furthermore, the quantitative analysis indicated that the moderate level of assessment literacy was directly affected by three of four hypothesised predictor variables: assessment-related experiences, in-service trainings, assessment courses in higher education. The statistical tests confirmed the impact of these factors, even though the magnitudes varied. This variation stemmed from multiple reasons as shown by qualitative data. For example, the weaker impact of experience among principals was attributed to the greater focus on administrative tasks rather than academic (including assessment-related) responsibilities, a mismatch between past assessment coursework and current assessment policy led to negative relationship between the assessment courses in higher education and assessment literacy, and in-service trainings demonstrated stronger positive impact especially when tailored to the actual needs of the principals and teachers. Both quantitative and qualitative

analysis showed no direct impact of assessment-related local support on assessment literacy.

Concerning the indirect impacts in both models, while the statistical tests indicated very low indirect effects, qualitative data suggested that indirect effects of the predictors on assessment literacy through other predictors existed.

The qualitative findings from phase two provided a more nuanced understanding of assessment literacy of principals and teachers. The key indicators of assessment literacy through the enactment of assessment reform include the ability to: identify key features of assessment reform, recognise challenges and barriers to implementing assessment reform (both broadly and within their specific contexts), develop contextually relevant strategies to address these challenges, and reflect on their assessment literacy, particularly after engaging with reform.

Based on these indicators, school principals and teachers across regions demonstrated sufficient and relatively diverse levels of assessment literacy. Notably, the principals and teachers exhibited the ability to adapt the national assessment policy to their local contexts. However, variation in implementation arose due to differences in professional experience, various trainings, collaborative interactions with peers and leaders, socio-cultural dynamics shaping their implementations.

Further, two paired interviews underscored the reciprocal relationship between principals' assessment literacy and leadership and teachers' assessment literacy. The principals' assessment-related decision makings and actions directly influenced their teachers' assessment literacy. Whilst the principals in both schools shared similar approaches in driving assessment reform, differences aroused based on their individual assessment literacy and that of their teachers. These paired interviews reinforced the overall interview findings, demonstrating that the nature of principal-teacher interaction further shape assessment literacy.

Disparities in enacting assessment reform can be interpreted from theory of practice (Bourdieu, 1977), as the results of the interplay of habitus, capital (cultural, social, and symbolic) and field. The disparities might be seen as inequalities

between regions. However, from the lens of multiplicity of enactment theory, such variations in enacting assessment reform should be seen as the principals and teachers' conscious decision makings and action reflecting their agency in responding to the demands of national policy, while navigating the unique structural features of their local contexts.

Chapter Seven: Conclusion

1. Introduction

This chapter summarises the research and presents the conclusions, demonstrating how the main findings address the overarching research questions by responding to the four research questions as outlined in section three in this chapter. It discusses the significance of the study in terms of theoretical, methodological, and empirical contributions to knowledge, building on and extending the existing knowledge and understanding of principal and teacher assessment literacy. This is then followed by an evaluation of the strengths and limitations of the study. Further, this chapter elaborates the potential implications for education policy, practices, and future research, before proposing a concluding remark.

2. Summary of the research

This study derives from an interest of supporting the enactment of Indonesian assessment reform in a new national curriculum under the *Emancipated Learning* policy. The most notable distinctive feature of the current assessment policy compared to the previous policy is the much greater autonomy and flexibility school principals and teachers have in conducting student assessment practices. Whilst this shift might seem beneficial, it requires the principals and teachers to possess adequate levels of assessment literacy, as now they have the authority to decide students' competency attainment, progression to the next grade, and graduation eligibility from primary and secondary levels.

The autonomy and flexibility basically introduce new challenges as the policy requires teachers to adopt diverse assessment methods and techniques, aligned with assessment for learning principles. The variety of assessment approaches is intended to generate authentic data on student competencies, which serves as the basis for improving subsequent learning processes. Furthermore, the new policy mandates that assessment practices should accommodate students' diverse

characteristics and learning needs. Therefore, principal and teacher assessment literacy are vital in fulfilling the demand of the reform and, more importantly, in generating valid and reliable assessment information, serving as a critical foundation for various important educational decisions.

Given the fundamental importance of principal and teacher assessment literacy, the current study focused its investigation on this critical area. In addition, studies regarding subject-specific teacher assessment literacy, including social studies subject, and principal assessment literacy as a distinct set of capabilities from that of teachers, both viewed from the lens of contemporary concept, remain scarce, making this a legitimate area for study.

Referring to the contemporary view of assessment literacy as a situated and multifaceted professional competence and social practice, this study aimed to explore Indonesian principals' and social studies teachers' assessment literacy, acknowledging it as one of the key elements of the current education policy transformation. This study considered that both assessment technical knowledge and skills and its application within a context wherein the assessment reform is enacted are equally important to understanding the principals' and teachers' assessment literacy. Hence, to answer the overarching research question, the study has formulated four research questions and addressed them systematically by employing a mixed-method research design. The use of mixed-method approach has enabled this study to uncover the principals' and teachers' assessment literacy from the contemporary viewpoint, with the combination of quantitative and qualitative data, ensuring robust data generation and analysis which incorporates the complexity of the principals' and teachers' local contexts.

The main findings of the study pertinent to each of the research questions are as follows.

3. Answering research questions

RQ1: *What is the level of Indonesian principals' and social studies teachers' assessment knowledge and skills?*

Assessment knowledge and skills represent Bernstein's vertical discourse of knowledge, informing part of Willis et al.'s (2013) concept of assessment literacy. This study suggests that assessment knowledge and skills constitute a fundamental component of assessment literacy and that, when combined with the application of assessment concepts within local contexts, they shape and reflect principals' and teachers' assessment literacy. This notion led the investigation to begin by measuring the levels of assessment knowledge and skills in both groups. This study also maintains that principal and teacher assessment literacy are two distinct constructs, given their differing areas of focus and levels of implementation. However, they are closely associated with each other and share several overlapping elements. Consequently, this study has developed and used two different instruments to measure the assessment knowledge and skills levels of each group.

The findings have demonstrated that the level of social studies teacher assessment knowledge and skills was fifty-nine (out of maximum 100), which is considered as a moderate level. This finding confirms many previous studies that employed similar assessment literacy instrument (e.g., Alkharusi et al., 2011; Yamtim and Wongwanich, 2014; Pffeifer-Hoens, 2017; and Privoskaya, 2018) and reported that relatively few teachers in various parts of the world have achieved an expert level regarding assessment knowledge and skills.

Among the six teacher assessment literacy dimensions, Analysing and Interpreting Assessment Results (T-AIA) was found to be the weakest area. The interviewed teachers acknowledged difficulties in analysing and interpreting the results of diagnostic assessment as well as in incorporating the data from three different learning domains. Limited experience, unclear concepts, and lack of guidance have been highlighted as the reasons for the low scores.

The study has identified that the principals had slightly lower levels of assessment literacy compared to that of the teachers with average score of fifty-four. Unlike previous studies that compared principal and teacher assessment literacy using identical measures (e.g., Lee-Perry, 2013; Rosas, 2014; and Pffeifer-Hoens, 2017), this study employed a different instrument for principals, although some elements overlap with that designed for teachers.

The quantitative findings indicate that the weakest area among the principals was Assessment Support for Teachers (P-SUP). The quantitative finding contradicts the qualitative data showing that supporting teachers was the most prominent strategy of the principals in ensuring the successful implementation of the assessment reform in their schools. Looking back to the items measuring the P-SUP dimension, one of the five items in this dimension which is directly related to teachers' classroom assessment practices scored very low, bringing down the overall score of this dimension. The principal assessment responsibilities, extending beyond classroom assessment practices, might explain the very low score on the item. On the other hand, the principals reported a wide range of strategies of supporting teachers, which most of them were not directly related to the technical aspects of classroom assessment.

RQ2: *How do assessment-related experiences, assessment courses in higher education, in-service assessment professional development programmes, and local support contribute to Indonesian principals' and social studies teachers' assessment knowledge and skills?*

This study hypothesised a causal model representing the direct and indirect relationships between principals' and social studies teachers' assessment literacy and four predictors respectively. The four predictors were hypothesised to have direct effects on assessment literacy. In addition, assessment courses and local support were also hypothesised to have indirect effects through both in-service assessment trainings and assessment-related experiences.

The results showed that the overall teacher assessment literacy (TAL) model fits the data as indicated by model fit statistic criteria. It suggests that the model adequately reflects the underlying pattern among the variables. Further, the statistical results revealed that while experiences, assessment courses, and in-service trainings had moderate direct impacts on teacher assessment knowledge and skills, local support did not.

The direct effect of assessment-related experiences on teacher assessment knowledge and skills was positive and moderate with standardised estimate of 0.29. The interview data justify the statistical results, revealing that the experiences to some extents have helped the teachers adapt to and bridge the transition processes. Previous studies found mixed findings regarding the impact of experience on assessment literacy, which might indicate that its impact is interconnected with other factors (Livingston and Hutchinson, 2017).

In-service trainings in assessment during the teachers' career had a stronger direct effect on teacher assessment knowledge and skills, with standardised estimates of 0.42. Some teachers in phase two reported they developed specific skills from the trainings, which helped them implement the mandated reform in their classrooms. Teachers from pilot schools interviewed in this study highlighted that the workshops and trainings they received as part of the pilot programme were designed to address their specific needs in enacting the reform and were highly useful. This finding supports previous studies (e.g., Lysaght and O'Leary, 2017 and Christoforidou and Kyriakides, 2021) that in-service professional development programmes tailored to specific needs of teachers are much more effective compared to one-size-fits-all models.

The statistical results further revealed that assessment courses in higher education had a negative direct impact on teacher assessment literacy (-0.48). This is not very surprising for two reasons. First, most of the courses were taken more than twenty years ago, indicated by the average age of the participants of over forty-three. At the time, the courses' emphasis was mainly on psychometric principles, measurement theory, and grading practices, while current policy focuses more on

using assessment data for improving student learning and on social aspects of assessment. Secondly, some teachers admitted that the assessment courses they undertook during pre-service education were mainly theoretical, not significant, and not relevant with current policy. DeLuca and Bellara (2013) and Xu and Brown (2016) suggested that the misalignment and lack coverage of the assessment courses in higher education have not prepared teachers adequately for their real-world assessment practices. The statistical analyses also indicated that the indirect impact of assessment courses on teacher assessment literacy, mediated through in-service trainings, was extremely low, rendering it impractical and can be disregarded. Additionally, no indirect impact through assessment experiences was observed.

Furthermore, even though the SEM analyses results showed no direct impact of the local support system on teacher assessment literacy and only very small indirect impacts through in-service trainings and assessment experiences, the statistical findings also revealed that local support had positive direct effects on both in-service trainings and assessment-related experiences, which in turn contributed to enhanced teacher assessment literacy. These latter findings were supported by the qualitative data, which indicated the existence of indirect impacts of local support on assessment literacy through trainings and experiences.

Likewise, the principal assessment literacy (PAL) model achieved the goodness of fit criteria, suggesting that the model is statistically consistent with the observed data and represents the causal relationships in the model without major discrepancies. Compared to the TAL model, most of the causal relationships between the variables in the PAL model were slightly lower.

The direct impact of assessment experiences of principals (0.19) on assessment competencies was lower than that of teachers. The interview responses revealed that the principals' experiences helped them adapt the reform initiatives into their schools and enabled them to support teachers enacting the mandated changes, which reflect assessment leadership role in guiding school-wide assessment practices (Stiggins and Duke, 2008). The quantitative findings also highlighted that

the principals' participation in professional learning activities had moderate direct effect (0.39) on their assessment knowledge and skills, allowing them to provide insights and evaluate teachers' assessment practices (Brookhart and Moss, 2013). The interview data supported the quantitative findings.

The assessment courses in higher education had a negative relationship with the principal assessment literacy (-0.34), likely due to the same reasons with that of teachers. On the other hand, even though the statistical test indicated a very small indirect impacts on assessment literacy, principal assessment courses in higher education correlated positively with in-service trainings and assessment-related experiences, which in turn influenced the principals' assessment knowledge and skills level. The finding supports Popham (2011) who emphasised the role of in-service professional learning in closing the gaps of pre-service education programmes.

Local support for principals from teacher colleagues, principal professional community, and school superintendent had no direct effects on the principals' assessment literacy. The support from teachers was more toward their commitment to the enactment of the reform. The principals also reported that although the professional community sessions were useful, the focus was mainly on administrative issues. In addition, some principals perceived that their school superintendents' understanding of assessment was limited, only enabling them to give broad and non-specific recommendations.

The indirect analysis detected very small mediating effect of assessment-related experiences of the principals between local support and assessment literacy. Some interviewed principals, especially those from pilot schools, admitted that they gained trust from their fellow school leaders in the respective regions, who might see them as 'expert', and were therefore often asked them to be the resource persons at local workshops and sharing sessions. The experiences as a resource person appears to be the reason behind the indirect effect of local support on assessment literacy.

RQ3: *How do school principals and social studies teachers navigate the enactment of mandated assessment reform against the challenges within their unique local context?*

Buxton et al. (2015) theorise that a wide range of teachers' decisions and actions are the results of the interplay between their agency and the structural forces, in which they operate. As cognition forms individual agency (Bandura, 1986), the principals' and teachers' comprehension of the core idea of the assessment policy is essential in shaping the enactment of the current assessment reform in Indonesia. The core concepts of the assessment reform represent the structural features (i.e., policy) (Buxton et al., 2015), which is essential in shaping the enactment of the new policy. Therefore, this study argues that it is important to first explore the educators' understanding of the key features of the new assessment policy, which along with the familiarity of their school specific contexts, would potentially affect the way they enact the mandated reform. Simultaneously, all these understanding of the reform key features, contextual awareness, the decisions they made and the actions they took in enacting the reform reflect the essential elements of their assessment literacy.

This study found that both teachers and principals, to varying degrees, were able to articulate the key features of the assessment reform, such as the advantages for students and teachers. Some of them also could identify the distinctive feature of the reform that it explicitly recommends schools to accommodate local cultures into student learning. Most of them were also aware that despite granting greater autonomy and flexibility, the reform initiative also introduced new challenges. The findings suggest that the differences in the ability to recognise the main attributes of the new assessment policy stem for two main reasons: the length of exposure to the new concepts through experience and relevant trainings, or the proficiency level of the assessment knowledge and skills, or a combination of those factors.

Recognising the challenges, barriers, or obstacles within their schools and broader community is central for principals and teachers, as it allows them to strategically navigate actions and exert their agency in the implementation of the assessment

reform. The qualitative findings reveal three common challenges identified by both groups encompassing conceptual challenges, practical challenges, and attitude toward the new policy.

Two types of challenges were consistently identified concerning the conceptual aspects of the assessment policy: designing and interpreting diagnostic assessment and incorporating three types of learning domains (knowledge, skill, and behaviour) into a single score. Lack of clarity of the concepts has led to differential practices. In particular, teachers reported different approaches in applying student behaviour assessment and how to incorporate it with two other domains.

Both groups reported higher workload, diverse student abilities and characteristics, and teacher readiness as the practical challenges. Higher workload issues were related to administrative burdens, complexity of the new policy, large class sizes, and limited time allocation. Diverse student abilities and characteristics brought practical challenges as the consequence of mandated differentiated learning and assessment for learning approaches. Meanwhile, challenges regarding teacher readiness were mainly due to teacher resistance to implementing the new policy, either because they were reluctant to leave their comfort zone or they remain attached to the old assessment paradigm. The resistance due to the issue of paradigm shift in assessment mainly occurred in schools from rural areas, which arguably stemmed from the teachers' habitus and existing cultural capital (Bourdieu, 1977).

In addition, several participants mentioned that practical challenges also came from local education authorities, which either arose from inadequate support or inconsistencies between local regulations and national policy. The former issue was reported by participants from eastern parts of Indonesia, while the latter occurred in Jawa Island, the western parts of Indonesia. This different attitude of local authorities might associate with disparity in education quality between eastern and western regions of Indonesia (Azzizah, 2015; and Dharmawan and Suryadarma, 2021).

The principals' and teachers' levels of assessment knowledge and skills, understanding of the core concepts of assessment reform, and awareness of challenges and potential barriers within and beyond their local contexts, collectively led to a wide range of approaches and strategies in enacting the assessment reform, manifesting as and explained by Buxton et al.'s (2015) multiplicity of enactments theory.

The principals adopted three main strategies namely establishing school-level assessment policy, providing support to teachers, and selecting potential and enthusiastic teachers at the initial stages of the implementation. The school-level policy includes developing a guidebook for teachers; looping strategy by which subject teachers remain with the same student cohort throughout three years of learning and setting school regulation on final grade calculation. However, some schools may have potential assessment validity issues concerning the latter policy, particularly pertinent to the integration of behaviour assessment and non-academic achievement into the final grade.

Providing support for teachers comes in several forms. The most prominent form is organising various capacity building for their teachers, including workshops, sharing sessions, teacher collaborations, and encouraging participation in training programmes. Further, principals also support teachers through intensive supervision and providing teachers with freedom to explore innovative assessment ideas and implement student assessment creatively.

Another interesting strategy that the principals performed was selecting potential teachers at the initial stage of the assessment reform enactment. They did this because they need to make sure that they have a strong and reliable team to ensure a smooth transition from the previous to the current prescribed assessment practices.

There was no clear distinction of strategies and approaches between principals from rural and urban areas. The interview responses also revealed that only two

principals, both from urban school, engaged parents more in student assessment as part of school policy.

Teachers adopted classroom assessment strategies, which primarily in response to students' diverse characteristics, such as focusing on students' individual distinct characteristics, adjustment to local context, flexible assessment formats, and peer tutoring. Most teachers, regardless from rural or urban schools, at classroom level, focused on students' characteristics by tailoring instruction and assessment practices to students' individual conditions (such as ability level or learning styles) and learning needs. Some teachers also adjusted their assessment strategies, addressing the disparity in education quality between eastern and western parts of Indonesia and cultural differences.

As the consequence of accommodating student diverse characteristics and social-cultural differences, teachers employed flexible assessment formats when evaluating their students' learning progress. In addition, a peer-tutoring strategy was chosen, where teachers distributed high-achieving students to all groups of students with mixed abilities in a classroom, instead of dividing the class into high, medium, and low ability groups, as practiced by some other teachers.

RQ4: *How do the Indonesian principals' assessment literacy and leadership in assessment affect social studies teachers' assessment literacy?*

As the two interconnected direct executors of the new educational policy on the field, examining each group's assessment literacy and how they affect one another are equally important. The latter part was investigated through research question four. To answer the research question, this study used two paired interviews: school A and B, following the engagement of both the principals and teachers in these schools throughout both phases.

In school A, despite achieving a low-level category in assessment knowledge and skills, the principal demonstrated his assessment literacy better through a wide range of strategies of supporting teachers enacting the assessment reform:

improving teacher assessment capacity and supporting teacher autonomy. Creating a guidebook for teachers and engaging teachers deemed to possess good level of assessment expertise in the initial stage of reform also marked the relationship between them. In particular, the assessment leadership actions of principal A were directed toward shifting teachers' paradigm as it may hinder the successful implementation of the new assessment policy. From the perspective of social constructivism theory (Pritchard & Woollard, 2013) this principal appears to adopt horizontal peer-to-peer interaction with his teachers as his assessment leadership style.

In school B, the principal stepped beyond only facilitating workshops and in-house trainings for teachers. He took the role as one of the resource persons in the capacity building sessions. This represented clearly his role as an assessment leader, which might be referred to as knowledgeable authority in assessment (Smith and Engelsen, 2013). His strong assessment leadership role might stem from his initiatives to proactively study the new assessment approaches before sharing them with teachers. Additionally, like principal A, principal B also established a school-level assessment policy guiding his teachers in enacting the mandated assessment reform. The principal-teacher pair in school B further reveals a close relationship between principal and teacher assessment literacy. However, he tended to apply the vertical expert-novice interaction in his assessment leadership approach (Pritchard & Woollard, 2013).

The leadership in assessment of both principals was an effort to ensure that teachers have the capacity and commitment to realise the reform, thus bridging the connection between principal and teacher assessment literacy. Furthermore, the necessity to understand assessment literacy beyond merely the levels (i.e., scores) of assessment knowledge and skills become more apparent through these paired interviews. The scores alone might not show any relationship but when practices of assessment leadership and other assessment-related activities within schools were examined, the association between the two becomes clear.

Overarching Research Question:

What is the current state of Indonesian principals' and social studies teachers' assessment literacy in enacting the assessment reform in diverse school contexts?

From the responses to research questions 1-4 collectively, it can be concluded that, in general, Indonesian school principals and social studies teachers have moderate levels of assessment literacy, though there is variation across regions and schools. These differing competencies are influenced by four factors namely assessment-related experiences, in-service trainings in assessment, assessment coursework in higher education, and local support, either directly or indirectly.

The average scores of assessment knowledge and skills among principals and social studies teachers, which fall under moderate categories, along with supporting evidence from the interview data demonstrate that school leaders and educators in Indonesia already have reasonably adequate assessment literacy to enact the assessment reform in their schools, albeit not yet perfect. This is evident through the diverse strategies they employed to implement the new assessment policy and respond to various challenges, obstacles, and local conditions.

This study suggests that through the enactment of assessment reform the assessment literacy of school leaders and teachers have the potential to be further improved. Therefore, to ensure that educational reform policies are implemented more effectively, supported by increasingly strengthened competencies of the school principals and teachers, this study provides several recommendations, which are elaborated in other section in this chapter.

4. Contributions of the study to knowledge

The findings from this study make several contributions to the field of principal and teacher assessment literacy encompassing theoretical, methodological and empirical contributions. The primary contribution of the current study is the development of a conceptual framework to understand principal and teacher assessment literacy. This research also contributes to the development of instruments to investigate social studies teacher and school principal assessment

literacy respectively. In addition, this study expands the knowledge of assessment literacy by providing empirical evidence of principals and teachers, particularly social studies teachers, from Indonesian context.

4.1 Theoretical contribution to knowledge

Based on the contemporary conception of assessment literacy as negotiated, multi-faceted contextual dependent professional and social competencies and practices (Willis et al., 2013; DeLuca et al., 2019; Pastore and Andrade, 2019), this study highlighted two central elements of the concept, namely mastery level of assessment knowledge and skills and the application of assessment responsive to social and cultural context wherein it is practiced. Consequently, to thoroughly understand and evaluate principal and teacher assessment literacy, both elements are equally important and need to be considered.

The findings indicate that, for both groups, three factors directly influence assessment literacy which are assessment courses in higher education, in-service professional development programmes in assessment, and assessment-related experience. In addition, local support in assessment from fellow teachers, the leader, and local professional community has an indirect impact on assessment literacy mediated through assessment experiences and assessment in-service trainings.

Regarding the second element of the concept, the research has also shown that the principals and teachers have enacted the assessment reform in various degrees and forms (e.g., diagnostic assessment, addressing students' distinct characteristics, and combining assessment results from three domains into a single grade) as the results of the interplay between their agency and the structural features, shaping the contexts in which they are working (Buxton et al., 2015; Bourdieu, 1977). This study highlights that the principals' and teachers' assessment literacy underpin their agency in enacting the assessment reform. They exhibited different approaches and strategies in response to the diverse challenges and barriers they encountered within their local contexts. On the other hand, the findings suggests that while the

schools, especially those from the same regions, may have similar social, cultural, and political features, the principals and teachers demonstrate different levels of engagement with the mandated reform, which were partly influenced by the mastery level of assessment knowledge and skills, especially concerning the principles of assessment for learning concepts. Therefore, this study uses the term "*adapted and differential enactments*" to describe the principal and teacher enactment of the assessment reform within their schools.

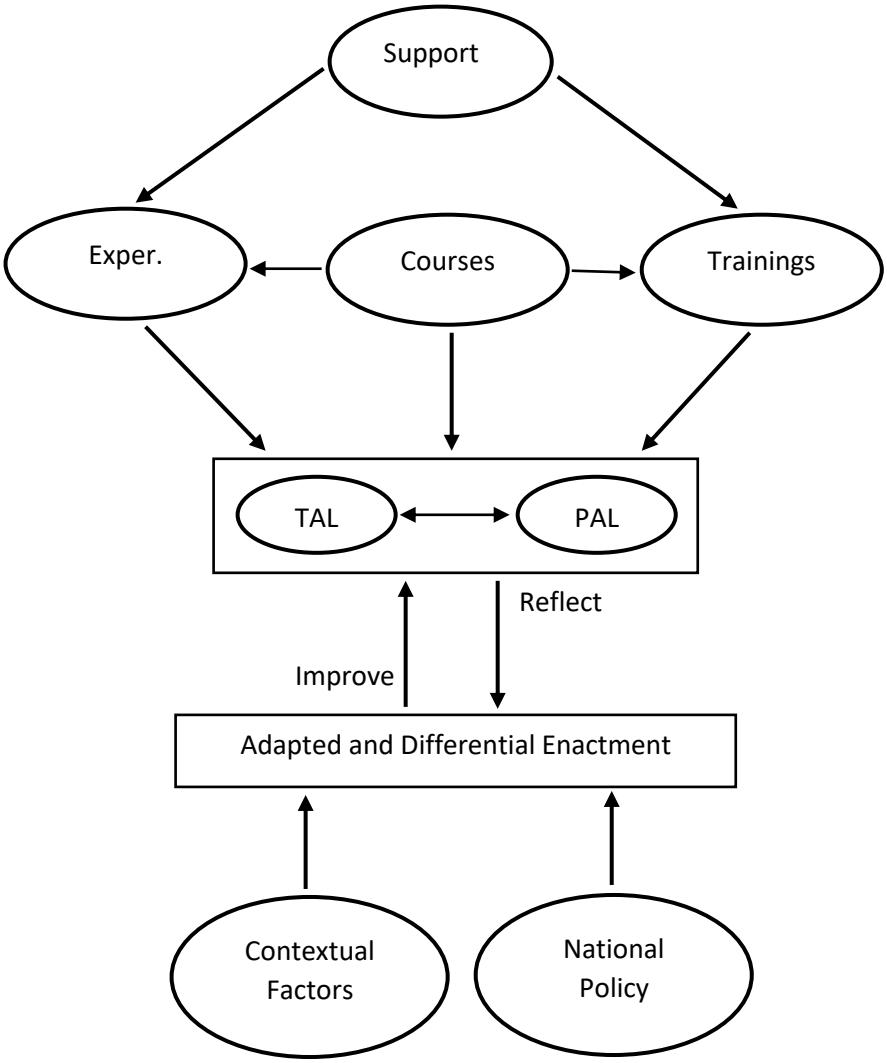
The findings have also shown that the enactment of the assessment reform has increased the assessment literacy of principals and teachers. This study further indicates that, despite their distinct role in student assessment, there is a close relationship between principal and teacher assessment literacy, influencing one another.

Taken together, this research suggests that the adapted and differential degrees and forms of enactment of the school principals and teachers need to be viewed as the manifestation of their assessment literacy rather than the representation of the fidelity or infidelity of implementation of the mandated assessment reform, while at the same time recognises the essential role of assessment knowledge and skills mastery level. Hence, this study further suggests that while the adapted and differential degrees and forms of assessment reform enactment may indicate the principals' and teachers' purposeful agentic actions in responding to the complex local structural features, caution is needed not to use the idea of multi-faceted policy implementations (i.e., lack or superficial implementations) as an excuse of merely insufficient assessment expertise.

The following diagram depicts the conceptual framework this study introduced in understanding principal and teacher assessment literacy.

Figure 14

Conceptual Framework of Understanding Principal and Teacher Assessment Literacy



While previous studies have developed conceptual frameworks relating assessment literacy to educational assessment and teacher training (Xu and Brown, 2016), approaches to classroom assessment (DeLuca, LaPointe-McEwan, and Luhanga, 2016), teacher identity as assessors (Looney, Cumming, van Der Kleij, and Harris, 2108), and three dimensions of assessment literacy (conceptual, praxeological, and socio-emotional dimensions) (Pastore and Andrade, 2019), this current study introduced a conceptual framework to view and link the assessment literacy of practicing principals and teachers, combining the knowledge and skills and context-dependent practices elements of assessment.

4.2 Methodological contribution to knowledge

This study makes a clear distinction between principal and teacher assessment literacy, where a principal assessment literacy covers dimensions beyond technical aspects of student and classroom assessment. This study developed five dimensions of principal assessment literacy namely purpose of assessment, supporting teacher assessment literacy, assessment communication, ethics and fairness in assessment, and school assessment policy. Six assessment dimensions of teacher in this study are purpose of assessment, constructing and administering assessments, analysing and interpreting assessment, communicating assessment results, reliability and validity of assessment, and ethics and fairness in assessment. Based on that notion, this study has developed school principal and social studies teacher assessment literacy measurement instruments, which can be useful for future investigations.

Another methodological contribution of this study relates to the development of the conceptual framework. While previous assessment literacy frameworks have been developed through a literature scoping review (Xu and Brown, 2016), systematic literature review (Looney et al., 2018), and Delphi inquiry method using expert feedback (Pastore and Andrade, 2019), this study developed its framework using empirical data from school principals and teachers. DeLuca et al.'s (2016) approaches to classroom assessment literacy inventory instrument also utilised empirical data from participant when developing their instruments. However, they establish the framework only from quantitative data. The present study utilised both quantitative and qualitative data in developing the framework.

4.3 Empirical contribution to knowledge

This study expands the body of knowledge regarding teacher assessment literacy by focusing on social studies teacher. Literature (Brookhart, 2011; Xu and Brown, 2016; Pastore and Andrade, 2019) emphasise the importance of binding teacher assessment literacy to a specific subject knowledge as student assessment practices need to be tailored according to a particular subject content due to its different characteristics.

Globally, studies about social studies/social sciences teacher assessment literacy are exceptionally scarce in the literature. So far, a careful literature search of international education journal databases such as ERIC, Web of Science, and Scopus only revealed two studies related to social studies teachers (Mertler, 2005; Clark, 2015). However, both studies investigated only one part of social studies teacher assessment literacy concept used in this current study. Whilst Mertler (2005) examined social studies teachers' level of assessment knowledge and skills using general teacher assessment literacy instruments and excluding the specific characteristics of social studies school subject, Clark (2015) focused solely on pre-service social studies student teachers' assessment literacy through their practices.

In the Indonesian context, research about teacher assessment literacy is mostly related to EFL teachers (Arrafii, 2021b; Aria et al., 2021; Rasyidah et al., 2020; Zulaiha et al., 2020; Umam and Indah, 2020; Zulaiha and Mulyono, 2020). While only a few studies have examined science teacher assessment literacy (such as Efendi et al., 2018), no research specifically focuses on social studies teachers.

Furthermore, this study specifically adds to the empirical knowledge about principal assessment literacy by incorporating the assessment knowledge and skills and the assessment leadership practice of principals, as the review of literature revealed that previous studies mostly focused only on either the knowledge and skills element (Lee-Perry, 2013; Rosas, 2014; Pffeifer-Hoens, 2017; Ball, 2017) or the socio-contextual assessment leadership practices of the principal (Renihan and Noonan, 2018; Smith, 2019; and Charteris and Smardon, 2021). In addition, this research is the first study investigating principal assessment literacy in Indonesia.

5. Implications of The Study

The results of this research have a number of important implications for practice, policy and future research as described in the following three sections.

5.1 Implications for practice

The data shows that teachers and principals, mostly those coming from non-pilot schools and rural areas, encountered some difficulty due to the lack understanding of some new concepts, such as diagnostic assessment and the incorporation of the assessment results from three learning domains into single score. Hence, before being formally implemented nationwide, it is important for the government to allow an adequate transition period from the previous curriculum and assessment policy, especially for rural and remote areas, to ensure the direct executors of the new policy have a sufficient understanding of the mandated new policy. This might ensure that the basic idea of the transformation being enacted into practice more effectively.

Secondly, referring to the different attitudes of local education authorities in responding to the *Emancipated Learning* policy, and the fact that structurally they are under the control of the Indonesian Ministry of Home Office, it is crucial for the Ministry of Education to not only maintain collaboration with the Ministry of Home Office but also directly with the local authorities through any MoE branches spreading across regions. This strategy could build the same vision and strengthen collaboration between institutions. By doing this, it is expected that schools will have better support both from the Ministry of Education and the local government.

Another important practical implication is that, given the diverse social and cultural features of local contexts across Indonesia, the evaluation of the extent to which a school have implemented the education policy should be formative in nature instead of judgemental. The formative approach of the evaluation could facilitate the development of principal and teacher assessment literacy and other pedagogical skills, as any treatment programmes, such as workshops, sharing best practices, and trainings, provided by the government, local authorities, or other providers will focus on and address mostly the weak areas of the reform. In this way, the schools are encouraged to make continuous progression of how they implement the reform. Thus, the emphasis of the current education policy on the assessment for learning principles, which is applied to the practice of student

assessment, might even be of greater benefit when applied at a higher level (i.e., school), in evaluating the so-called policy adherence at school level.

5.2 Implications for policy

Drawing on the findings, this study suggests some important policy recommendations concerning principal and teacher pre-service education and in-service trainings. Firstly, the findings suggest that past assessment education in higher education did not prepare current practicing educators adequately due to misalignment of the curriculum content with current emphasis of assessment policy and/or insufficiently covered within the curriculum. Given that assessment policy in the long term may continuously evolve, shifting from one focus to another, the assessment topics in pre-service teacher education should cover all necessary foundational elements equitably, preparing prospective educators with a strong knowledge base, by which it can be expected to remain relevant, as long as possible, for future principals and teachers in carrying out their assessment responsibilities, despite changes in assessment policies. Even though it will likely have a consequence for adding more credit/unit allocation for assessment topics, highlighting the critical function of student assessment to enhance learning, this change is highly recommended.

Secondly, the findings from phase one and two of the study indicate that experience has a positive direct impact on principal and teacher assessment literacy. Therefore, pre-service teacher education programmes should provide adequate and equal amount of practical experience across pedagogic and assessment areas for student teachers during their teaching practicum period.

Another implication of the current study concerns in-service professional development programmes, which evidently affect assessment literacy of principals and teachers. This research provides evidence that schools encounter diverse challenges and potential barriers in enacting the assessment reform relating to local contexts. In response to such challenges, government-funded in-service teacher trainings should design its programmes addressing the contextual elements of

assessment practices. The trainings should provide enough space to accommodate teachers' specific needs. By employing the approaches, it can be expected that the trainings will be more relevant and grounded, in supporting teachers carrying out their assessment tasks.

In addition, while this study documented evidence of the importance of principal assessment literacy in fulfilling assessment leadership roles, existing school principal preparation programmes in Indonesia focus only on topics of general leadership, management, and school administration. The present study suggests that the government should highly consider including assessment-related topics into the principal preparation programmes.

6. Limitation of the study and suggestions for future research

The present study developed two new quantitative data collection instruments to investigate principal and teacher assessment knowledge and skills respectively. In general, both instruments have fulfilled most of the reliability and validity requirements, except the assessment-related experiences as the predictor variable of assessment literacy. It had reliability and validity coefficients slightly below the recommended criteria.

After close inspection to the relevant questionnaire items (indicators), the potential source of the reliability and validity issue lies in the way the items were written (i.e., the sentences and/or the word choices). For example, one item of the experience construct/variable, asked the respondents to indicate how many times they have served as an assessment "instructor", with 10-point scale. The average value of this item was extremely low, leading to a highly skewed distribution, which subsequently lower the overall reliability and validity of the variable. It seems that the word choice "instructor" might have affected the participants' responses. The participants apparently may have interpreted the word as being a highly formal role in a training setting, whilst in fact they were expected to count their experience as an assessment instructor even in a casual situation, such as being a speaker in a sharing best practice informal session among teacher colleagues in a school.

In addition, a few items in the new scales had somewhat low factor loadings to the relevant constructs, despite the CFA (measurement) models fulfil all the goodness of fit criteria. Again, after careful examination of the items, it seems that the range within each scale-point might have caused the responses to positively skew, leading to low factor loading. Therefore, future studies are expected to address the issue of item construction to improve the instruments' reliability and validity.

In terms of the qualitative data, future inquiries investigating a similar topic as the current study may also include the voice of students which will allow triangulation in data analysis from three different sources: teachers, school principals, and students. Such a move is reasonable since improving student learning outcomes is often the focus of many educational policy reforms, thus student voice could provide richer data of the phenomena under investigation. Arguably, the age of the student needs to be mature enough so their answers can be meaningful and reliable. One of the reasons this study did not engage students was because it involved mostly year-7 teachers, whose students were 12-13 years old. Additionally, further research is required to explore the assessment literacy of principals and teachers from other geographical areas to better capture the huge complexity of regions in Indonesia.

7. Concluding Remark

This study explored Indonesian school principals' and social studies teachers' assessment literacy based on a contemporary conception of assessment literacy as a negotiated multifaceted and context-dependent professional and social competence and practice (Willis et al., 2013, DeLuca et al., 2019). In doing so, this study employed a mixed-method research design, through which it examined the assessment knowledge and skills levels of the principals and teachers in the quantitative phase, followed by investigating the enactment of assessment reform in the qualitative phase. This approach provides a thorough picture of assessment literacy beyond merely the assessment knowledge and skills level but also include the practical component of assessment literacy.

The study shows that the Indonesian principals' and social studies teachers' assessment knowledge and skills levels were moderate. This study also reveals that the principals and teachers from diverse geographical areas demonstrated adapted and differential degrees and forms of enactment of the assessment reform, as the results of the interplay between their agency and the structural features within their local contexts. Findings indicate that the mastery of assessment knowledge and skill influenced the enactment of mandated assessment policy. Hence, this study suggests that both the mastery level of assessment knowledge and skills along with the application of the assessment policy within local contexts, simultaneously reflect principals' and teachers' assessment literacy.

The primary contribution of the current study to the body of knowledge is the conceptual framework developed to understand and link principal and teacher assessment literacy, that combines the knowledge and skills and context-dependent practices elements of assessment. Through the development of two data collection instruments of assessment knowledge and skills of principals and teachers this study contributes to future research wishing to investigate similar topics. Additionally, this study adds to the empirical literature of assessment literacy by focusing on social studies teachers and school principals in Indonesia.

This study makes some recommendations for the implementation of the assessment reform concerning the transition period from the previous policy, collaboration with local authorities, and monitoring and evaluation approach of the reform. This study also recommends some transformation of assessment education programmes both in higher education and in-service professional development programmes. Lastly, this study provides direction for future research concerning principal and teacher assessment literacy and practice.

References

Absolum, M., Flockton, L., Hattie, J., Hipkins, R., & Reid, I. (2009). *Directions for Assessment in New Zealand: developing students' assessment capabilities*. http://assessment.tki.org.nz/content/download/176/958/version/1/file/Directions+for+Assessment+in+New+Zealand.PDF

Adams, K. A., & Lawrence, E. K. (2019). *Research methods, statistics, and applications* (2nd ed.). SAGE Publications.

Adams, P. (2006). Exploring social constructivism: theories and practicalities. *Education 3-13*, *34*(3), 243-257. https://doi.org/10.1080/03004270600898893

Adie, L., Addison, B., & Lingard, B. (2021). Assessment and learning: an in-depth analysis of change in one school's assessment culture. *Oxford review of education*, *47*(3), 404-422. https://doi.org/10.1080/03054985.2020.1850436

Adie, L., Stobart, G., & Cumming, J. (2020). The construction of the teacher as expert assessor. *Asia-Pacific Journal of Teacher Education*, *48*(4), 436-453. https://doi.org/10.1080/1359866X.2019.1633623

Alkharusi, H., Kazem, A. M., & Al-Musawai, A. (2011). Knowledge, skills, and attitudes of preservice and inservice teachers in educational measurement. *Asia-Pacific Journal of Teacher Education*, *39*(2), 113-123. https://doi.org/10.1080/1359866x.2011.560649

Anis, M., & Anwar, C. (2020). Self-organized learning environment teaching strategy for ELT in Merdeka Belajar concept for high school student in Indonesia. *Jounal of English Education*, *5*(2), 199-205.

Ara, N., & Saeed, M. (2020). Relationship Between Primary School Teachers' Assessment Literacy and Classroom Assessment Practices. *Global Regional Review*, *V*(1), 9. https://doi.org/10.31703/grr.2020(V-I).53

Aria, D., Sukyadi, D., & Kurniawan, E. (2021). Teacher Assessment Literacy: Indonesian EFL Secondary Teachers' Self-Perceived on Classroom-Based Ass.Practices. *Jounal of English Education*, *10*(1), 15-26. https://doi.org/10.25134/erjee.v10i1.5349

Arrafii, M. A. (2020). *Towards Formative Assessment: Exploring EFL Teachers' Conceptions and Practices of Assessment in Indonesia* [Traditional, University of Leicester]. Leicester.

Arrafii, M. A. (2021a). Assessment reform in Indonesia: contextual barriers and opportunities for implementation. *Asia Pacific Journal of Education*, 1-16. https://doi.org/10.1080/02188791.2021.1898931

Arrafii, M. A. (2021b). Indonesian teachers' conceptions of values and dimensions of assessment practice: The effect of teachers' characteristics. *Teaching and Teacher Education*, *98*. https://doi.org/10.1016/j.tate.2020.103245

Azis, A. (2015). Conceptions and Practices of Assessment: A Case of Teachers Representing Improvement Conception. *TEFLIN Journal - A publication on the teaching and learning of English*, *26*(2). https://doi.org/10.15639/teflinjournal.v26i2/129-154

Azzizah, Y. (2015). Socio-Economic Factors on Indonesia Education Disparity. *International Education Studies*, *8*(12), 13.

Bandura, A. (1986). *Social foundations of thought and action : a social cognitive theory*. Prentice-Hall.

Bernstein, B. (1999). Vertical and Horizontal Discourse: An essay. *British journal of sociology of education*, *20*(2), 157-173. https://doi.org/10.1080/01425699995380

Bielick, S. (2016). Survey and Questionnaires. In D. Wyse, N. Selwyn, E. Smith, & L. E. Suter (Eds.), *The BERA/SAGE Handbook of Educational Research* (Vol. 2). SAGE Publication Ltd.

Bielsa, G. (2010). Pragmatism and the Philosophical Foundations of Mixed Methods Research. In A. Tashakkori & C. B. Teddlie (Eds.), *SAGE Handbook of Mixed Methods in Social & Behavioral Research* (2nd ed., pp. 95-118). SAGE Publications, Inc.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for Learning, Putting it into Practice*. Open University Press.

Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in education : principles, policy & practice*, *5*(1), 7-74. https://doi.org/10.1080/0969595980050102

Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in education : principles, policy & practice*, *25*(6), 551-575. https://doi.org/10.1080/0969594X.2018.1441807

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch Model, Fundamental Measurement in the Human Sciences*. Routledge Taylor & Francis.

Bostrom, E., & Palm, T. (2020). Expectancy value theory as an explanatory theory the effect of professional development programmes in formative assessment on teacher practice. *Teacher Development*, *24*(4), 539-558. https://doi.org/10.1080/13664530.2020.1782975

Bourdieu, P. (1977). *Outline of a Theory of Practice* (R. Nice, Trans.). Cambridge University Press.

bps.go.id. (2023). *Hasil Sensus Penduduk*. Badan Pusat Statistik.

Brookhart, S. M. (2011). Educational Assessment Knowledge and Skills for Teachers. *Educational Measurement: Issues and Practice*, *30*(1), 3-12.

Brookhart, S. M., & Moss, C. M. (2013). Leading by Learning. *4*(8). https://journals.sagepub.com/doi/10.1177/003172171309400804

Brown, G. T. (2008). Assessment Literacy Trainin and Teacher Conception of Assessment. In C. M. Rubie-Davies & C. Rowlinson (Eds.), *Challenging Thinking about Teaching and Learning* (pp. 285-302). Novapublisher.

Bryman, A. (2016). *Social research methods* (5th ed.). Oxford University press.

Buxton, C. A., Allexsaht-Snider, M., Kayumova, S., Aghasaleh, R., Choi, Y.-J., & Cohen, A. (2015). Teacher agency and professional learning: Rethinking fidelity of implementation as multiplicities of enactment: TEACHER ENGAGEMENT AND TEACHER ENACTMENT. *Journal of research in science teaching*, *52*(4), 489-502. https://doi.org/10.1002/tea.21223

Byrne, B. M. (2010). *Structural Equation with AMOS: Basic Concept, Application, and Programming* (Second ed.). Routledge Taylor and Francis Group, LLC.

Cagasan Jr., L., Luo, R., Robertson, P., & Care, E. (2016). *Formative Assessment Project: Phase 2 Research Report, Assessment Curriculum and Technology Research Centre (ACTRC)*. https://actrc.org/wp-content/uploads/2017/04/formative-assessment-phase-2-research-report_v4_lc.pdf

Centre of Assessment and Learning. (2022). *Panduan Pembelajaran dan Asesmen Jenjang Pendidikan Dasar dan Menengah (Guidance of Learning and Assesssment for Primary and Secondari Education)*. Jakarta, Indonesia: Ministry of Education

Chang, M. C., Shaeffer, S., Samarrai, S. A., Ragatz, A. B., Ree, J. d., & Stevenson, R. (2014). *Teacher Reform in Indonesia, The Role of Politics and Evidence in Policy Making*. T. W. Bank.

Charteris, J., & Smardon, D. (2021). Leadership for assessment Capability: Dimensions of Situated leadership Practice for enhanced Sociocultural assessment in Schools.

Leadership and Policy in Schools, 1-13. https://doi.org/10.1080/15700763.2021.1910715

Cherryholmes, C. H. (1992). Notes on Pragmatism and Science Realism. *Educational Researcher*, *21*(6), 13-17.

Cheung, G. W., Cooper-Thomas, H. D., Lau, R. S., & Wang, L. C. (2024). Reporting reliability, convergent and discriminant validity with structural equation modeling: A review and best-practice recommendations. *Asia Pacific Journal of Management*, *41*, 39. https://doi.org/doi.org/10.1007/s10490-023-09871-y

Choi, J. (2017). Understanding elementary teachers' different responses to reform: The case of implementation of an assessment reform in South Korea. *International electronic journal of elementary education*, *9*(3), 581-598.

Christoforidou, M., & Kyriakides, L. (2021). Developing teacher assessment skills: The impact of the dynamic approach to teacher professional development. *Studies in educational evaluation*, *70*, 101051. https://doi.org/10.1016/j.stueduc.2021.101051

Clark, A. K., Nash, B., & Karvonen, M. (2022). Teacher Assessment Literacy: Implications for Diagnostic Assessment Systems. *Applied measurement in education*, *35*(1), 17-32. https://doi.org/10.1080/08957347.2022.2034823

Clark, J. S. (2015). "My assessment didn't seem real": The Influence of Field Experiences on Preservice Teachers' Agency and Assessment Literacy. *Journal of Social Studies Education Research*, *6*(2), 91-111.

Clifton, G. J. (2023). *Language Assessment Literacy and Formative Assessment In Indonesian EFL Education: A Study of Assessment Policy, Teaching Materials, and Teacher Practices under Kurikulum Merdeka* University of Maryland]. Baltimore County, USA.

Cohen, J. (1988). *Statistical Power Analysis for The Behavioral Sciences* (2nd ed.). Hillsdale, N.J. : Lawrence Erlbaum Associates.

Cohen, L., Manion, L., & Morrison, K. (2018). *Research Methods in Education* (Eighth edition. ed., Vol. 1). Routledge. https://doi.org/10.4324/9781315456539

Collins, K. M. T. (2016). Sampling Decisions in Educational Research. In D. Wyse, N. Selwyn, E. Smith, & L. E. Sutter (Eds.), *The BERA/SAGE Handbook of Educational Research* (Vol. 1). SAGE Publication Ltd.

Coombs, A., DeLuca, C., LaPointe-McEwan, D., & Chalas, A. (2018). Changing approaches to classroom assessment: An empirical study across teacher career stages. *Teaching and Teacher Education*, *71*, 134-144. https://doi.org/10.1016/j.tate.2017.12.010

Cortina, J. M., Sheng, Z., Keener, S. K., Keeler, K. R., Grubb, L. K., Schmitt, N., Tonidandel, S., Summerville, K. M., Heggestad, E. D., & Banks, G. C. (2020). From Alpha to Omega and Beyond! A Look at the Past, Present, and (Possible) Future of Psychometric Soundness in the Journal of Applied Psychology. *Journal of Applied Psychology*, *105*(12), 31. https://doi.org/dx.doi.org/10.1037/apl0000815

Creswell, J. W. (2012). *Educational Research, Planning, Conducting, and Evaluating Quantitative and Qualitative Research* (P. A. Smith, Ed. 4th ed.). Pearson Education Inc.

Creswell, J. W. (2014). *RESEARCH DESIGN, Qualitative, Quantitative, and Mixed Method Approaches* (4th ed.). SAGE.

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE.

Crotty, M. (1998). *Foundations of Social Research, Meaning and perspective in the research process*. Routledge. https://doi.org/https://doi.org/10.4324/9781003115700

Dangwal, R., & Kapur, P. (2008). Children's learning processes using unsupervised 'hole in the wall' computers in shared public spaces. *Australasian Journal of Educational Technology*, *24*(3), 339-354.

Darling-Hammond, L., Oakes, J., Wojcikiewicz, S. K., Hyler, M. E., Guha, R., Podolsky, A., Kini, T., Cook-Harvey, C. M., Mercer, C. N. J., & Harrell, A. (2019). *Preparing teachers for deeper learning*. Harvard Education Press.

Day, C., Sammons, P., Hopkins, D., Harris, A., Leithwood, K., Gu, Q., & Brown, E. (2010). *Day et al._10 strong claims about successful school leadership*. https://dera.ioe.ac.uk/id/eprint/2082/1/10-strong-claims-about-successful-school-leadership.pdf

DeLuca, C., & Bellara, A. (2013). The Current State of Assessment Education: Aligning Policy, Standards, and Teacher Education Curriculum. *Journal of teacher education*, *64*(4), 356-372. https://doi.org/10.1177/0022487113488144

DeLuca, C., Chavez, T., & Cao, C. (2013). Establishing a foundation for valid teacher judgement on student learning the role of pre service assessment education. *Assessment in education : principles, policy & practice*, *20*(1). https://doi.org/10.1080/0969594X.2012.668870

DeLuca, C., Coombs, A., McGregor, S., & Rasooli, A. (2019). Toward a Differential and Situated View of Assessment Literacy: Studying Teachers' Responses to Classroom Assessment Scenarios. *Frontiers in Education*, *4*(94). https://doi.org/10.3389/feduc.2019.00094

DeLuca, C., & Klinger, D. A. (2010). Assessment literacy development: identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice*, *17*(4), 419-438. https://doi.org/10.1080/0969594x.2010.516643

DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016a). Approaches to Classroom Assessment Inventory: A New Instrument to Support Teacher Assessment Literacy. *Educational Assessment*, *21*(4), 248-266. https://doi.org/10.1080/10627197.2016.1236677

DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016b). Teacher assessment literacy: a review of international standards and measures. *Educational Assessment, Evaluation and Accountability*, *28*(3), 251-272. https://doi.org/10.1007/s11092-015-9233-6

DeLuca, C., Valiquette, A., Coombs, A., LaPointe-McEwan, D., & Luhanga, U. (2018). Teachers' approaches to classroom assessment: a large-scale survey. *Assessment in Education: Principles, Policy & Practice*, *25*(4), 355-375. https://doi.org/10.1080/0969594x.2016.1244514

Deneen, C. C., Brown, G. T. L., & Parker, D. C. (2016). The impact of conceptions of assessment on assessment literacy in a teacher education program. *Cogent Education*, *3*(1). https://doi.org/10.1080/2331186x.2016.1225380

Denzin, N. K., & Lincoln, Y. S. (2018). *The SAGE Handbook of Qualitative Research* (5th ed.). SAGE Publication, Inc.

Dharmawan, G., & Suryadarma, D. (2021). Education Quality across Indonesia's Districts. *Journal of South East Asia Economies*, *38*(3), 26. https://doi.org/10.1355/ae38-3g

Dunlai, L., & You, S. (2015). An investigation of chinese middle school english teachers' assessment literacy. *Indonesian EFL Journal*, *1*(1).

Edwards, F. (2017). A rubric to track the development of secondary pre-service and novice teachers' summative assessment literacy. *Assessment in education : principles, policy & practice*, *24*(2), 205-227. https://doi.org/10.1080/0969594X.2016.1245651

Efendi, R., Rustaman, N. Y., & Karniawati, I. (2018). Profile of Prospective physics teachers on assessment literacy. *Journal of Physics: Conference Series*, *1013*. https://doi.org/10.1088/1742-6596/1013/1/012047

Febriani, A., Azizah, Y., & Setiawati, M. (2022). Analisis Perubahan Kurikulum Terhadap Hasil Belajar Peserta Didik Kelas X di MAN 1 Solok (*Analysis of curriculum change on the learning outcomes ofyear 10 student at MAN 1 Solok*). *Jurnal Pendidikan dan Ilmu Sosial*, *1*(4), 9. https://jurnal.jomparnd.com/index.php/jp

Ferary, D. (2021). On Ki Hajar Dewantara's philosophy of education. *Nordic Journal of Comparative and International Education*, *5*(2), 14. http://doi.org/10.7577/njcie.4156

Fornell, C., & Larcker, D. F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, *18*(1), 12. https://doi.org/10.2307/3151312

Fowler, F. J. (2014). *Survey research methods* (5th ed.). SAGE.

Fulmer, G. W., Lee, I. C. H., & Tan, K. H. K. (2015). Multi-level model of contextual factors and teachers' assessment practices: an integrative review of research. *Assessment in education : principles, policy & practice*, *22*(4), 475-494. https://doi.org/10.1080/0969594X.2015.1017445

Goretzko, D., Siemund, K., & Sterner, P. (2024). Evaluating Model Fit of Measurement Models in Confirmatory Factor Analysis [Original Research Article]. *Educational and Psychological Measurement*, *84*(1), 22. https://doi.org/10.1177/00131644231163813

Gotch, C. M., & French, B. F. (2014). A Systematic Review of Assessment Literacy Measures. *Educational measurement, issues and practice*, *33*(2), 14-18. https://doi.org/10.1111/emip.12030

Gravetter, F. J., & Wallnau, L. B. (2017). *Statistics for the Behavioral Sciences* (10th ed.). Cengage Learning.

Grissom, J. A., Egalite, A. J., & Lindsay, C. A. (2021). *How Principals Affect Students and Schools: A Systematic Synthesis of Two Decades Research*. http://www.wallacefoundation.org/principalsynthesis.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate Data Analysis* (7th ed.). Pearson.

Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review*, *31*(1), 25. https://doi.org/10.1108/EBR-11-2018-0203

Hair Jr., J. H., Hult, G. T. M., Ringle, C. M., Sarstedt, M., Danks, N. P., & Ray, S. (2021). *Partial Least Squares Structural Equation Modeling (PLS-SEM) Using R, A Workbook*. Springer. https://doi.org/10.1007/978-3-030-80519-7

Halim, A. (2020). The Indonesian curriculum: does it retain culturally responsive teaching? *Journal of English Language and Culture*, *11*(1), 10. https://doi.org/10.30813/jelc.v11i1.2399

Harker, R., Mahar, C., & Wilkes, C. (1990). *An Introduction to the Work of Pierre Bourdieu: The Practice of Theory*. The McMillan Press Ltd.

Hasim, Z., Di, S., & Barnard, R. (2018). Eliciting teachers' understanding and their reported practices on school-based formative assessment: Methodological challenges. *Indonesian Journal of Applied Linguistics*, *8*(1). https://doi.org/10.17509/ijal.v8i1.11476

Heng, T. T., Song, L., & Tan, K. (2021). Understanding the interaction of assessment, learning and context: Insights from Singapore. *Educational research (Windsor)*, *63*(1), 65-79. https://doi.org/10.1080/00131881.2021.1874248

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academic Marketing Science*, *43*, 21. https://doi.org/10.1007/s11747-014-0403-8

Heritage, M. (2010). Formative Assessment and Next Generation of Assessment; Are We Losing Opportunity. In. Washington DC: Council of Chief School State Officers, Washington DC.

Herppich, S., Praetorius, A.-K., Förster, N., Glogger-Frey, I., Karst, K., Leutner, D., Behrmann, L., Böhmer, M., Ufer, S., Klug, J., Hetmanek, A., Ohle, A., Böhmer, I., Karing, C., Kaiser, J., & Südkamp, A. (2018). Teachers' assessment competence: Integrating knowledge-, process-, and product-oriented approaches into a competence-oriented conceptual model. *Teaching and Teacher Education*, *76*, 181-193. https://doi.org/10.1016/j.tate.2017.12.001

Herzog-Punzenberger, B., Altrichter, H., Brown, M., Burns, D., Nortvedt, G. A., Skedsmo, G., Wiese, E., Nayir, F., Fellner, M., McNamara, G., & O'Hara, J. (2020). Teachers responding to cultural diversity: case studies on assessment practices, challenges and experiences in secondary schools in Austria, Ireland, Norway and Turkey. *Educational Assessment, Evaluation and Accountability*, *32*(3), 395-424. https://doi.org/10.1007/s11092-020-09330-y

Hipkins, R., & Cameron, M. I. (2018). *Trends in assessment: An overview of themes in the literature*. https://www.nzcer.org.nz/research/publications/trends-assessment-overview-themes-literature

Hollingworth, L. (2012). Why leadership matters: empowering teachers to implement formative assessment. *Journal of Educational Administration*, *50*(3), 365-379. https://doi.org/10.1108/09578231211223356

Hu, L.-t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modelling: A Multidisciplinary Journal*, *6*(1), 56. https://doi.org/10.1080/10705519909540118

Hunaepi, & Suharta, I. G. P. (2024). Transforming Education in Indonesia: The Impact and Challenges of Merdeka Belajar Curriculum. *Path of Science*, *10*(6), 15. https://doi.org/10.22178/pos.105-31

Irawati, D., Najili, H., Supiani, S., & Zaqiah, Q. Y. (2022). Merdeka Belajar Curriculum Innovation and Its Application in Education Units. *Edumaspul: Jurnal Pendidikan*, *6*(2), 9. https://doi.org/https://doi.org/10.33487/edumaspul.v6i2.4603

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed Method Research; A Research Paradigm Whose Time Has Come. *Educational Researcher*, *33*(7), 14-26.

Kaden, U., & Patterson, P. P. (2014). Changing Assessment Practices of Teaching Candidates and Variables that Facilitate that Change. *Action in teacher education*, *36*(5-6), 406-420. https://doi.org/10.1080/01626620.2014.977700

Kane, M. T. (2013). Validating the Interpretations and Uses of Test Scores. *Journal of educational measurement*, *50*(1), 1-73. https://doi.org/10.1111/jedm.12000

Kanjee, A., & Mthembu, J. (2015). Assessment literacy of foundation phase teachers: An exploratory study. *South African Journal of Childhood Education (SAJCE)*, *5*(1). https://doi.org/https://doi.org/10.4102/sajce.v5i1.354

Kemdikbud.go.id. (2018). *Badan Bahasa Petakan 652 Bahasa di Indonesia*. Kementerian Pendidikan dan Kebudayaan Republik Indonesia. Retrieved 17 February from

Kemdikbud.go.id. (2020). *Daftar Tanya Jawab Kebijakan Ujian Nasional*. Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi. Retrieved 23 February from

Kline, R. B. (2016). *Principles and Practice of Structural Equation Modelling* (5th ed.). The Guilford Press.

Kline, T. J. (2005). *Psychological Testing: A Practical Approach to Design and Evaluation*. SAGE Publications, Inc.
Lee-Perry, M. (2013). *Teacher and Principal Assessment Literacy_disertasi* [Doctoral Thesis, University of Montana].
Lee, M., & Hallinger, P. (2012). National contexts influencing principals' time use and allocation: economic development, societal culture, and educational system. *School Effectiveness and School Improvement*, *23*(4), 18.
Leong, W. S., Ismail, H., Costa, J. S., & Tan, H. B. (2018). Assessment for learning research in East Asian countries. *Studies in educational evaluation*, *59*, 270-277. https://doi.org/10.1016/j.stueduc.2018.09.005
Lincoln, Y. S., Lynham, S. A., & Guba, E. G. (2018). Paradigmatic Controversies, Contradictions, and Emerging Confluences, Revisited. In N. K. Denzin & Y. S. Lincoln (Eds.), *The SAGE Handbook of Qualitative Research* (5th ed.). SAGE Publications, Inc.
Little, T. D., Rhemtulla, M., Gibson, K., & Schoemann, A. M. (2013). Why the Items versus Parcels Controversy Needn't Be One. In N. I. o. Health (Ed.), *Psychol Methods* (pp. 29).
Livingston, K., & Hutchinson, C. (2017). Developing teachers' capacities in assessment through career-long professional learning. *Assessment in education : principles, policy & practice*, *24*(2), 290-307. https://doi.org/10.1080/0969594X.2016.1223016
Looney, A., Cumming, J., van Der Kleij, F., & Harris, K. (2018). Reconceptualising the role of teachers as assessors: teacher assessment identity. *Assessment in education : principles, policy & practice*, *25*(5), 442-467. https://doi.org/10.1080/0969594X.2016.1268090
Louis, K. S., Mayrowetz, D., Smiley, M., & Murphy, J. (2009). The Role of Sensemaking and Trust in Developing Distributed Leadership. In *Distributed Leadership* (pp. 157-180). https://doi.org/10.1007/978-1-4020-9737-9_9
Lysaght, Z., & O'Learry, M. (2017). Scaling up writ small using an assessment learning audit instrument to stimulate site based professional development one school at a time. *Assessment in education : principles, policy & practice*, *24*(2), 271-289. https://doi.org/10.1080/0969594X.2017.1296813
Maisyaroh, Juharyanto, Bafadal, I., Wiyono, B. B., Ariyanti, N. S., Adha, M. A., & Qureshi, M. I. (2021). The principals' efforts in facilitating the freedom to learn by enhancing community participation in indonesia. *Cakrawala pendidikan : CP*, *40*(1), 196-207. https://doi.org/10.21831/cp.v40i1.36119
Mertler, C. A. (2005). Secondary Teachers' Assessment Literacy: Does Classroom Assessment Make a Different? *American Secondary Education*, *33*(2).
Mertler, C. A., & Campbell, C. (2005). *Measuring teacher's knowledge and application of classroom assessment concept; development of ALI* Annual Meeting of American Educational Research Association, Montreal, Quebec, Canada.
Militello, M., Bass, L., Jackson, K., & Wang, Y. (2013). How Data Are Used and Misused in Schools: Perceptions from Teachers and Principals. *Education Sciences*, *3*(2), 98-120. https://doi.org/10.3390/educsci3020098
Ministry of Foreign Affairs. (2018). *Sekilas Tentang Indonesia (A Glance about Indonesia)*. CONSULATE GENERAL OF THE REPUBLIC OF INDONESIA IN FRANKFURT, THE FEDERAL REPUBLIC OF GERMANY. Retrieved 17 February from
MoE. (2016). *Ministry of Education Regulation 23/2016 about Standards of Assessment*. Jakarta, Indonesia: Ministry of Law and Human Rights
MoE. (2021). *Ministry of Education Regulation on National Assessment*. Jakarta, Indonesia: Ministry of Education, Culture, Research, and Technology

MoE. (2022a). *Ministry of Education Regulation on Assessment Standards for Pre-School, Primary, and Secondary Education*. Ministry of Education, Culture, Research, and Technology

MoE. (2022b). *Ministry of Education Regulation regarding Teacher Education Standards*. Jakarta, Indonesia: Legal Bureau, Ministry of Education, Culture, Research, and Technology

Morgan, D. L. (2007). Paradigm Lost & Pragmatic Regained. *Journal of Mixed Methods Research*, *1*(1), 48-76.

Morris, C. E. E. (2017). *Toward Assessment Leadership: Study of Assessment Practice among School and Classroom Leaders* University of Kentucky, USA]. Kentucky, USA.

Morris, R., Perry, T., & Wardle, L. (2021). Formative assessment and feedback for learning in higher education: A systematic review. *Review of Education*, *9*(3), 26. https://doi.org/10.1002/rev3.3292

Nikmard, F., & Mohamadi Zenouzagh, Z. (2020). Designing and validating a potential assessment inventory for assessing ELTs' assessment literacy. *Language Testing in Asia*, *10*(1). https://doi.org/10.1186/s40468-020-00106-1

Nuradhawati. (2019). Dinamika Sentralisasi dan Desentralisasi di Indonesia. *Jurnal Akademia Praja*, *2*(1), 19. https://doi.org/ https://doi.org/10.36859/jap.v2i01.90

Ogan-Bekiroglu, F., & Suzuk, E. (2014). Pre-service teachers' assessment literacy and its implementation into practice. *The Curriculum Journal*, *25*(3), 344-371. https://doi.org/10.1080/09585176.2014.899916

Oo, C. Z., Alonzo, D., Asih, R., Pelobillo, G., Lim, R., San, N. M. H., & O'Neill, S. (2024). Implementing school-based assessment reforms to enhance student learning: a systematic review. *Educational Assessment, Evaluation and Accountability*, *36*(1), 7-30. https://doi.org/10.1007/s11092-023-09420-7

Osses-Vargas, A., Schwantner, U., Adams, R. J., & Wernet, N. (2022). *Relationship between students' economic, cultural, and social status, school climate and student achievement in Indonesia* International Conference on Assessment and Learning (ICAL), Bali, Indonesia.

Pastore, S. (2023). Teacher assessment literacy: a systematic review. *Frontiers in education (Lausanne)*, *8*. https://doi.org/10.3389/feduc.2023.1217167

Pastore, S., & Andrade, H. L. (2019). Teacher assessment literacy: A three-dimensional model. *Teaching and Teacher Education*, *84*, 128-138. https://doi.org/10.1016/j.tate.2019.05.003

Pastore, S., Azzolini, D., Marzadro, S., & Ottaviani, F. (2025). Preparing assessment literate teachers: A cross-national trend analysis. *Studies in educational evaluation*, *86*, 101485. https://doi.org/10.1016/j.stueduc.2025.101485

Patton, M. Q. (2015). *Qualitative research & evaluation methods / Michael Quinn Patton* (V. Knight, K. Bierach, Y. McDufee, & D. C. Felts, Eds. 4th edition ed.). SAGE Publication Inc.

Pfeiffer-Hoens, M. (2017). *Investigation of School Leader and Teacher Level of Assessment Literacy* [Doctoral, Middle Tennessee State University].

Plake, B. S., Impara, J. C., & Fager, J. J. (1993). Assessment Competencies of Teachers: A National Survey. *Educational Measurement: Issues and Practice*, *12*(4), 10-39.

Popham, W. J. (2011). Assessment Literacy Overlooked: A Teacher Educator's Confession. *The Teacher Educator*, *46*(4), 265-273. https://doi.org/10.1080/08878730.2011.605048

Popham, W. J. (2014). The right test for the wrong reason. *Phi Delta Kappan*, *96*(1), 46-52. https://doi.org/10.1177/0031721714547862

Poskitt, J. (2014). Transforming professional learning and practice in assessment for learning. *The Curriculum Journal*, *25*(4), 542-566. https://doi.org/10.1080/09585176.2014.981557

Pritchard, A., & Woollard, J. (2013). *Psychology for the classroom: Constructivism and social learning*. Routledge.

Prizovskaya, V. (2018). Identifying New Jersey Teachers' Assessment Literacy as Precondition for Implementing Student Growth Objectives. *Journal of Education and Learning*, *7*(1), 184-207. http://nottingham.idm.oclc.org/login?url=https://www.proquest.com/scholarly-journals/identifying-new-jersey-teachers-assessment/docview/2011265881

Puad, L. M. A. Z., & Ashton, K. (2020). Teachers' views on classroom-based assessment: an exploratory study at an Islamic boarding school in Indonesia. *Asia Pacific Journal of Education*, *41*(2), 253-265. https://doi.org/10.1080/02188791.2020.1761775

Purnamasari, R., Safitri, N., & Kurnia, D. (2023). Literacy and numeracy item test development of Minimum Assessment Competency test for year five students. *Research and Learning in Elementary Education*, *7*(1), 11. https://jbasic.org/index.php/basicedu

Putra, I. P. (2023). *153,621 More Schools are Applying for the New Curriculum in 2023*. Medcom.id. Retrieved 28 March 2025 from https://www.medcom.id/pendidikan/news-pendidikan/eN4L2x3N-153-621-sekolah-daftar-implementasi-kurikulum-merdeka-di-tahun-ajaran-2023-2024

Randel, B., Apthorp, H., Beesley, A. D., Clark, T. F., & Wang, X. (2016). Impacts of professional development in classroom assessment on teacher and student outcomes. *the Journal of Educational Research*, *109*(5), 491-502. https://doi.org/10.1080/00220671.2014.992581

Rasyidah, U., Triana, N., & Saukah, A. (2020). The Teachers' Assessment Knowledge and Practice; Contribution of Past-Time Experience and Present-Time Decision. *The Qualitative Report*, *25*(7), 1738-1753.

Renihan, P., & Noonan, B. (2018). Principals as Assessment Leaders in Rural Schools. *Rural Educator*, *33*(3).

Rezagah, K. (2022). Iranian EFL teachers' assessment literacy knowledge: The impact of the educational system on teachers' classroom assessment practices. *International Journal of Language Testing*, *12*(2).

Rizaty, M. A. (2023). *The Ministry of Education Recorded that There are 436.707 in Indonesia in 2023/2024 Academic Year*. DataIndonesia.id. Retrieved 28 March 2025 from https://dataindonesia.id/pendidikan/detail/kemendikbud-catat-436707-sekolah-di-indonesia-pada-20232024

Rogers, A. P., Reagen, E. M., & Ward, C. (2020). Preservice teacher performance assessment and novice teacher assessment literacy. *Teaching Education*. https://doi.org/10.1080/10476210.2020.1840544

Rogers, C. R. (1969). *Freedom to Learn*. Charles E. Merril Publishing Company.

Rosas, R. (2014). *Elementary School Teachers' and Principals' Formative Assessment Beliefs, Practices, and Assessment Literacy* California State University]. California, United States of America.

Rubiyantoro, Y. (2024). *Leadership in Early Childhood Education: Case Studies of Indonesian Public Kindergartens* University of Nottingham]. Nottingham.

Rukoyah, Hermita, N., Vebrianto, R., Mujtahid, I. M., Sulistyo, U., & Samsudin, A. (2023). Reflection of Indonesian Educators on the Implementation of the Merdeka Curriculum. *Mimbar Sekolah Dasar*, *10*(3), 17. https://doi.org/https://doi.org/10.53400/mimbar-sd.v10i3.64864

Saldana, J. (2016). *The Coding Manual for Qualitative Researcher* (J. Seaman, Ed. 3rd ed.). SAGE Publication Inc.

Saputra, E., Hamied, F. A., & Suherdi, D. (2020). The development of beliefs and practices of language assessment literacy: does a professional learning community help? *Journal of Education for Teaching*, *46*(3), 414-416. https://doi.org/10.1080/02607476.2020.1761250

Sardareh, S. A. (2016). Formative Feedback in a Malaysian Primary School ESL Context. *Malaysian online journal of educational sciences*, *4*(1), 1.

Schneider, C., & Bodensohn, R. (2017). Student teachers appraisal of the importance of assessment in teacher education and self reports on the development of assessment competence.pdf>. *Assessment in education : principles, policy & practice*, 127-146. https://doi.org/10.1080/0969594X.2017.1293002

Schneider, C., DeLuca, C., Pozas, M., & Coombs, A. (2020). Linking personality to teachers' literacy in classroom assessment: a cross-cultural study. *Educational research and evaluation*, *26*(1-2), 53-74. https://doi.org/10.1080/13803611.2021.1902354

Scott, S. (2016). Reconceptualising Instructional Leadership: Exploring the Relationships Between Leadership, Instructional Design, Assessment, and Student Needs. In S. Scott, D. E. Scott, & C. F. Webber (Eds.), *Leadership of Assessment, Inclusion, and Learning* (Vol. 3). Springer.

Smith, K., & Engelsen, K. S. (2013). Developing an assessment for learning (AfL) culture in school: the voice of the principals. *International Journal of Leadership in Education*, *16*(1), 106-125. https://doi.org/10.1080/13603124.2011.651161

Squire, P. J. (2010). Student Assessment Procedure in Bostwana_Squire. *Current Research Journal of Social Sciences 2*(6), 296-300.

Steedle, J. T. (2016). Descriptive Statistics. In D. Wyse, N. Selwyn, E. Smith, & L. E. Suter (Eds.), *The BERA/SAGE Handbook of Educational Research* (Vol. 1). SAGE Publication Ltd.

Stiggins, R. J. (2001). The Principal's Leadership Role in Assessment. *NASSP Bulletin*, *85*(621), 13-26.

Stiggins, R. J., & Duke, D. (2008). Effective Instructional Leadership Requires Assessment Leadership. *Phi Delta Kappan*, *90*(4), 285-291. https://doi.org/10.1177/003172170809000410

Sullivan, A., & Calderwood, L. (2016). Longitudinal, Cross-Sectional, and Trend Studies. In D. Wyse, N. Selwyn, E. Smith, & L. E. Suter (Eds.), *The BERA/SAGE Handbook of Educational Research* (Vol. 1). SAGE Publication Ltd.

Tay, H. Y., & Tan, K. (2019). Assessment Leadership. In B. Wong, S. Hairon, & P. T. Ng (Eds.), *School Leadership and Educational Change in Singapore*. Springer Nature Switzerland AG 2019. https://doi.org/https://doi.org/10.1007/978-3-319-74746-0

Tayag, J., & Ayuyao, N. (2020). Exploring the relationship between school leadership and teacher professional learning through structural equation modeling. *International Journal of Educational Management*, *34*(8), 16. https://doi.org/10.1108/IJEM-11-2018-0372

Teddlie, C. B., & Tashakkori, A. (2010). Overview of Contemporary Issues in Mixed Methods Research. In A. Tashakkory & C. Teddlie (Eds.), *SAGE Handbook of Mixed Methods in Social & Behavioral Research* (2 ed., pp. 1). SAGE Publications, Inc. https://doi.org/10.4135/9781506335193.n1

Umam, A., & Indah, Y. A. (2020). Exploring In-Service TEYL Teachers' Assessment Literacy: Implication For Continuing Professional Development. *JEES (Journal of English Educators Society)*, *5*(1), 47-51. https://doi.org/10.21070/jees.v5i1.364

Utomo, A. (2024). *Laying the Foundations of Systemic Change in Education: Reflections on Indonesia's Merdeka Belajar Reform* World Bank Education Global Practice seminar, Washington D.C.

Volante, L., & Fazio, X. (2007). Exploring Teacher Candidates' Assessment Literacy: Implications for Teacher Education Reform and Professional Development [Abstract]. *Canadian journal of education*, *30*(3), 749-770. https://doi.org/10.2307/20466661

Vygotsky, L. S. (1978). *Mind in Society: Development of Higher Psychological Processes* (1 ed.). Harvard University Press.

Warwick, P., Shaw, S., & Johnson, M. (2015). Assessment for Learning in International Contexts: exploring shared and divergent dimensions in teacher values and practices. *Curriculum journal (London, England)*, *26*(1), 39-69. https://doi.org/10.1080/09585176.2014.975732

Webber, C. F., Scott, S., Aitken, E. N., Lupart, J., & Scott, D. E. (2013). Leading assessment for enhanced student outcomes. *School Leadership & Management*, *33*(3), 240-255. https://doi.org/10.1080/13632434.2013.773885

Wiese, E., & Nortvedt, G. A. (2023). Teacher assessment literacy in culturally and linguistically diverse classrooms: A Norwegian case study. *Teaching and Teacher Education*, *135*, 104357. https://doi.org/10.1016/j.tate.2023.104357

Willis, J., Adie, L., & Klenowski, V. (2013). Conceptualising teachers' assessment literacies in an era of curriculum and assessment reform. *The Australian Educational Researcher*, *40*(2), 241-256. https://doi.org/10.1007/s13384-013-0089-9

Wyatt-Smith, C., Alexander, C., Fishburn, D., & McMahon, P. (2017). Standards of practice to standards of evidence: developing assessment capable teachers. *Assessment in education : principles, policy & practice*, *24*(2), 250-270. https://doi.org/10.1080/0969594X.2016.1228603

Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, *58*, 149-162. https://doi.org/10.1016/j.tate.2016.05.010

Yamtim, V., & Wongwanich, S. (2014). A Study of Classroom Assessment Literacy of Primary School Teachers. *Procedia - Social and Behavioral Sciences*, *116*, 2998-3004. https://doi.org/10.1016/j.sbspro.2014.01.696

Yan, X., Zhang, C., & Fan, J. J. (2018). "Assessment knowledge is important, but …": How contextual and experiential factors mediate assessment practice and training needs of language teachers. *System (Linköping)*, *74*, 158-168. https://doi.org/10.1016/j.system.2018.03.003

Ye, W. (2023). Urban public-school teachers' assessment literacy in China's Zhengzhou city. *Education 3-13*, *51*(5), 836-849. https://doi.org/10.1080/03004279.2021.2025128

Yin, R. K. (2018). *Case Study Reseach and Application: Design and Methods* (6th ed.). SAGE Publication Inc.

Ysenbaert, J., Van Houtte, M., & Van Avermaet, P. (2020). Assessment policies and practices in contexts of diversity: unravelling the tensions. *Educational Assessment, Evaluation and Accountability*, *32*(2), 107-126. https://doi.org/10.1007/s11092-020-09319-7

Zhang, Z., & Burry-Stock, J. A. (2003). Classroom Assessment Practices and Teachers' Self-Perceived Assessment Skills. *Applied measurement in education*, *16*(4), 323-342. https://doi.org/10.1207/S15324818AME1604_4

Zolfaghari, F., Ahmadi, A., & Hui, S. K. F. (2016). Assessment literacy components across subject matters. *Cogent Education*, *3*(1). https://doi.org/10.1080/2331186x.2016.1252561

Zolfaghari, S., & Ashraf, H. (2015). The Relationship between EFL Teachers' Assessment Literacy, Their Teaching Experience, and Their Age: A Case of Iranian EFL Teachers. *Theory and Practice in Language Studies*, *5*(12). https://doi.org/10.17507/tpls.0512.16

Zuhdi, M. (2015). Pedagogical Practices in Indonesia. In E. H.-F. Law & U. Miura (Eds.), *Transforming Teaching and Learning in Asia and Pacific, Case Studies from Seven Countries* (pp. 142-160). UNESCO.

Zulaiha, S., Mulyono, H., & Ambarsari, L. (2020). An Investigation into EFL Teachers' AssLit; Indo Teachers' Perceptions and Classroom Practice. *European Journal of Contemporary Education*, *9*(1), 189-201. https://doi.org/10.13187/ejced.2020.1.189

Appendix

Appendix 1: Survey questionnaire for teachers

Social Studies Teachers' Assessment Literacy Questionnaire

Thank you for choosing to participate in this survey. Please answer the following questions to the best of your knowledge. Your candid and thoughtful responses will be greatly appreciated. Your identity and responses will be kept completely confidential and will not affect you in any way.

Please tick (√) the response options or answer the questions that apply to you.

1. Background Information

A.	**Demographic Information**									
1.	Gender	O Male	O Female							
2.	Age	…… years								
3.	Province	O Riau			O East java			O East Kalimantan		
		O South Sulawesi			O Papua			O other: ……		
4.	Regency/Municipality	……								
5.	Degree attained	O Doctor			O Masters			O Bachelor		
6.	Teaching experience (year)	O	O	O	O	O	O	O	O	O	O
		0-3	>3-6	>6-9	>9-12	>12-15	>15-18	>18-21	>21-24	>24-27	>27
7.	Assessment instructor experience (year)	O	O	O	O	O	O	O	O	O	O
		0-1x	2-3x	4-5x	6-7x	8-9x	10-11x	12-13x	14-15x	16-17x	>17x
8.	Test items developer experience (year)	O	O	O	O	O	O	O	O	O	O
		0-1x	2-3x	4-5x	6-7x	8-9x	10-11x	12-13x	14-15x	16-17x	>17x

B.	**Assessment Education and Training**									
1.	**Assessment-related courses in HE**									
	a. Total credit	O	O	O	O	O	O	O	O	O	O

	0-6	7-12	13-18	19-25	26-32	33-39	40-46	47-53	54-60	>60
b. Relevance with current assessment responsibilities	O 1	O 2	O 3	O 4	O 5	O 6		O 7	O 8	O 9	O 10
	highly irrelevant ⇐					⇒ highly relevant				
c. Effectiveness of the course(s)	O 1	O 2	O 3	O 4	O 5	O 6		O 7	O 8	O 9	O 10
	highly ineffective ⇐					⇒ highly effective				

2. Assessment-related Professional Development (PD)										
a. Total hours (1 day = 8 hours)	O 0-8	O 9-17	O 17-25	O 26-34	O 35-43	O 44-52	O 53-61	O 62-70	O 71-79	O > 79
b. Relevance with current assessment responsibilities	O 1	O 2	O 3	O 4	O 5	O 6	O 7	O 8	O 9	O 10
	highly irrelevant ⇐						⇒ highly relevant			
c. Effectiveness of PD	O 1	O 2	O 3	O 4	O 5	O 6	O 7	O 8	O 9	O 10
	highly ineffective ⇐						⇒ highly effective			
3. Local assessment-related support										
a. Support from school colleagues	O 1	O 2	O 3	O 4	O 5	O 6	O 7	O 8	O 9	O 10
	never ⇐						⇒ always			
b. Support from subject professional community	O 1	O 2	O 3	O 4	O 5	O 6	O 7	O 8	O 9	O 10
	never ⇐						⇒ always			
c. Support from school principal	O 1	O 2	O 3	O 4	O 5	O 6	O 7	O 8	O 9	O 10
	never ⇐						⇒ always			

2. Assessment Literacy Inventory

Please tick one correct/best answer on each of the questions below.

Scenario 1:
Mr. Adi, a year 8 Social Studies teacher, has just finished teaching a unit about Production, Consumption, and Distribution of Goods and Services. As a part of his

teaching activities, he needs to conduct assessment of his students' learning on the unit.

a. Mr. Adi wants to see whether his students correctly understand the concept of factors of production and gives a quiz at the end of sub-unit Production. The purpose of this assessment is …

- ☐ Summative
- ☐ Formative
- ☐ Diagnostic

b. To examine students' comprehension about the relationship between the concept of consumption and the concept of scarcity, Mr. Adi asks his students to conduct small scale research in their local environment and write a short report about the situation which reflect the relationship. To give his students consistent and accurate grade, Mr. Adi is advised to ….

- ☐ develop a scoring rubric based on the students' level of understanding
- ☐ ask his colleagues about the criteria that has been used in their classrooms
- ☐ use assessment criteria provided in teacher's guidebook

c. Mr. Adi conducts an end-of-unit test to the class and most students score 80 or above in the scale of 100. It can be inferred from the results that …

- ☐ 80% or more of the students answer the questions correctly
- ☐ most students answer the test items 80% or more correctly
- ☐ most students understand 80% or more of the content in the unit

d. In a project-based assessment on concept of production factors, Mr. Adi found a few students could not differentiate between the concept of capital and fund. To improve the students' learning in that area, the best strategy Mr. Adi can use in communicating the result of the project is ….

- ☐ informing students what to expect before the project begins.
- ☐ describing the area(s) students have problem and suggesting improvement.
- ☐ giving fair scores according to what students have done in the project.

e. Mr. Adi administers end-of-unit multiple choice test with 25 items. The results can be said valid when he …

- ☐ writes or chooses test items drawn from material he has taught in class
- ☐ uses test items at the end of unit from commercially published student book
- ☐ uses test items from standardized test item bank related to the unit

f. Following are examples of assessment fairness when Mr. Adi asks his students to write an essay on a topic of distribution of goods and services in digital era, EXCEPT ….

- ☐ Explaining the assessment criteria that will be used to judge the essay
- ☐ Using rubric consistently when assessing all the students' essays
- ☐ Considering students' behaviour during the lesson when giving scores

Scenario 2:

This semester, Ms. Ida will teach a unit of Social Interaction to her year nine students. She wants her students to be able to apply what they learn in class to their everyday lives instead of merely memorise concepts and facts. Therefore, she decides to use more project-based and problem-based learning approaches throughout the semester.

a. A three-week-end-of-semester project about types of social interaction will be used as a part of final mark which reflect students' achievement toward pre-determined learning targets. Ms. Ida uses this assessment for ….

- ☐ summative purpose
- ☐ formative purpose
- ☐ diagnostic purpose

b. The following steps are necessary for Ms. Ida to ensure that student assessment in social studies are consistent, fair, and using shared criteria across year 9 social studies parallel classes.

- ☐ Ms. Ida develops her assessment plan and make a report to the school principal.
- ☐ Ms. Ida talks to other year 9 social studies teachers to achieve agreement regarding assessment methods, scoring, and criteria.
- ☐ Ms. Ida creates scoring rubric for assessing students' project and choose relevant test items from test item bank.

c. One of the learning targets of the unit, which assesses students' higher order thinking skills, states 'students are able to attribute correct message from a passage about land transformation from farmland into industrial land'. The validity of a true-false test item in assessing this learning target is low due to …

- ☐ The probability to get correct answer by guessing is high.
- ☐ True-false item only measures low order thinking skills.
- ☐ The question will be too easy to answer

d. After an end-of-unit summative test, Ms. Ida announces that the test result of the whole class will be printed out and stick on the board in front of the class tomorrow morning before the earliest regular classroom lesson schedule because Ms. Ida will

have to attend a professional development program early in the morning. What ethical issue arise from such situation?

- ☐ Ms. Ida action is considered in-appropriate because it is lack of formality.
- ☐ Students that absent at the day will not be able to see their test result.
- ☐ Mr. Ida does not keep students' privacy right regarding the test results.

Scenario 3:

Ms. Ani is teaching a social studies unit 'Space and Spatial Interconnectivity' to her year 8 students. The latest assessment policy from the Government mandates that teachers should emphasize Assessment for Learning (AfL) approaches in their assessment practices and use various assessment methods to increase students' learning outcomes.

a. Which of the following assessment practice example that is NOT the recommended assessment practice?

- ☐ Ms. Ani uses self-assessment and peer-assessment more frequently to develop students' ownership in assessment.
- ☐ Ms. Ani writes detailed description on students' assignments when giving written feedback to the students individually.
- ☐ Ms. Ani organize an extra one-hour lesson after school to drill students with standardized test taken from item bank.

b. One of the learning targets in the unit is 'student can analyse the positive and negative impact of spatial interconnectivity'. If Ms. Ani wants to use a multiple-choice item to measure this learning target, she should …

- ☐ uses more homogenous options.
- ☐ uses a stimulus before the stem of the item
- ☐ writes a long stem for the item.

c. Another learning targets in the unit is 'students can criticise an example of a solution to a social problem caused by migration based on certain criteria'. According to cognitive taxonomy of Anderson et al. (Bloom revised version), which cognitive level Ms. Ani will assess for the learning target?

- ☐ Understanding
- ☐ Analysing
- ☐ Evaluating

d. After a series of quizzes, essays, and a mid-semester test, Ms. Ani found that 4 of her students are far behind compared to other students. Which of Ms. Ani's following action that reflects the practice of fairness principle in Assessment for Learning in communicating the assessment results to students?

- ☐ Tell the students about the results, discuss about their strengths and weaknesses, and adjust the learning target.
- ☐ Inform the parents about the assessment results and suggest for private lesson for the students to improve learning.
- ☐ Tell the students about the assessment results and ask them to do more practice using a compilation of standardized test items.

e. Ms. Ani and 2 of her colleagues are going to write parallel but different end-of-unit-test for year eight social studies classes. What should they do to increase the reliability of the test?

- ☐ The teachers should develop and use the same item specification.
- ☐ The teachers should use items from standardized test item bank.
- ☐ Each of the teachers should modify test items from the student book.

f. Which of the following action is considered lack of ethical responsibility in conducting assessment of the unit?

- ☐ Randomly selecting test items from test items at the end of unit.
- ☐ Conducting a test without informing the students in advanced.
- ☐ Discussing with colleagues about students' learning difficulties anonymously.

Scenario 4:

Mr. Ali teaches year seven social studies class (junior secondary education: year 7-9), whose students are just graduated from primary schools (year 1-6). To identify the students' prior knowledge, strengths, and weaknesses, he administers a multiple-choice pre-test at the first day of the class after some fun introduction activities at the beginning of the class. He will use the test results to develop lesson plans which meet the students' level of competences and learning needs.

a. What is the purpose of the test Mr. Ali conducts?

- ☐ Formative
- ☐ Diagnostic
- ☐ Summative

b. With such purpose as described in the scenario 4, what level of cognitive skills will Mr. Ali most likely measure through the multiple-choice test?

- ☐ Remembering and understanding
- ☐ Applying and analysing
- ☐ Evaluating and creating

c. Mr. Ali wants to use the test next year for new students. Therefore, he used the test results to analyse the test items. The result shows that the average difficulty of the items is 0.14. What does this number imply?

- ☐ The test is easy
- ☐ The test is valid
- ☐ The test is difficult

d. After getting initial data regarding students existing competences, strengths, and weaknesses, Mr. Ali develops lesson plans for the semester accordingly. At the end of the semester, Mr. Ali needs to communicate students learning progress with their parents. Which of the following action should be AVOIDED in writing the report?

- ☐ Attaching evidence to support the written report.
- ☐ Using scientific and complex sentences to give detail information.
- ☐ Combining numbers and description of learning progress.

e. Which of the following is considered lack of ethical responsibility action, in using the results from the multiple-choice test at the beginning of semester?

- ☐ Mr. Ali put the list of members of groups using labels of “Low”, “Middle”, and “High” in front of the class.
- ☐ Mr. Ali kept the list of the group for himself and regularly mixes the groups as a peer-tutor strategy.
- ☐ Mr. Ali shared the test result with his colleagues and ask for their experiences in dealing with various level of competences.

Scenario 5:

Ms. Ica is a social studies teacher in an urban-industrial area school. Approximately 30% of the students are coming from other provinces and islands who have different social and cultural background. This semester, the students are studying about Hindu and Buddha cultural heritage from 5-7 BC which the local students are more familiar with. Since Government promotes contextual teaching and learning approaches, Ms. Ica used many examples from local context to teach the topics. In the middle of semester, Ms. Ica just realized that the non-local students are left behind compared to the locals.

a. The following assessment methods are appropriate in responding to both the students diverse social-cultural background and the mandated assessment for learning approach EXCEPT ...

- ☐ Ms. Ica uses group project assessment, in which all the groups comprise local and non-local students.
- ☐ Ms. Ica asks the non-local students practice more standardized tests related to local culture topics after school hours.
- ☐ Ms. Ica gives essay assignment which have 5-6 alternative issues which the students may choose the one that suit them the best.

b. Ms. Ani wants her students to appreciate the values of their traditional cultural heritage. Which of the following test item will best measure the learning target?

- ☐ The biggest Hindu temple in Indonesia is ...
  a. Borobudur
  b. Prambanan
  c. Singhasari
  d. Penataran
- ☐ Identify the theme of the stories on the top level of Borobudur temple.
- ☐ Describe how your local cultural values contribute to the society.

c. The following action is in-line with the concept of Assessment for Learning (AfL) in giving feedback to students.

- ☐ Increasing students' motivation by giving short praises such as "good", "well-done", "don't give up".
- ☐ Informing students, the scores of every assessment the teacher have administered.
- ☐ Giving written rich description of students learning progress and areas to improve.

d. The action below can increase the validity of Ms. Ica's end-of-semester student evaluation.

- ☐ Ms. Ica combines mid-semester and end-of-semester multiple-choice tests as the final mark for students.
- ☐ Ms. Ica uses two different groups of learning targets for the two groups of students.
- ☐ Ms. Ica increases the multiple-choice items in the end-of-semester test from 30 to 40 items.

Prize draw and availability for interview session

1.	Are you interested to join a prize draw? (Gift vouchers worth £8 each are available for 25% of lucky respondents)	☐ yes	☐ no	If yes, please provide your email address:
2.	Are you willing to participate in the interview session regarding assessment enactment? (The interview will last for about 1 hour and conducted either face to face or online at your convenient time. Each interviewee will receive a small internet/transport fund £13)	☐ yes	☐ no	If yes, please provide your email address:

Appendix 2: Survey questionnaire for principal

Principals' Assessment Literacy Questionnaire

Thank you for choosing to participate in this survey. Please answer the following questions to the best of your knowledge. Your candid and thoughtful responses will be greatly appreciated. Your identity and responses will be kept completely confidential and will not affect you in any way.

Please tick (√) the response options or answer the questions that apply to you.

3. Background Information

A.	**Demographic Information**									
9.	Gender	O Male	O Female							
10.	Age	…… years								
11.	Province	O Riau			O East java		O East Kalimantan			
		O South Sulawesi			O West Papua		O other: ……			
12.	Regency/Municipality	……								
13.	Degree attained	O Doctorate			O Masters		O Bachelor			
14.	Teaching experience (year)	O	O	O	O	O	O	O	O	O	O
		0-3	>3-6	>6-9	>9-12	>12-15	>15-18	>18-21	>21-24	>24-27	>27
15.	Assessment instructor experience (year)	O	O	O	O	O	O	O	O	O	O
		0-1x	2-3x	4-5x	6-7x	8-9x	10-11x	12-13x	14-15x	16-17x	>17x
16.	Test items developer experience (year)	O	O	O	O	O	O	O	O	O	O
		0-1x	2-3x	4-5x	6-7x	8-9x	10-11x	12-13x	14-15x	16-17x	>17x

B.	**Assessment Education and Training**									
1.	**Assessment-related courses in HE**									
	d. Total credit	O	O	O	O	O	O	O	O	O	O
		0-6	7-12	13-18	19-25	26-32	33-39	40-46	47-53	54-60	>60
		O	O	O	O	O	O	O	O	O	O

e.	Relevance with current assessment responsibilities	1 2 3 4 5 6 7 8 9 10 highly irrelevant ⇐ ⇒ highly relevant
f.	Effectiveness of the course(s)	O O O O O O O O O O 1 2 3 4 5 6 7 8 9 10 highly ineffective ⇐ ⇒ highly effective

2.	**Assessment-related Professional Development (PD)**	
d.	Total hours	O O O O O O O O O O 0-8 9-17 17-25 26-34 35-43 44-52 53-61 62-70 71-79 > 79
e.	Relevance with current assessment responsibilities	O O O O O O O O O O 1 2 3 4 5 6 7 8 9 10 highly irrelevant ⇐ ⇒ highly relevant
f.	Effectiveness of PD	O O O O O O O O O O 1 2 3 4 5 6 7 8 9 10 highly ineffective ⇐ ⇒ highly effective

3.	**Local assessment-related support**	
d.	Support from school colleagues	O O O O O O O O O O 1 2 3 4 5 6 7 8 9 10 never ⇐ ⇒ always
e.	Support from principal community	O O O O O O O O O O 1 2 3 4 5 6 7 8 9 10 never ⇐ ⇒ always
f.	Support from school superintendent	O O O O O O O O O O 1 2 3 4 5 6 7 8 9 10 never ⇐ ⇒ always

4. Assessment Literacy Inventory

Please tick one correct/best answer on each of the questions below.

Part 1. (Purpose of Assessment)

1. Mr. Adi, a year 8 Social Studies teacher, wants to see whether his students correctly understand the concept of factors of production by giving a quiz at the end of sub-unit Production. The purpose of this assessment is …

- ☐ Summative
- ☐ Formative
- ☐ Diagnostic

2. A three-week-project with a topic "energy around us" will be used as a part of final mark which reflect students' achievement toward pre-determined learning targets. Ms. Ida uses this assessment for ….
   - ☐ summative purpose
   - ☐ formative purpose
   - ☐ diagnostic purpose

3. Ms. Ainy teaches year seven Math class (junior secondary education: year 7-9), whose students are just graduated from primary schools (year 1-6). To identify the students' prior knowledge, strengths, and weaknesses, she administers a multiple-choice test at the first day of the class after some fun introduction activities at the beginning of the class. She will use the test results to develop lesson plans which meet the students' level of competences and learning needs. What is the purpose of the test Ms. Ainy conduct?
   - ☐ Formative
   - ☐ Diagnostic
   - ☐ Summative

4. Ms. Ica, an Indonesian Language teacher, analysed most of assessment results she has administered up to the mid-semester and use the results to revise the lesson plans that she has developed at the beginning of the semester. The purpose of Ms. Ica's action is …
   - ☐ Formative
   - ☐ Diagnostic
   - ☐ Summative

<u>Part 2. (Supporting teachers to improve assessment literacy)</u>

Mr. Andi, an junior secondary school principal, wants to evaluate and assist his teachers on how they enact the assessment for learning principles in their teaching practices. To do this he initially analyses lesson plans, observes classrooms, discusses with teachers, and asks some of the students.

5. In an effort to provide effective professional development programs for his teachers, accurately addressing the teachers' lack of assessment knowledge and skills, M. Andi can use the following strategies to identify teachers training needs, EXCEPT …
   - ☐ Analyse teachers' lesson plans
   - ☐ Organise reflective dialogue with teachers

- ☐ Observe teaching practice in one classroom

6. Mr. Andi is having a discussion with Mr. Joko on the steps in his lesson plan of concepts of location and interconnectivity to year 7 students. Which of the following suggestions given by Mr. Andi indicate assessment for learning principles?

   - ☐ The teacher forms discussion groups comprising students with various competence levels and asks to share their knowledge to each other.
   - ☐ The teacher asks the students about what they have been doing during the weekend to make a comfort start.
   - ☐ The teacher gives test practices to prepare the students for the next month end-of-year standardized test.

7. When looking at a teacher's lesson plans, the following things are the initial indicators that the teacher is implementing the assessment for learning principles EXCEPT ...

   - ☐ various assessment methods and techniques
   - ☐ the use of rubric in assessing essays
   - ☐ instruments for student self-assessment

8. Which step in the teacher's lesson plan shows that the teacher encourages students to be the owner of their own learning?

   - ☐ The teacher develops a scoring rubric based on the students' level of understanding.
   - ☐ The teacher asks the students to form group discussion.
   - ☐ The teacher shares learning targets and discusses success criteria with students.

9. After several discussion with teachers, Mr. Andi found out that when conducting end-of semester tests, many teachers use directly test items from commercially published student books or collection of national exam items books without matching the items with what have been taught during the semester. What topic in professional learning community meeting(s) should be included to address the problem?

   - ☐ Validity of assessment.
   - ☐ Interpreting assessment results.
   - ☐ Test item analysis.

<u>Part 3. (Assessment Communication)</u>

10. In a project-based assessment on concept of production factors, Mrs. Reny found a few students could not differentiate between the concept of capital and fund. To improve the students' learning in that area, the best strategy Mrs. Reny can use in communicating the assessment results is ….

- ☐ informing students what to expect before the project begins.
- ☐ describing the area(s) students have problem and suggesting improvement.
- ☐ giving fair scores according to what students have done in the project.

11. At the end of a semester, Mr. Yudi needs to communicate students' learning progress with their parents. Which of the following action should be AVOIDED in writing the report?
    - ☐ Attaching evidence to support the written report.
    - ☐ Using scientific and complex sentences to give detail information.
    - ☐ Combining numbers and description of learning progress.

12. After a series of quizzes, essays, and a mid-semester test, Mr. Anto found that 4 of her students are far behind compared to other students. Which of Mr. Anto's following action reflects the practice of Assessment for Learning in communicating the assessment results to students?
    - ☐ Tell the students about the results, discuss about their strengths and weaknesses, and adjust the learning target if necessary.
    - ☐ Inform the parents about the assessment results and suggest for private lesson for the students to improve learning.
    - ☐ Tell the students about the assessment results and ask them to do more practice using a compilation of standardized test items.

13. The following action is in-line with the concept of Assessment for Learning (AfL) in giving feedback to students.
    - ☐ Increasing students' motivation by giving short praises such as "good", "don't give up".
    - ☐ Informing students, the scores of every assessment the teacher have administered.
    - ☐ Giving written rich description of students learning progress and areas to improve.

<u>Part 4. (Assessment Fairness and Ethics)</u>

14. After an end-of-unit summative test, Ms. Wida announces that the test result of the whole class will be printed out and stick on the board in front of the class the next morning before the earliest regular classroom lesson schedule because Ms. Wida will have to attend a professional development program early in the morning. What ethical issue arise from such situation?
    - ☐ Ms. Wida action is considered in-appropriate because it is lack of formality.
    - ☐ Students that absent at the day will not be able to see their test result.
    - ☐ Mr. Wida does not keep students' privacy right regarding the test results.

15. Which of the following action is considered lack of ethical responsibility in conducting assessment of the unit?

- ☐ Randomly selecting test items from test items at the end of unit.
- ☐ Conducting a test without informing the students in advanced.
- ☐ Discussing with colleagues about students' learning difficulties anonymously.

16. Which of the following is considered lack of ethical responsibility action, in using the results from the multiple-choice test to map students' initial competences at the beginning of semester?

- ☐ A teacher puts the list of members of groups using labels of "Low", "Middle", and "High" in front of the class.
- ☐ A teacher kept the list of the group for himself and regularly mixes the groups as a peer-tutor strategy.
- ☐ A teacher shared the test result with his colleagues and ask for their experiences in dealing with various level of competences.

17. A school located in an urban area comprises of approximately 30% students coming from other provinces, who have different social and cultural background. Mr. Fauzan, a social studies teacher is teaching a topic of Hindu and Buddha cultural heritage from 5-7 Century which the local students are more familiar with. Mr. Fauzan used many examples from local context to teach the topics to apply contextual learning approach. In the middle of semester, Mr. Fauzan just realized that the non-local students are left behind compared to the locals.

    The following assessment methods are appropriate in responding to both the students diverse social-cultural background and the mandated assessment for learning approach EXCEPT ...

- ☐ Mr. Fauzan uses group project assessment, in which all the groups comprise local and non-local students.
- ☐ Mr. Fauzan gives essay assignment which have 5-6 alternative issues which the students may choose the one that suit them the best.
- ☐ Mr. Fauzan gives extra facts and concepts memorization tests to the non-local students to increase their familiarity with local cultures.

18. Which of the following action reflects fairness in conducting classroom assessment?

- ☐ Giving students opportunity to negotiate with teachers, what type of assessment will best demonstrate their learning.
- ☐ Developing learning targets at the beginning of semester to evaluate all students learning at the end of semester.

- ☐ Combining students' assessment from paper-and-pencil tests, essays, and projects with results from attitude observation.

Part 6. (School Assessment Policy)

One of the main features in the *Emancipated Learning* policy is that each school should develop their own unique school-level operating curriculum. The curriculum should accommodate the characteristics of the society where the school is located such as the social, cultural, and historical aspects. The school also needs to provide guideline regarding the school assessment policy as part of the curriculum.

19. The general guideline of the school assessment plan that reflect the *Emancipated Learning* policy should be …
    - ☐ applying priority classes for top students based on their previous achievements
    - ☐ setting timeline and types of assessment for mid- and end- of semesters.
    - ☐ emphasizing the use of assessment as part of student learning.

20. In response to the *Emancipated Learning* policy, what statement in the school assessment policy that shows the school are practicing fair practices regarding students' various level of competencies and learning needs?
    - ☐ Teachers should consider students' behaviour during classroom activities as part of final grade.
    - ☐ Teachers need to conduct diagnostic assessments at least at the beginning of academic year.
    - ☐ Teachers need to combine the assessment results from all formative and summative assessment in determining the final grade.

21. The following examples represent school assessment policy regarding strategies that teachers could use to improve student learning, which align with the *Emancipated Learning* policy, EXCEPT …
    - ☐ teachers use assessment results to rank students based final mark.
    - ☐ teachers use assessment results to identify students' difficulties.
    - ☐ teachers use assessment results to adjust teaching strategies.

22. One of the school assessment policy states that teachers teaching the same subject at the same grade need to establish parallel but different end-of-unit-semester tests. What guideline should be given in the assessment policy document to increase the reliability of the test?
    - ☐ The teachers should develop and use the same item specification.
    - ☐ The teachers should use items from standardized test item bank.
    - ☐ Each of the teachers should modify test items from the student book.

Prize draw and availability for interview session

1. Are you interested to join a prize draw? (Gift vouchers worth £8 each are available for 25% of lucky respondents) ☐ yes ☐ no — If yes, please provide your email address:

2. Are you willing to participate in the interview session regarding assessment enactment? (The interview will last for about 1 hour and conducted either face to face or online at your convenient time. Each interviewee will receive a small internet/transport fund £13) ☐ yes ☐ no — If yes, please provide your email address:

Appendix 3: Interview protocol for teachers

Open-ended interview items with selected social studies teachers

Identity of interviewee

Interviewee code : ……..

School Identity (Region and Province) : …….

Primary interview items for selected teachers are as follows:

1. Explain the extent to which assessment reform in the emancipatory curriculum has been enacted in classroom.
2. What are the aspects of assessment reform that are difficult to implement? why?
3. What are the factors that hinder the implementation of assessment reform in the classroom level?

   Possible prompting questions:

   - Classroom/students conditions (cognitive, engagement, economic issues).
   - Social-cultural-historical aspects of the society.
   - Local government/school policy.
   - School facilities.
   - Support from parents.
4. What are the factors that support the implementation of assessment reform in classroom level?

   Possible prompting questions:

   - Support from school colleagues in the form of informal/formal professional learning or other form of support.
   - Support from subject teacher community (social studies).
   - Support from school principal.
   - Support from school superintendent.
   - Support from other parties.
5. What are the strategies and adjustment, or modification made in enacting the assessment reform at classroom level that are deemed necessary?
6. Explain the impact of assessment experience in enacting the assessment reform.
7. Explain the impact of assessment education in higher education in enacting the assessment reform.
8. Explain the impact of assessment in-service trainings in enacting the assessment reform.

9. To what extent the enactment process of assessment reform has improved (or not) your assessment literacy?

Appendix 4: Interview protocol for principals

Open-ended interview items with selected principals

Identity of interviewee

Interviewee code : ……..

School Identity (Region and Province) : …….

Primary interview items for selected principals are as follows:

1. Explain the school policy regarding the implementation of assessment reform, if any. Elaborate the details.
2. What are the types of support provided to support the implementation of assessment reform?
3. To what extent the school has implemented mandated assessment reform. Give examples?
4. What are the obstacles that hinder the implementation of assessment reform?

   Possible prompting questions:

   - Social-cultural-historical aspects of the society
   - Students' conditions (economic/social-cultural)
   - Parents' support
   - Teachers' condition
   - School facilities
   - Local government policy
5. What are the factors that support the implementation of assessment reform at school level?

   Possible prompting questions:

   - Support from local government
   - Support from school superintendent
   - Support from school principal community
   - Support from teacher colleagues
   - Support from other parties
6. Explain the strategies and adjustment or modification in enacting the assessment reform at school level that are deemed necessary.
7. Explain the impact of assessment experience in enacting the assessment reform.

8. Explain the impact of assessment education in higher education in enacting the assessment reform.
9. Explain the impact of assessment in-service trainings in enacting the assessment reform.
10. To what extent the enactment process of assessment reform has improved (or not) your assessment literacy?

Appendix 5: Research ethics approval



University of Nottingham
UK | CHINA | MALAYSIA

School of Education
University of Nottingham
The Dearing Building
Jubilee Campus
Wollaton Road
Nottingham
NG8 1BB

educationresearchethics@nottingham.ac.uk

14/12/2022

Dear Yasser

Thank you for your research ethics application for your project:

Principle and Social Studies Teacher Assessment of Literacy and the Enactment of Assessment Reform in Diverse School Contexts

Thank you for addressing the minor points we raised in a letter dated 6 December 2022.

Following receipt of your revised documentation we are pleased to confirm your research is now:

- **Approved**

We wish you well with this research.

If your research overruns the intended completion date, please contact the Ethics Team to arrange an extension and update on any additions/changes to your work.

Please note that from September 2021 applications given a favourable ethical opinion will be subject to an annual audit in line with the *Code of Practice for Research Ethics Committees* (2021).



Prof Mary Oliver
On behalf of the Ethics Committee

+44 (0)115 9514470
educationadmin@nottingham.ac.uk
nottingham.ac.uk/education

Appendix 6: Formal request letter to local education authorities for accessing research site and participants

Hal : Permohonan ijin pengambilan data penelitian

Kepada
Yth. Kepala Dinas Pendidikan Kota Manado
Di Tempat

Bersama ini saya:
Nama : Yasser Awaluddin
Tempat Studi : University of Nottingham, United Kingdom

Mengajukan permohonan untuk melakukan pengambilan data penelitian sebagai bagian dari program pendidikan jenjang Doktoral di University of Nottingham UK. Adapun judul penelitian, target dan jumlah responden, dan jenis data penelitian yang akan saya gali adalah sebagai berikut:

Judul penelitian :

Literasi Asesmen Kepala Sekolah dan Guru IPS dalam Penerapan Reformasi Penilaian pada Berbagai Konteks Sekolah di Indonesia

Target dan jumlah responden :

1. Survey: 20 – 25 Kepala SMP dan 40 – 60 Guru IPS SMP di setiap Kota/Kabupaten.
2. Wawancara: 1 KS dan 1 Guru pada 5 Kota/Kabupaten terpilih.

Jenis data penelitian :

1. Survey: Literasi asesmen Kurikulum Merdeka pada KS dan Guru IPS
2. Wawancara: Implementasi asesmen dalam Kurikulum Merdeka

Demikian surat permohonan ini saya ajukan. Besar harapan saya agar permohonan ini dapat disetujui. Terima kasih atas perhatian dan bantuannya.

Nottingham, 9 February 2023

Yasser Awaluddin

Tim Peneliti

Mahasiswa PhD : Yasser Awaluddin (yasser.awaluddin@nottingham.ac.uk)

Pembimbing : 1. Assoc. Prof. Vicente Reyes, Jr. (Vicente.reyes@nottingham.ac.uk)
2. Assoc. Prof. Jenny Eliot (jenny.elliot@nottingham.ac.uk)

School of Education Research Ethics Coordinator:
educationresearchethics@nottingham.ac.uk

Appendix 7: Information sheet for survey participants

Information Sheet for Potential Survey Respondent

Title of the research:

Principal and Social Studies Teacher Assessment Literacy and the Enactment of Indonesian Assessment Reform in Diverse School Contexts.

Introduction:

This research is a PhD research project in the School of Education, University of Nottingham, United Kingdom. You are invited to participate in this project as a survey respondent because you are either a school principal or a social studies teacher, and your school is currently implementing the new Emancipatory curriculum of *Emancipated Learning* policy.

Before you decide to take part in this study, it is essential for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Take time to decide whether you wish to take part.

Purpose of study:

This project has four objectives:

1. To provide a picture of principal and social studies teacher assessment literacy that influence the success of the *Emancipated Learning* policy.
2. To examine how much teaching experience, assessment-related education and training, and local assessment learning support from the local context contribute to principals' and social studies teachers' assessment literacy.
3. To reveal how the school principals and social studies teachers enact the assessment reform as within their schools and whether the enactment process enhances their assessment literacy.
4. To examine whether principal assessment literacy affects teacher assessment literacy.

Description of participation:

This research has three phases, data collection instrument development (phase 1), quantitative (phase 2) and qualitative (phase 3). This is an invitation to phase 2 of the study. To participate, you will respond to an online questionnaire. The questionnaire consists of 8 questions of background information, 9 Likert-type questions of assessment education and support, and 22 (for principals) or 25 (for teachers) true-false multiple-choice questions. Approximately it will take 30 minutes to answer all the questions. Your responses will be used mainly to gain understanding regarding research objective number one and two.

At the end of the questionnaire, there will be a question asking whether you are interested in taking part in the second phase of this study in a 1-hour interview session. If you agree, you will be asked to give your email address so the researcher can contact you later.

Nature of participation:

Participation in this study is entirely voluntary, you may refuse to participate without prejudice and any adverse consequences because of your decision. Likewise, no potential risks nor harm will occur by participating in this research. If you wish to participate you can click the button "I am willing to take part as a respondent in this survey" at the end of this page. You may also withdraw at any point during the study without giving any reason. To allow the researchers to identify and delete your responses if you decide to withdraw when you have completed the survey, at the beginning of the survey you are asked to provide a word code (e.g., a combination of letter(s) and number(s)) with which the researcher can identify your data and delete it from the dataset. Withdrawing from the study will not affect you in anyway.

Benefit of Participation

Whilst there is not any immediate benefit for the respondents by participating in the study, respondents will have the opportunity to contribute to improving the educational reform in Indonesia. Your responses and information through this study can be valuable feedback for the implementation of the current educational reform. The findings of this study might be potentially used as the basis for improvement of the aspects related to the educational reform, such as improving the quality of in-service professional development for teachers and/or strengthening the support for local teacher professional communities both at school-based and regency/city-based.

Confidentiality and security of information:

This research ensures the confidentiality and anonymity of its participants. You will not be asked to give your identity or your school identity. Respondents of this study will not be identified from any reports and publication of the study. Further information regarding data privacy, confidentiality, and legal basis for collecting, storing, archiving, and sharing research data can be checked in a separated *privacy notice* document.

Data management, preservation, and sharing:

All the anonymous survey data will be stored in the University of Nottingham SharePoint, a secure cloud storage technology, until the research is finished and subsequently will be transferred, retained, preserved, and published through the University of Nottingham Research Data Management Repository under a CC-BY license for 7-25 years to allow data users to validate the research findings as part of the research integrity commitment. The survey data in SharePoint will be deleted permanently after it is transferred to the University of Nottingham Data Management Repository.

Contact details:

No.	Name	Role	Email address
1.	Associate Professor Vicente Reyes, Jr	Principal Investigator – PhD research supervisor	Vicente.reyes@nottingham.ac.uk
2.	Associate Professor Jenny Elliot	PhD research supervisor	Jenny.elliot@nottingham.ac.uk
3.	Yasser Awaluddin	PhD student	Yasser.awaluddin@nottingham.ac.uk
4.	Research Ethics Coordinator		educationresearchethics@nottingham.ac.uk

Consent statement:

By clicking the button below, survey respondents consent to their participation and of the data being used for the intended research purposes and being retained, preserved and published through the University of Nottingham Research Data Repository under a CC-BY license for 7 to 25 years. Please click the button below if you decide to participate:

O I am willing to take part as a respondent in this survey

Appendix 8: Information sheet for interview participants

Information Sheet for Interview Participant

Title of the research:

Principal and Social Studies Teacher Assessment Literacy and the Enactment of Indonesian Assessment Reform in Diverse School Contexts.

Introduction:

Thank you for your willingness to participate in the interview session of this research. Before you decide to take part in this interview session, please take time to read the following information carefully and discuss it with others if you wish. Take time to decide whether you wish to take part. If there is any unclear information, do not hesitate to contact the research team.

This interview is a part of my PhD research project in School of Education, University of Nottingham, United Kingdom. You are invited to participate in this project as an interview participant because you are either an junior secondary school principal or social studies teacher and your school is currently implementing the new Emancipatory curriculum of *Emancipated Learning* policy.

Purpose of study:

This project has four objectives:

1. To provide a picture of principal and social studies teacher assessment literacy that influence the success of the *Emancipated Learning* policy.
2. To examine how much teaching experience, assessment-related education and training, and local assessment learning support from the local context contribute to principals' and social studies teachers' assessment literacy.
3. To reveal how the school principals and social studies teachers enact the assessment reform as within their schools and whether the enactment process enhances their assessment literacy.
4. To examine whether principal assessment literacy affects teacher assessment literacy.

Description of participation:

This is an invitation of phase 3 of the research. To participate, you will be interviewed regarding the enactment of the assessment reform in your school/classroom. Your answers will be used to address the research objective number three and four. The researcher would like to ask about your experience in enacting the assessment reform, such as how far it has been implemented, what kind of support and difficulties you had, and what modification or adjustment you have made and why.

The interview will be either face to face or online using Microsoft Teams, whichever convenient with you, and will be conducted at your school or other places you feel more comfortable at your chosen time. This interview will last approximately one hour and will be recorded to be analysed later on. The audio/video files will be transcribed verbatim manually by the researcher and then will be translated to English.

Nature of participation:

The participation in this study is fully voluntary, you may refuse to participate or decide to withdraw at any time during the project without prejudice or any adverse consequences because of your decision. Likewise, no potential risks nor harm will occur by participating in this research. If you agree to participate, you will be asked to sign a consent letter. Should you decide to withdraw after the interview is conducted and transcribed, all the interview data related to you will be identified from the storage system and permanently deleted at once. Withdrawing from the study will not affect you in anyway.

Benefit of Participation

Whilst there is not any immediate benefit by participating in the study, participants will have the opportunity to contribute to improving the educational reform in Indonesia. By participating in this study, you will have an opportunity to express your ideas, objection, or obstacles which might become important information for the policy maker to improve of the *Emancipated Learning* policy.

Confidentiality and security of information:

This research ensures the confidentiality and anonymity of its participants. Pseudonymisation techniques to all identifiable information such as name, school identity, age, degree, and places will be applied to de-identify the interview participants in the final English-version transcripts. Any summary interview content, or direct quotations from the interview, that are made available through project reports and academic publications will also be subject to the same rigorous pseudonymisation. As such, all interview participants cannot be identified in all shared and published research outputs.

During the course of the research, all files containing personal data (name, school identity, image, voice, and other identifiable information) will be stored securely in password-protected folders in University of Nottingham cloud system (SharePoint) and have restricted access only to member of the research team. All the interview data containing identifying information kept during the project will be deleted permanently after PhD is completed.

Further information regarding data privacy, confidentiality, and legal basis for collecting, storing, archiving, and sharing research data can be checked in a separated *privacy notice* document.

Data management, preservation, and sharing:

The pseudonymised English-version of the interview transcripts will be transferred to University of Nottingham Research Data Repository, where it will be retained, preserved, and published under a CC-BY license for 7 to 25 years to allow data users to validate the research findings as part of the research integrity commitment. The principal investigator,

as the Data Owner, will retain overall responsibility for managing data, data security and data quality.

Contact details:

No.	Name	Role	Email address
1.	Associate Professor Vicente Reyes, Jr	Principal Investigator – research supervisor	Vicente.reyes@nottingham.ac.uk
2.	Associate Professor Jenny Elliot	Research supervisor	Jenny.elliot@nottingham.ac.uk
3.	Yasser Awaluddin	PhD student	Yasser.awaluddin@nottingham.ac.uk
4.	Research Ethics Coordinator		educationresearchethics@nottingham.ac.uk

Appendix 9: Interview participant consent form



University of Nottingham
UK | CHINA | MALAYSIA

INTERVIEW PARTICIPANT CONSENT FORM

Project title: Principal and Teacher Assessment Literacy and the Enactment of Indonesian Assessment Reform in Diverse School Contexts

Researcher's name: Yasser Awaluddin

Supervisor's name: Associate Professor Vicente Reyes, Jr.
Associate Professor Jenny Elliot

- I have read the Participant Information Sheet, and the nature and purpose of the research project has been explained to me. I understand and agree to take part.
- I have read the project's Privacy Notice.
- I understand the purpose of the research project and my involvement in it.
- I understand that I may withdraw from the research project at any stage and that this will not affect my status now or in the future.
- I understand that while information gained during the study may be published, I will not be identified, and my personal results will remain confidential.
- I understand that I will be audio/video recorded during the interview.
- I understand that my personal data (name, school identity, image, voice, and other identifiable information) will be stored securely in password-protected folders in University of Nottingham one-drive and only accessible by the researcher.
- I understand that I may contact the researcher or supervisor if I require further information about the research, and that I may contact the Research Ethics Coordinator of the School of Education, University of Nottingham, if I wish to make a complaint relating to my involvement in the research.

Signed …… (research participant)

Print name …… **Date** ……

Contact details

Researcher: Yasser.awaluddin@nottingham.ac.uk

Supervisor:

School of Education Research Ethics Coordinator:
educationresearchethics@nottingham.ac.uk

Appendix 10: Codes from first cycle coding process

First Cycle Coding

No.	Main codes	No.	Second and third level codes
1.	% of enactment		
2.	Challenges of enactment	1.	Challenges from local administrators
		2.	Inconsistency in concept application
		3.	Involving parents
		4.	More complex than previous assessment:
			1. Administrative demand
			2. School project
			3. Clarity of new concepts
			4. Combining 3 aspect for final mark
			5. Differentiated learning
			6. Lack of support from colleagues
			7. Low student groups
			8. More work
		5.	Parents complain of the use of digital assessment
		6.	Resistance of change
		7.	Teacher readiness
		8.	Time
		9.	Teacher creativity
		10.	Workload for small/private schools
		11.	Teacher student ratio
3.	Potential barriers of new policy		
		12.	Disagreement with new policy
			9. Reasons why national exam should be maintained
			10. Thoughts on the use of national exam
		11.	Changing policy
		12.	False perception on new curriculum and assessment
		13.	Learning new concepts slowly
4.	Gap between pilot and non-pilot schools		
5.	General aspects of enactment		
		14.	Diverse student SES
		15.	Knowledge sharing of pilot schools

6.	Similarity with previous concepts		
7.	Transition stage in 1st year		
8.	Self-perceived assessment literacy		
9.	Difficult aspects		
		16.	Diagnostic assessment
10.	Impact of covid		
11.	Strategies of implementation		
		17.	Timely communication of policy changes
		18.	Following guidance from MoE fully
		19.	Engaging parents
		20.	Different approach between high and low classes
		21.	Emphasizing diagnostic assessment
		22.	Emphasizing FA at the end of every session
		23.	Encouraging students' engagement
		24.	Flexible assessment time according to students' learning speed
		25.	Flexible format of summative assessment
		26.	Focusing on student distinct characteristics
			11. Doubt in categorising students
			12. Student can choose type of assessment
		27.	Holistic assessment through project
		28.	Peer tutor in student groups
		29.	Intensive supervision
		30.	Regular update of school operating curriculum
		31.	Selected materials for summative assessment
		32.	Semi credit approach
		33.	Choosing Innovative teachers as the initial team
		34.	Devoting a team to ensure effective implementation (non-pilot, private school)
		35.	Emphasizing teacher to teach students to be clever rather than intelligent
		36.	Encouraging teacher creativity
		37.	Focus on teacher IT skills
		38.	Keeping the same teachers for students
		39.	Providing materials needed

		40.	Regular meetings and sharing good practices
		41.	School CoP
		42.	Teacher collaboration
12.	School assessment plan		
		43.	Combining academic and non-academic achievement
		44.	Combining summative and formative assessment for final mark
		45.	Cut off scores
		46.	Reward as assessment
		47.	School made guide
13.	Adjustment, adaptation, and modification		
14.	Teacher empathy		
15.	Digital based assessment		
16.	Strategizing against local policy		
17.	School policy based on discussion and agreement		
18.	Combining new policy with existing school policy		
19.	Strength of new policy		
		48.	Accommodating local culture easier with project
		49.	Better knowing of students
		50.	Better learning outcome
		51.	Higher motivation
		52.	More interesting to teachers and students
		53.	More just for students
		54.	Simpler
20.	Support of enactment		
		55.	Parent support
		56.	Flexibility from local authority
		57.	Monitoring from local authority
		58.	Support from local teacher community
		59.	Mentoring program
		60.	PMM (teaching resources online platform)
		61.	Facilitating Trainings
		62.	Financial support from school operational fund (Bostrom & Palm)
		63.	High motivation teachers
		64.	Young and tech savvy teachers

21.	Students' diverse characteristics in line with new policy		
22.	Variation in formative and summative assessment		
23.	Variation in conducting diagnostic assessment		
		65.	Multiple choice for diagnostic assessment