



**An inquiry into teacher assessment literacy: the development of an assessment
literacy framework and instrument to measure the assessment literacy of
teachers in Saudi Arabia**

Abdullah Alzhrany

A thesis submitted to fulfil requirements for the degree of

Doctor of Philosophy

School of Education and Social Work

Faculty of Arts and Social Sciences

The University of Sydney

2023

Statement of Originality

This is to certify that to the best of my knowledge; the content of this thesis is my own work. This thesis has not been submitted for any degree or other purposes.

I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.

Signature*

Name:

Abdullah Ahmad Ali Alzhrany

Table of Contents

Statement of Originality	i
Table of Contents	ii
List of Figures.....	vi
List of Tables	viii
List of Abbreviations	x
Acknowledgements	xii
Abstract.....	xiii
Chapter 1: Setting the Scene.....	1
1.1 Introduction	1
1.2 Educational Reforms and Issues in Saudi Arabia.....	2
1.3 Statement of the Problem.....	2
1.4 Teacher Assessment Literacy Research: International Contexts	4
1.5 Assessment Literacy and its Measures	6
1.6 Teacher Assessment Literacy Research: The KSA Context.....	8
1.7 Development and Validation of the Assessment Literacy Framework and Questionnaire	12
1.8 Significance of the Study and research Questions	13
1.9 Organisation of the Thesis	16
Chapter 2: Literature Review	17
2.1 Introduction.....	17
2.2 Saudi Arabia: A Social, Economic and Political Overview	17
2.2.1 An Overview of the Education System in Saudi Arabia: Historical and Recent Developments.....	19
2.2.2 Education Structure and Management in Saudi Arabia	25
2.2.3 MOE and Vision 2030: The Role of Education in Remaking a Nation.....	28
2.2.4 Teaching and Teacher Education Policies in Saudi Arabia	33
2.3 Curriculum and Assessment.....	36
2.4 Assessment Literacy: A Review of the Literature.....	42
2.5 Measuring Teacher Assessment Literacy.....	53
2.6 Assessment Literacy Studies in the Context of Saudi Arabia	55
2.7 Teacher Assessment Literacy Professional Learning	57
2.8 Summary	60
Chapter 3: Methodology and Methods	63
3.1 Introduction	63
3.2 Methodology	64
3.3 Delineating the Teacher Assessment Standards with the Professional Knowledge and Understanding to Develop the Assessment Literacy Framework	65
3.4 Research Ethics	66
3.5 The Sample	68

3.6 Description of the Stages	71
3.6.1 Stage 1: Development and Validation of the Assessment Literacy Framework	71
3.6.1.1 Step 1: Unpacking and Elaborating the Various NPST Assessment Standards ...	71
3.6.1.2 Step 2: Refining the Assessment Literacy Framework.....	72
3.6.1.3 Step 3: Focus Group Sessions.....	72
3.6.2 Stage 2: Development and Validation of the Assessment Literacy Questionnaire	73
3.6.3 Stage 3: The Final Review of the Instrument (MTALQ) and Implementation	75
3.6.3.1 Step 1: The Use of Teacher Assessment Literacy Framework.....	75
3.6.3.2 Step 2: The Feedback Sessions from the Participants.....	76
3.6.3.3 Step 3: The Finalisation of MTALQ.....	76
3.7 Analysis using Classical Test Theory	77
3.8 Administration of the Survey.....	78
3.9 Collection of Validity Evidence.....	78
3.10 Summary	79
Chapter 4: Results.....	80
4.1 Introduction.....	80
4.2 Survey Responses	80
4.3 Primary Teachers' Focus Group Responses – Elaborations requiring changes	84
4.3.1 Elaboration 29.....	85
4.3.2 Elaboration 57.....	86
4.3.3 Elaboration 23.....	87
4.3.4 Elaboration 43.....	88
4.3.5 Elaboration 14.....	89
4.3.6 Elaboration 16.....	90
4.3.7 Elaboration 51.....	91
4.4 Primary Teachers' Focus Group Responses - Elaborations not Requiring Amendments ..	93
4.4.1 Elaboration 20.....	93
4.4.2 Elaborations 21, 22, 24 and 25.....	94
4.4.3 Elaborations 30 and 31	95
4.4.4 Elaboration 32.....	95
4.4.5 Elaboration 36.....	96
4.4.6 Elaboration 37.....	96
4.4.7 Elaboration 44.....	97
4.4.8 Elaboration 47.....	98
4.4.9 Elaboration 56.....	98
4.5 Secondary and Intermediate Teachers' Focus Group Responses - Elaborations	
Requiring Amendments.....	99
4.5.1 Elaborations Requiring Amendments.....	99
4.5.2 Elaborations not Requiring Amendments.....	99
4.5.3 Elaboration 21.....	100
4.5.4 Elaboration 22.....	100
4.5.5 Elaboration 30.....	101
4.5.6 Elaboration 31.....	101
4.5.7 Elaboration 32.....	102
4.5.8 Elaboration 36.....	102
4.5.9 Elaboration 44.....	103

4.5.10 Elaboration 56.....	104
4.6 Aligning Items of the Original TALQ with the Elaborations that Comprise the Refined Assessment Literacy Framework	105
4.7 Building the Modified TALQ-MTALQ.....	120
4.7.1 TALQ Standard 1: Choosing Assessment Methods (Standard 10.1 in NPST)	122
4.7.2 TALQ Standard 2: Developing Assessment Methods (Standard 10.1 in NPST)	122
4.7.3 TALQ Standard 4: Using Assessment Results for Decision-making (Standard 10.4 in NPST).....	123
4.7.4 TALQ Standard 5: Using Assessment in Grading (Standard 10.2 in NPST).	124
4.7.5 TALQ Standard 6: Communicating Assessment Results (Standard 10.5 in NPST).	125
4.7.6 TALQ Standard 7: Modern Assessment, or Relevant Knowledge and Skills for Assessment (Standard 10.3 in NPST).	127
4.8 Finalising the MTALQ	129
4.9 Chapter Summary	134
Chapter 5: Results from Analysis of the Modified Teacher Assessment Literacy Questionnaire (MTALQ).....	137
5.1 Introduction.....	137
5.2 Demographics of the Respondents to the MTALQ	137
5.3 The MTALQ Results	139
5.4 Summary of Item Statistics by Standards	141
5.4.1 MTALQ Standard 1: Choosing Assessment Methods (Standard 10.1 in NPST).	143
5.4.2 MTALQ Standard 2: Developing Assessment Methods (Standard 10.1 in NPST)... ..	147
5.4.3 MTALQ Standard 3: Administering, Scoring and Interpreting Assessment Results (Standard 10.2 in NPST).	149
5.4.4 MTALQ Standard 4: Using Assessment Results for Decision-making (Standard 10.4 in NPST).	153
5.4.5 MTALQ Standard 5: Using Assessment in Grading (Standard 10.2 in NPST).	157
5.4.6 TALQ Standard 6: Communicating Assessment Results (Standard 10.5 in NPST).	162
5.4.7 MTALQ Standard 7: Modern Assessment, or Relevant Knowledge and Skills for Assessment (Standard 10.3 in NPST).	165
5.5 Chapter Summary	167
Chapter 6: Discussion	169
6.1 Introduction.....	169
6.2 Research Question 1: The Assessment Literacy Framework	170
6.2.1 Utility of the Refined Assessment Literacy Framework	187
6.3 Research Question 2: The Validity of the TALQ for Assessing NPST.....	189
6.4 Validity of the MTALQ	203
6.5 Research Question 3: Validity of the MTALQ for Assessing the NPST.	206
6.6 Chapter Summary	217
Chapter 7: Conclusion.....	219
7.1 Introduction.....	219
7.2 Summary of Results.....	219
7.3 Implications and Recommendations of the Study	221
7.4 Assessment Literacy: A Vital Cog in the Efficacy of the KSA Educational System	221
7.5 Contributions of the Study	227

7.6 Areas for Future Research	229
7.7 Limitations of the Study.....	229
7.8 Concluding Remarks.....	231
References.....	235
Appendices.....	250
Appendix A: Ethics Approval.....	250
Appendix B: The Likert Survey.....	253
Appendix C: The Assessment Literacy Framework	274
Appendix D: The Refined Assessment Literacy Framework	291
Appendix E: The NPST	308
Appendix F: The Original TALQ	310
Appendix G: The Modified Teacher Assessment Literacy Questionnaire	326
Appendix H: Invitation to School Principal.....	339
Appendix I: Letter to Ministry of Education	341
Appendix J: Participant Consent.....	344
Appendix K: Teacher Participant Plain Language Statement.....	349
Appendix L: The Ministry of Education Approval.....	352
Appendix M: The Safety Protocol	353

List of Figures

Figure 4.1 Item 21 of the Original TALQ.....	110
Figure 4.2 Item 19 of the Original TALQ.....	111
Figure 4.3 Item 10 of the Original TALQ.....	113
Figure 4.4 Item 21 of the Original TALQ.....	114
Figure 4.5 Item 22 of the Original TALQ.....	114
Figure 4.6 Item 24 of the Original TALQ.....	115
Figure 4.7 Item 26 of the Original TALQ.....	117
Figure 4.8 Item 30 of the Original TALQ.....	117
Figure 4.9 Item 33 of the Original TALQ.....	118
Figure 4.10 Item 14 of the Original TALQ.....	118
Figure 4.11 Item 16 of the Original TALQ.....	119
Figure 4.12 Item 35 of the Original TALQ.....	120
Figure 4.13 Item 6 of the MTALQ	122
Figure 4.14 Item 12 of the MTALQ	123
Figure 4.15 Item 18 of the MTALQ	124
Figure 4.16 Item 21 of the MTALQ	124
Figure 4.17 Item 28 of the MTALQ	125
Figure 4.18 Deleted Item No. 28 of the Original TALQ	126
Figure 4.19 Item 30 of the MTALQ	127
Figure 4.20 Item 6 of the MTALQ	131
Figure 4.21 Item 33 of the MTALQ	132
Figure 4.22 Item 35 of the MTALQ	133
Figure 4.23 Item 36 of the MTALQ	133
Figure 5.1 Item 1 of the MTALQ	144
Figure 5.2 Item 4 of the MTALQ	145
Figure 5.3 Item 7 of the MTALQ	148
Figure 5.4 Item 8 of the MTALQ	149
Figure 5.5 Items 14 and 15 of the MTALQ	150
Figure 5.6 Item 18 of the MTALQ	154

Figure 5.7 Item 20 of the MTALQ	155
Figure 5.8 Item 19 of the MTALQ	156
Figure 5.9 Item 25 of the MTALQ	158
Figure 5.10 Item 24 of the MTALQ	159
Figure 5.11 Item 27 of the MTALQ	160
Figure 5.12 Item 28 of the MTALQ	161
Figure 5.13 Item 31 of the MTALQ	163
Figure 5.14 Item 32 of the MTALQ	164
Figure 5.15 Item 34 of the MTALQ	167

List of Tables

Table 4.1	Range of Means for Elaborations Obtained from the Judgement of Primary Teachers on the Relative Importance of the Elaborations of the Practitioner and Advanced Career Levels of Standard 10 of the Saudi Arabian Professional Standards for Teachers.....	82
Table 4.2	Range of Means for Elaborations Obtained from the Judgement of Secondary and Intermediate Teachers on the Relative Importance of the Elaborations of the Practitioner and Advanced Levels of Standard 10 of the Saudi Arabian Professional Standards for Teachers.....	83
Table 4.3	Aligning a Sample of the Original TALQ Items with Elaborations of the Refined Assessment Literacy Framework and the NPST Standards.....	106
Table 4.4	Items to be Deleted in Response to the Refined Assessment Literacy Framework...	116
Table 4.5	The Number of Items Assigned to the Five Standards in the NPST.....	121
Table 5.1	Demographic and Career Profiles of KSA Teacher Respondents	138
Table 5.2	Summarised Highlights of Results.....	140
Table 5.3	Summary of Statistical Results by Item	142
Table 6.1	Connections Between the NPST, Items in the Original TALQ, Elaborations in the Refined Assessment Literacy Framework, Corrected Item-To-Total Score Correlation and Cronbach's Alpha if an Item is Deleted.....	192
Table 6.2	Item 6 in the Original TALQ, Corrected Item-To-Total Score Correlation/Cronbach's Alpha if Item Deleted, and Elaborations in the Refined Assessment Literacy Framework.....	198
Table 6.3	Items 8 And 9 in the Original TALQ, Corrected Item-To-Total Score Correlation/Cronbach's Alpha if Item Deleted, and Elaborations in the Refined Assessment Literacy Framework.....	202
Table 6.4	Items 6 in the MTALQ, Corrected Item-To-Total Score Correlation/Cronbach's Alpha if Item Deleted, and Elaborations in the Refined Assessment Literacy Framework	209

Table 6.5 Items 2 and 8 in the MTALQ, Corrected Item-to-total Score Correlation/Cronbach's Alpha if Item Deleted, and Elaborations in the Refined Assessment Literacy Framework.....	214
Table 6.6 Comparison of TALQ and Similar Instruments' Reliability.....	216

List of Abbreviations

AfL	Assessment for Learning
AITSL	Australian Institute for Teaching and School Leadership
ALI	Assessment Literacy Inventory
APST	Australian Professional Standards for Teachers
CALI	Classroom Assessment Literacy Inventory
CTT	Classical Test Theory
DDDM	Data-Driven Decision-Making
ETEC	Education and Training Evaluation Commission
GAT	General Achievement Test
GDP	Gross Domestic Product
HREC	Human Research Ethics Committee
IRT	Item Response Theory
ITP	Initial Teacher Preparation
KSA	Kingdom of Saudi Arabia
MENA	Middle East and North African countries
MOE	Ministry of Education
MTALQ	Modified Teacher Assessment Literacy Questionnaire
NAP	National Assessment Program
NTP	National Transformation Program
NPST	National Professional Standards for Teachers
OECD	Organisation for Economic Co-operation and Development
PIRLS	Progress in International Reading Literacy Study
PISA	Program for International Student Assessment
PLC	Professional Learning Community
SAAT	Scholastic Achievement Admission Test
SRL	Self-regulated Learning
TALIS	Teaching and Learning International Survey
TALQ	Teacher Assessment Literacy Questionnaire

Tatweer	King Abdullah bin Abdulaziz's Public Education Development Project
TCAQ	Teacher Competencies Assessment Questionnaire
TIMSS	Trends in International Mathematics and Science Study
UNESCO	United Nations Educational, Scientific and Cultural Organization
US	United States of America
ZPD	Zone of Proximal Development

Acknowledgements

Through my PhD journey, I have learned that a PhD cannot be accomplished without the support of many individuals. This supports needs in various forms mainly inspiration, encouragement, and technical assistance. I am immensely grateful to everyone who contributed to my project, particularly thankful to Professor Jim Tognolini and Dr. Jon Twing as without their help, I would not have been able to complete this study.

A valuable lesson I have learned from Professor Jim is that the educational assessment, measurement, and evaluation improves student learning ultimately all over the world. Apart from studies, Professor Jim fostered my networks and personality. Dr Jon, a remarkable statistical expert, inspired my commitment to professionalism.

I am grateful to my friends for their incredible support during my PhD journey. Their supportive presence helped me in resolving administrative and academic issues and in celebrating the occasional successes during studies.

I would like to express my gratitude to CEMA members for their inspiring support in enhancing the education system. I also acknowledge the participants who assisted me in data collection for their cooperation.

Also, I acknowledge the support of my family, especially my wife Safiah, who cared for our three children throughout my PhD journey. Safiah, your incredible contribution is beyond words. To my daughter Mira and sons Naser and Basel, you are my eternal sources of joy and love, and you empowered me to complete this work. You are the most beautiful part of my life, and I cherish you always.

Abstract

In the era of increasing educational reforms, teacher assessment literacy has increasingly been recognised as an essential competency of effective teachers because effective teaching practices guided by evidence and information like that collected during assessment are integral to improving student outcomes. The increasing emphasis in teacher assessment literacy can also be attributed to the recognition that sound assessment practices are a cornerstone of efficacious educational reforms. As part of its own recent educational reforms, the Kingdom of Saudi Arabia (KSA) established the National Professional Standards for Teachers (NPST) for its teachers to meet. The incorporation of assessment as one of the standards in the NPST recognises the importance of assessment literacy for KSA teachers. In fact, a recent Organisation for Economic Co-operation and Development (OECD) report indicated that KSA teachers have low levels of assessment literacy. When set against KSA's aim to raise its quality of education through its educational reforms such findings are concerning. Moreover, while the NPST incorporates assessment as one of its standards, the delineation of assessment skills and knowledge are relatively brief and not contextualized.

Given this environment, this study developed a comprehensive assessment literacy framework that aims to help Practitioner and Advanced teachers in the KSA to achieve the National Professional Standards for Teachers (NPST). In addition to developing the assessment literacy framework, the researcher also worked collaboratively with KSA teacher participants to develop an assessment literacy instrument or questionnaire, that is, the modified version of Teacher Assessment Literacy Questionnaire TALQ, called the MTALQ. In addition to focus

group interviews, the study used Classical Test Theory to analyse the results, focus on collecting internal consistency reliability, evidence, and investigated simple structure via correlations.

The findings indicate that KSA teachers considered most of the skills incorporated into the assessment literacy framework (based on the NPST) as essential skills and knowledge, or at least desirable knowledge, for teachers to demonstrate that they have met the NPST and have the competency to effectively enhance student learning outcomes. However, despite the robust methodology used to enhance the validity of the instrument (MTALQ), the results indicate that KSA teachers faced a degree of difficulty in answering the items correctly which impacts aspects of the instrument's reliability evidence and requires additional longitudinal evidence of validity. In addition, the results also corroborate (albeit with a small purposively selected group of teachers) OECD findings that KSA teachers have a rather low level of assessment literacy.

The key outcomes of this study are the development of a refined assessment literacy framework and MTALQ. In KSA's ongoing educational reforms, the refined assessment literacy framework can be employed to unpack the NPST in relation to the kinds of skills and knowledge KSA teachers need to meet the NPST. MTALQ can also be used as a tool to diagnose teachers' assessment literacy levels and the areas in which teachers might be lacking. With both the refined assessment literacy framework and MTALQ, professional learning around assessment literacy in the KSA can be more structured and targeted to prepare pre-service teachers and upskill in-service teachers to meet the NPST and enhance student learning outcomes.

Chapter 1: Setting the Scene

1.1 Introduction

Internationally, educational systems have attached great importance to the role of assessment because of its crucial function in the learning and teaching processes and its contribution to quality learning outcomes (Black & Wiliam, 1998). Developing countries have been urged to adopt strategies and programs based on current best practice in assessment to meet short- and long-term objectives and improve the quality of their educational systems with the goal of achieving parity with the rest of the world (Al-Sadan, 2000). Despite recent educational reforms and the importance of assessment in student learning outcomes, little is known about the state of teachers' assessment literacy and the capacity of teachers in using assessment to enhance student learning outcomes to achieve policy aims, though there is much discussion that assessment literacy for teachers is a relatively neglected area of training. This is particularly true in the KSA where it has little or no attention.

Informed by this context, one essential aim of this study is to construct an assessment literacy framework for developing and validating a teacher assessment literacy questionnaire which can be used to measure the assessment literacy of teachers. The proof of concept will be conducted using schools in the Kingdom of Saudi Arabia (KSA). The assessment literacy framework will specify what it is teachers need to know and be able to do to demonstrate proficiency in relation to the assessment standards prescribed in the KSA National Professional Standards for Teachers (NPST). The assessment literacy questionnaire will be developed and administered to a small sample of KSA teachers with the intention of validating the questionnaire and providing some insights into the state of KSA teachers' assessment literacy. This will serve as an example of how the questionnaire can be used to assess teacher assessment literacy in

developing countries. Having a better understanding of teachers' assessment literacy will provide evidence upon which to base policy recommendations regarding how teachers can be better supported in terms of building and consolidating their assessment knowledge and skills. It will also assist in designing the kinds of professional learning in assessment literacy that are best suited to enhance knowledge and skills gaps which teachers may have in this area.

1.2 Educational Reforms and Issues in Saudi Arabia

In 2013, the KSA founded its Education, Training Evaluation Commission (ETEC), a government body that specialises in assessment, measurement, and evaluation. In 2016, the ETEC partnered with King Abdullah bin Abdulaziz's Public Education Development Project (TATWEER) to create the National Professional Standards for Teachers (NPST) in KSA. These Standards encompass three areas: Professional Knowledge, Professional Practices (of which assessment is the third component of this level) and Professional Values and Responsibilities (see Appendix E for the NPST). In 2016, the KSA officially announced the landmark 'Saudi Vision 2030'. This policy aims to reform and develop all aspects of the country—key of which is education. The vision promises substantial reforms to the entire education system and aims to improve the performance of KSA's educational system so that it is on par with other leading educational systems across the world. To this end, in 2019, the Ministry of Education (MOE) reformed its national curriculum and designed new standards which are aligned to the global educational agenda of developing 21st-century competencies (Tognolini, 2020).

1.3 Statement of the Problem

In the landmark Saudi Vision 2030 policy (discussed further in Chapter 2), reforms of the KSA educational system are an area which is part of the national reforms. In this era of increasing educational reform, teacher assessment literacy has been identified as an essential

competency of effective teachers (R. J. Stiggins, 1991a). The increasing interest in teacher assessment literacy can be attributed to the recognition of the importance of assessment as part of school and educational system reforms (Heitink et al., 2016). It is also seen as a vital part of teachers' work in fostering student learning through effective classroom assessments (DeLuca et al., 2013; DeLuca et al., 2015; M. F. Hill et al., 2010; Lam, 2015; Popham, 2011). However, the problem is that while there is an acknowledgement of the importance of teacher assessment literacy in systems around the world, and a commitment captured in requirements built into statements of professional standards, there does not appear to be ready measures of teacher assessment literacy. In order for systems to be able to impact on teacher assessment literacy they need to have measures of what the assessment literacy of teachers is at the moment (baseline), introduce an evidence-based strategy to improve the assessment literacy (if deemed necessary) and then have further measures to ascertain whether the strategies have helped close the gap (that is, from where the system currently is to where it wants to be in terms of the assessment literacy of the teachers).

A starting point for systems is a set of professional standards for teachers (if they are available). This study addresses the problem by taking an existing instrument of assessment literacy for teachers from one education system and demonstrates a process for contextualising and validating it for use in a different system. In the KSA there seems to be an absence of a clear articulation of an assessment literacy framework that informs teaching assessment practices against a teacher's experience level, as well as a tool to effectively measure KSA teachers' assessment literacy against the (NPST). Therefore, considering the paucity of studies in this key area and the absence of an assessment literacy framework and tool, this study will fill the gaps that have been identified and highlighted in the literature review in Chapter 2.

1.4 Teacher Assessment Literacy Research: International Contexts

KSA's interest in incorporating assessment literacy in its educational system (a represented by the NPST) is not new. Interest in teacher assessment literacy began several decades ago. Initial definitions of assessment literacy were centred around the more practical dimension of assessment, such as the strategies and skills needed to assess (Gotch & French, 2014; Schafer & Lissitz, 1987). Much of educational research in the area (Brown, 2008b; DeLuca et al., 2013; R. Stiggins, 2009; R. J. Stiggins, 1995) as well as educational policies around the world have attempted to inform and support teacher practice based on what is thought to be essential for effective teaching practice. For example, in Australia, the Australian Professional Standards for Teachers (APST) include the articulation of the competencies that teachers need to acquire in terms of assessment (Australian Institute for Teaching and School Leadership [AITSL], 2018). Similarly, in the KSA, the NPST also includes the kinds of knowledge and skills that teachers at different levels of experience ought to possess in terms of assessment literacy. Thus, on a broader level, teacher assessment literacy can be viewed as 'an individual's understandings of the fundamental assessment concepts and procedures deemed likely to influence educational decisions' (Popham, 2011, p. 267). A typical behaviour that indicates teacher assessment literacy would be the teacher using multiple assessments that are aligned with defined achievement goals and then inferring student performance against nominated criteria or standards.

Several other definitions that are based on the notions of standards and competencies have also been proffered in the literature. For example, according to R. J. Stiggins (1991a, 1991b), 'teacher assessment literacy is knowledge about educational assessment and the skills required to apply that knowledge to measure student learning'. Some of the actions he classified

as being key to effective assessment include explaining and sharing with students the reasons behind assessing; what is to be assessed; designing precise assessment processes; and communicating results of the assessment effectively. DeLuca and Klinger (2010) view assessment literacy as having the knowledge of the use of various assessment practices. They also highlight the importance of having the theoretical and philosophical knowledge that underpin the types of assessment of students' learning. In integrating both the micro- and macro-perspectives on educational and school assessment, Brookhart (2011) extended the notion of teacher assessment literacy and reconsidered the Seven Standards of Teacher Competencies in Educational Assessment of Students (American Federation of Teachers [AFT] et al., 1990).

The changing perspectives of assessment literacy has meant a growing focus on 'teacher capabilities to plan and implement quality assessment tasks, to interpret evidence and outcomes appropriate to the assessment purpose and type, and to engage students themselves as active participants in assessment of their own learning' (Looney et al., 2018, p. 2). Such changes are apparent in current teacher standards globally, which seem to now be more focused on formative assessments, support for teachers and the way assessment results are being reported and conveyed (DeLuca et al., 2016a). For instance, in Australia, the APST related to assessment is expressed as: 'assess, provide feedback and report on student learning' and the purpose of keeping accurate and reliable records of student achievement (AITSL, 2018, pp. 18-19). Similarly, apart from other standards that guide teacher professional practice, KSA's NPST has a standard that specifically relates to teachers' capacity in assessment which reflect definitions of assessment literacy (ETEC, 2016). Standard 10 of the NPST refers to 'evaluating the student's performance' and its sub-standards include: 'preparing evaluation tools; applying evaluation; engaging students in evaluation process and Preparing Evaluation Reports' (ETEC, 2016).

Rather, unsurprisingly, the notion of teacher assessment literacy has been gradually shifting over time. More recent conceptions of teacher assessment literacy also recognise the importance and the role of context in the development and enactment of the assessment knowledge and skills. It is also seen as part of teachers' professional identities in which teachers amalgamate their knowledge of assessment with that of content, pedagogy, and the broader learning context (Adie, 2013; Cowie et al., 2014; Looney et al., 2018; Scarino, 2013; Xu & Brown, 2016). Various contextual factors can have an impact on the nature and extent of assessment practices, and these include teacher experience, professional learning, classroom context and curriculum (Looney et al., 2018). It is such socio-cultural factors that can account for the diverse teacher experiences of assessment despite the uniformity in educational policies and professional learning (Tierney, 2006). Past studies (for example, Coombs et al., 2018; DeLuca et al., 2016a, 2016b) indicate that these include teachers' understandings of the purposes of assessment and the assessment processes. Therefore, while teacher capacities in assessment remain key, more recent conceptions of teacher assessment literacy view assessment in a more complex manner considering a host of socio-cultural factors that impact on what teachers think and do regarding assessment practices.

1.5 Assessment Literacy and its Measures

Reflecting the increasing international recognition that assessment literacy is a key area in determining student success, several studies have been conducted in the Western context and have sought to measure teacher assessment literacy using quantitative measures. Some of these measures used were informed by the original 1990 *Standards for Teacher Competence in Educational Assessment of Students* (AFT et al., 1990). For example, Plake et al. (1993, pp. 10–39) analysed the assessment literacy of 555 teachers and 268 administrators across 45 states in

the United States (US). A 35-item instrument, the Teacher Competencies Assessment Questionnaire (TCAQ) was designed to measure the seven competency standards. Findings indicated substantial weaknesses in teachers' understanding and use of assessments, as demonstrated by an average score of 66% of teachers across the 35 items. Respondents lacked competency in interpreting, integrating and communicating assessment results.

Plake et al.'s (1993) research was followed by that of O'Sullivan and Johnson (1993). Their study used the TCAQ with 51 graduate students in a teaching program. The objectives of their research were to examine the alignment of TCAQ items, the assessment tasks, and the level of students' comprehension of the teacher competencies standards. Their study results illustrated that there was a high degree of alignment between the performance tasks against the standards which in effect, added support to the validity of the TCAQ instrument.

Early in the 2000s, Campbell et al. (2002) used a revised version of the TCAQ in their study involving 220 undergraduate students in a pre-service teaching initial teacher education program. The results demonstrated that pre-service teachers' competencies in the seven standards varied quite significantly. Given this result, the authors suggested that pre-service teachers were lacking in proficiency in the essential areas of assessment when they went on to be in-service teachers. In another project, which involved 67 pre-service and 192 in-service teachers, Mertler (2003) arrived at results that echoed those of Campbell et al. (2002). The results from Mertler's research are also consistent with results found in the studies undertaken by Plake et al. (1993) and Campbell (2004, 2005).

In 2004, the TCAQ underwent changes initiated by Mertler and Campbell. This resulted in the Assessment Literacy Inventory (ALI) (Mertler & Campbell, 2004, 2005). The objective of the revisions was to recast the test items into questions that referred to scenarios. Doing so was

intended to create a more authentic method of measuring the performance against the standards (AFT et al., 1990). Mertler and Campbell (2004, 2005) used the ALI to measure student comprehension of the standards. Their study demonstrated that teacher assessment literacy in essential areas was lacking. This echoed other studies performed on this aspect of assessment literacy research.

1.6 Teacher Assessment Literacy Research: The KSA Context

The previously discussed studies demonstrate that some instruments have been developed and used to measure the assessment literacy of teachers in the context of the different education systems that have professional standards for teachers. However, in KSA, studies measuring teacher assessment literacy have never been undertaken. Yet, assessment literacy is undoubtedly essential to the success of its educational outcomes so, this is one glaring gap that this study intends to fill.

However, a few studies around assessment more generally in KSA have been conducted to date and they could point to some implications regarding KSA teachers' assessment literacy. One study on assessment was a recent comparative study between the Science Education Diploma Programme at Jeddah in Saudi Arabia and a United States university in Iowa and their implementation of performance skills lessons (Al-Shikh, 2017). The study found that students of science teachers in Saudi Arabia were significantly different from those in the US. Specifically, greater familiarity with the scientific content, better implementation of skills lessons, better teaching strategies and assessment knowledge were noted in the case of US teachers compared to KSA teachers. Based on the findings of the study, the following recommendations were proposed: a specialised diploma program for teachers should be introduced and the number of specialised courses and the number of study hours, as the requirement for a diploma, should be

increased. Notably, this study also suggested that teacher assessment literacy needs to be enhanced.

Another study by Althobeti (2018) examined the assessment methods and assessment culture in higher education among a sample of 171 faculty members at Shakra University and their relation to the quality of the outputs of university education. The study aimed to measure and assess the abilities of learners as well as to determine the extent to which behavioural outcomes and learning goals were achieved. It found that faculty members focused on traditional methods of assessing students; the use of modern assessment methods was low; and the level of diversity in the use of evaluation methods was medium. Similarly, this study implied the need for teacher assessment literacy to be developed.

A study by Yousuf (2018), focused on determining the obstacles faced in the implementation of authentic assessment in the learning of Islamic Education in the schools of the Tabuk Educational Region, indicated that there were significant obstacles to implementing authentic assessment. The magnitude of the obstacles faced by teachers significantly differed from the level of obstacles faced by supervisors. In another study on assessment, Albalawi (2018) found that mathematics teachers in Riyadh have a positive attitude towards formative assessment and that they used formative assessment prominently in their assessment practices. However, the study also found that the teachers faced some of the following types of issues when implementing formative assessment in the classroom: insufficient class time; accommodating all students in a single large classroom, and; being able to utilise modern assessment procedures. While the above studies relate to assessment more broadly, they nonetheless seem to suggest the state of teacher assessment literacy in KSA, which is useful in the context where there seems to be an absence of research around Saudi teachers' assessment literacy.

Recent government educational reforms are the key to transforming KSA's educational system to prepare its citizens for the 21st century. Yet, there has been very little research conducted around teacher assessment literacy, even though sound teacher assessment practices are key to the realisation of KSA's educational aspirations.

There has been research that has identified several issues in teacher assessment practices in Saudi Arabian schools (Alanazi, 2017). These issues include using assessments that only focus upon shallow learning (Alotabi, 2014) and a perceived lack of knowledge of contemporary assessment skills taught in the initial teacher preparation programs.

In 2017, the Ministry of Education conducted a study of assessment training in universities. This study found that faculty members when teaching assessment focused on traditional methods of assessing students; gave very little training on the use of modern assessment methods; and, only gave medium exposure to using appropriate assessment methods to accommodate different levels of diversity. Such findings lend support to the argument that there is a problem with training teachers about assessment in initial teacher education programs and that these problems are exemplified once teachers start teaching in the KSA system. They also suggest that the MOE sees the solution being closely aligned with professional learning that is designed to increase teachers' skills in assessment (i.e., teacher assessment literacy).

Despite the educational reforms that include improving the assessment system and practices in KSA no study has been done on assessment literacy of teachers to date. While there has been some research around assessment, little is known about KSA teachers' assessment literacy and their views on the kinds of assessment skills and knowledge that are considered essential and ultimately impact on their assessment practices (Barnes et al., 2015; Deneen & Brown, 2016; Fulmer et al., 2015). For example, pressing questions such as 'What are the

essential assessment knowledge and skills for Saudi teachers?’ ‘To what extent are KSA teachers’ assessment literate?’ have yet to be answered. In KSA’s educational context in which assessment is seen to have a central role in delivering student outcomes, finding the answers to these questions is imperative.

Considering these discussions and given the general lack of research about teacher assessment literacy in the KSA and the importance cited in the literature of assessment literacy for teachers and by the Saudi Ministry of Education’s establishment of standards for assessment literacy, this study would seem to be very important and timely. First, it will establish an assessment literacy framework linking the standards for assessment literacy (NPST) with meaningful interpretations by Saudi teachers of what is most important to measure via elaborations of the assessment literacy framework. This study will then seek to modify the original TCAQ document (further refined and referred to as the ‘Teacher Assessment Literacy Questionnaire’ [TALQ]) into an appropriate measure of assessment literacy for use in the KSA. Doing so will then allow teachers in the KSA to benchmark their current levels of skill and knowledge about assessment literacy to the NPST and, through professional learning, improve those skills over time. It is therefore anticipated that the proposed study will make an important contribution to determining the level of assessment literacy of KSA teachers and then, if required, enhancing KSA teachers’ assessment literacy.

This would be an important contribution because studies such as those by Black and Wiliam (1998) and Hattie (2008) show significant empirical evidence that teacher assessment practices are one of the most influential factors in enhancing student learning outcomes. More recent studies (e.g., Barnes et al., 2015; Deneen & Brown, 2016; Fulmer et al., 2015) also add

evidence to support the important role teacher assessment practices play in improving student learning outcomes.

1.7 Development and Validation of the Assessment Literacy Framework and Questionnaire

In addressing the research questions that guided the present study, focus group discussions were conducted with study participants to finalise the elaborations produced through an analysis of the literature in the area and from the NPST. These elaborations then formed the basis for the revised assessment literacy framework and it is this framework that explicitly delineates the knowledge and skills, which study participants agree, that KSA teachers need to demonstrate that they have met the assessment standards described in the NPST. The elaborations that participants identified were considered essential for teachers to demonstrate that they could meet the requirements in the NPST and they directly addressed the first research question.

Three stages of validation evidence collection were then carried out to build a validation argument to support the assessment literacy framework and resulting questionnaire used in the study. The following stages give a brief overview of the various stages used in the study (the processes in the development of this assessment literacy framework and instrument will be expanded further in Chapter 3).

Stage 1: The first stage of this study required the development of an assessment literacy framework which provided for elaborations and articulations of the NPST standards. Focus groups of Saudi teachers were used for this purpose. Participants helped select only the most important standards, clarify their meaning and organisation and link them to specific knowledge and skills.

This was carried out by reviewing the NPST and teasing out further details and probable sub-standards into what are called ‘elaborations’ (i.e., the elaborations are the skills, knowledge and understanding needed by teachers to demonstrate the standards) in this study. This was done because the NPST, while specifying standards around assessment literacy, is arguably short on details that pertain to more specific knowledge and skills around assessment literacy. The validation of this assessment literacy framework is further discussed in Chapter 3.

Stage 2: Using the refined assessment literacy framework as a guide, the Teacher Assessment Literacy Questionnaire (‘original TALQ’) was selected to develop the modified version of the Teacher Assessment Literacy Questionnaire, (‘the MTALQ’), which is the measurement instrument used for collecting information about teacher assessment literacy in KSA. This was done via focus groups of Saudi teachers aligning, retaining, modifying, and constructing new items that reflect the Saudi context and that better align to the refined assessment literacy framework and hence the assessment literacy standards (NPST).

Stage 3: Following the validation of the original TALQ in Stage 2 (again, see Chapter 3 in this regard), it was administered to a small purposive sample of Saudi teachers to gain further validity evidence regarding the MTALQ instrument (see Appendix G for The MTALQ) and indicate how the results could be used to gain understanding of the current assessment literacy of those KSA teachers who chose to be involved in the study.

1.8 Significance of the Study and research Questions

The aims of the current study are:

1. Develop an assessment literacy framework that aligns the assessment knowledge and skills demanded of teachers by the KSA NPST.

2. Evaluate the efficacy of an assessment literacy questionnaire, the TALQ, (Teacher Assessment Literacy Questionnaire; Alkharusi, 2011; Mertler & Campbell, 2004; Plake et al., 1993) to determine the extent it is appropriate for measuring teacher performance in relation to the assessment literacy framework representing the NPST of Saudi Arabia.
3. Design, construct and validate an assessment literacy questionnaire (MTALQ) derived from the original TALQ that is more aligned to the assessment literacy framework and NPST and can be used to assess KSA teachers' assessment skills, knowledge and understanding.

These aims will be addressed by answering the following research questions:

1. What assessment literacy knowledge and skills do KSA teachers need to demonstrate that they have met the assessment standards described in the KSA NPST?
2. To what extent is the TALQ (the original TALQ) a valid measure of the assessment knowledge and skills required by KSA teachers to demonstrate that they have met the standards described in the NPST?
3. What evidence is available to support the validity of a new questionnaire (MTALQ) based upon the original TALQ, but augmented by items that better align it with the requirements of the KSA assessment literacy framework?

The first question aims to explicitly articulate the range of knowledge and skills that KSA teachers need to be able to demonstrate the stipulated teacher professional standards. This will be accomplished by establishing an assessment literacy framework which incorporates the skills and knowledge that teachers judge they will need to demonstrate to meet the needs of the assessment literacy standards as contained in KSA's NPST.

The second question is designed to assess the extent to which the assessment literacy framework can be measured by an existing teacher assessment literacy questionnaire within the KSA context. It uses the previously developed original TALQ (Alkharusi, 2011; Mertler & Campbell, 2004; Plake et al., 1993) for this purpose.

The third question focuses on developing a teacher assessment questionnaire, the MTALQ, (by modifying the existing TALQ questionnaire) that is tailored to the requirements of the KSA teacher professional standards. Once developed, validity evidence for its use will be presented, showing that the MTALQ is a suitable measure for the assessment literacy of KSA teachers.

Potential contributions from this study include first the development of an assessment literacy framework that is specifically based on KSA teacher professional standards and that can inform policymakers and other educational stakeholders about the kinds of assessment skills and knowledge that teachers need to be able to perform and know respectively. Second is the development and the validation of a questionnaire that can effectively assess teachers' assessment literacy in the KSA educational context. Third, because there is no explicit research on assessment literacy relative to the NPST, this study should contribute substantively to all future assessment literacy research. Finally, once developed, the refined assessment literacy framework or questionnaire may be used by other Arab speaking countries in the Middle East and beyond. Indeed, the study outcomes are generalisable beyond the KSA context and may be of value internationally. For example, in terms of its methodological contribution, the proposed study's development of an assessment literacy framework from a set of teacher professional standards and associated assessment literacy questionnaire could be replicated in other international contexts.

1.9 Organisation of the Thesis

This first chapter has provided the background of the study and highlighted the study problem, study aims and study questions that guide the study. Against this context, the purpose of the study and its significance have been stated. The arguments as to why the study is imperative and how the study can contribute to the articulation of an assessment literacy framework and the development and validation of a bespoke teacher assessment literacy instrument for KSA teachers have also been put forward. This chapter also identified the study's significance and its potential contributions on various levels and its generalisability to contexts beyond the borders of KSA.

Considering the study aims and study context, Chapter 2 first presents a brief overview of the social, economic, political, and religious context of the KSA and its educational system. This includes its educational system structure and governance and a discussion of Saudi Arabia's teaching and teacher education policies. This chapter concludes with the review of literature on teacher assessment literacy and its related areas as well as identifying assessment literacy studies that have been conducted in the context of the KSA.

Chapter 3 explicates the methodology and methods employed to address the study questions. Justifications are presented for the selection of the study methodology, methods, study design, data analysis and the study participants. Chapter 3 also explicates the processes involved in the various stages of the development and validation of an assessment literacy framework and a questionnaire to assess KSA teachers' assessment literacy.

Chapters 4-7 present the results and findings, discussion and conclusions of the study drawing on the data collected, the analyses of this data and review of field notes taken during the investigation.

Chapter 2: Literature Review

2.1 Introduction

This chapter starts with a brief overview of the social, economic, political and religious context of the KSA and its educational system. It includes the structure and governance of the educational system and a discussion of Saudi Arabia's teaching and teacher education policies. It continues with a review of literature on teacher assessment literacy and its related areas, as well as assessment literacy studies that have been conducted in the context of Saudi Arabia. To understand the need for teacher assessment literacy in KSA, it is necessary to understand the educational context of KSA as well as the research links to assessment literacy relevant for this context. Providing such a context is pivotal for addressing the aims of this thesis.

2.2 Saudi Arabia: A Social, Economic and Political Overview

King Abdulaziz Bin Saud established the KSA, also known as Saudi Arabia, in 1932. It has a total area of 2.15 million square kilometres and is the largest country on the Arabian Peninsula (Ministry of Culture and Information, 2013). It comprises 13 provinces with Riyadh being its capital. KSA is one of the fastest-growing populations in the world, with an estimated 30 million people and its annual population growth rate is 3.5% (Onsman, 2010). About 90% of its people are of Arab origins with the rest comprised of Asians and Africans (Alhawsawi, 2013). The overwhelming majority of its people are Muslims, as is reflected in the official religion of Islam and official language of Arabic. Furthermore, and perhaps most significantly, the nation is recognised as the cradle of Islam and the guardian of the two holiest sites for Muslims, the Masjid al-Haram in Makkah and the Al-Masjid an-Nabawi in Medina. It is hardly a surprise that Islam heavily influences Saudi Arabian culture, habits, and beliefs given how important Islam is to Saudi citizens' daily life. Saudi Arabia's geography, history, culture, and

religion play an important role in being able to understand and interpret education policy and procedures, especially as they relate to teachers and teaching.

KSA is the wealthiest country in the region and much of its wealth and economy is founded on oil. Its oil industries generate about 80% of its national income and is the main driver of its rapid growth, which began in the 1970s when the oil economy first began to thrive. The immense revenues that came from its oil economy allowed KSA to fund infrastructure developments, such as transportation and housing. US \$301.6 billion was allocated to its 2014 state budget, demonstrating the wealth it possesses and the revenues that have been generated from oil over the decades (Ministry of Finance, 2014). Such wealth from oil not only transformed the country economically but also turned it into an economic and political power in the region (Alkharashi, 2012). KSA has 25% of global oil reserves and is a leading oil producing country. Its economy also remains one of the fastest growing in the Middle East and North Africa. Moreover, being the custodians of the two holiest places for Muslims, it receives millions of international visitors from across the world who perform religious pilgrimages, further contributing to the local economy.

The King of Saudi Arabia, which is a monarchy, serves as the head of state and the Crown Prince is the prime minister of the nation. Along with The Crown Prince, who is the next in line to the throne and serves as the King's deputy, the King assisted in ruling the nation by the Council of Ministers, also known as the King's Cabinet. The Cabinet is made up of 22 ministers who oversee the various departments like education, health, and finance and hold administrative power over the formulation and execution of national policy in all areas. The Consultative Council, that the King himself appoints and is comprised of 150 members, also supports him. They provide counsel to both the King and his Cabinet on issues such as education, the economy,

and Islamic affairs, and reviews various national policies. Although the Cabinet and the Council wield influence and authority in their respective roles, the King has the final say and is therefore seen as the final arbiter on national matters (Royal Embassy of Saudi Arabia in Washington, 2014). Being an Islamic country, KSA's constitution and legal framework are founded on shari'ah (Islamic law) and much of the country's governance is underpinned by Islamic tenets.

2.2.1 An Overview of the Education System in Saudi Arabia: Historical and Recent Developments

In Saudi Arabia, education has always been and still is centred on Islam. Before 1925, most education was provided through mosques and religious schools, where students were largely taught Arabic and the Quran. However, in 1925, the Directorate of Education established a formal educational system. Established by King Abdulaziz before he founded KSA in 1932, the Directorate was tasked to establish formal schools across the country.

The first government-run schools were established in 1930, but at that time, only male students were allowed to enrol (Alsharif, 2011; Wiseman, 2010). It was only in 1960 that government schools were established for females. The move initially attracted criticism from groups who believed that non-religious teaching would be meaningless for women. Such views did change subsequently when the education of females began to attract support (Almutairi, 2008). All schools in Saudi's educational system are single-sex and students are enrolled into schools according to their gender. While this may seem unusual to the uninitiated observer, such policies can be attributed to Islamic principles and beliefs and Arabic culture, all of which discourage the free intermingling of males and females. In Muslim-majority nations in the region of the Middle East, schools for one gender are commonplace. For instance, Bahrain solely has single-sex secondary and elementary schools (Fryer & Levitt, 2010). There are no substantive

differences between the education of boys and girls in Saudi Arabia, despite the country's single-sex school system. All students regardless of gender undergo the same levels of schooling (i.e., primary, intermediate and secondary), are provided with common school facilities and infrastructure and learn the same curriculum with minor differences catering to gender (Al-Johani, 2009).

Since the beginning of its formal education system in 1925, the government of Saudi Arabia has continued to recognise the important role education plays in the development of the country. In recent decades, this has become increasingly evident. The Saudi government has consistently invested in the various sectors of its education system, allocating a significant part of its annual budget to education. For instance, in 2013, education was allocated 25% of the total budget of more than US \$300 billion (Ministry of Finance, 2013) - a percentage which far outstrips the expenditure on education of most developed countries. Furthermore, the MOE is the only government ministry that has not experienced budget cuts. Indeed, the budget for education has been rising ever since 2000, demonstrating the importance that the Saudi government places on it for the advancement of the nation.

The government provides free education to all Saudi residents from pre-school to university as part of its investment in education (Alamri, 2011; Alsharif, 2011; Onsman, 2010). Schools and colleges are also provided with free books along with health services and information technology infrastructure. Moreover, to encourage citizens to complete higher education, undergraduates receive a monthly governmental allowance of about US \$250. Additionally, the government provides financial aid for postgraduate studies abroad and in 2013 alone, the government provided scholarships to more than 148,000 citizens to pursue higher study in countries such as Australia, the UK and the US.

The government's dedication to education is demonstrated by recent educational changes in the KSA. Education is seen as a key instrument in transforming the country in the 21st century. Education is also considered to be a vital instrument in reducing the Saudi economy's dependence on oil and supports the development of a knowledge-based economy, which can be a more sustainable means of economic growth and progress in the country (Bashehab & Buddhapriya, 2013).

One significant project launched earlier in the millennium is the \$2.4 billion King Abdullah Public Education Development Project, which is also known as 'Tatweer' (meaning 'to develop'). Launched in 2007, its objective is to cultivate a labour force characterised by productivity and efficiency in an adaptive and creative economic system and improve the education system (Tatweer, 2008), thereby contributing to the improvement of the general school experience and fostering a generation of KSA citizens ready to take an active part in the development of the country and its people. To achieve its objectives, Tatweer called for new teacher professional learning initiatives, the implementation of new classroom learning technologies and greater teacher autonomy in state schools.

In 2013, the KSA founded the Education and Training Evaluation Commission (ETEC), which specialises in assessment, measurement and evaluation. ETEC published its first handbook for the assessment and evaluation system in 2014 (see www.etec.gov.sa), and in 2016, ETEC partnered with King Abdullah bin Abdulaziz's Public Education Development Project (Tatweer). This partnership led to the creation of the NPST in KSA, which encompasses three areas: Professional knowledge, Professional practices (assessment is the third component of this level) and Professional values and responsibilities.

In 2016, KSA officially announced the ‘SAUDI VISION 2030’, an initiative geared towards changing and developing all aspects of the country. This Vision also promised substantial reforms to the entire education system and in 2019, the KSA’s MOE reformed the structure of the standards of public education and the curricula itself. Such changes were aimed at raising KSA’s educational system to be in line with other progressive educational systems around the world. It also aimed at reforming assessment practices to create a more assessment-based culture in the educational system (Assalahi, 2021).

In January 2019, another educational development was announced by the Minister of Education at an educational leadership conference in which crucial problems in the assessment system in primary schools were identified. This led to a review that saw the Council of Ministers authorise the MOE to reform the continuous assessment system and in 2019, the Ministry announced that the continuous assessment system would be supplemented by central examinations for primary school levels at grade levels 3–6 (MOE,2022). In 2019, the MOE also reformed its national curriculum and designed new standards to ensure that they are consistent with the global educational imperative of developing 21st century competencies.

With the slew of reforms in recent years, it can be said that education in KSA is at an inflection point. Governmental educational accomplishments such as providing universal access to education across a highly dispersed and vast school-age population constitute a significant milestone. Yet, this seems to have come at a cost. Providing high-quality education has seen the resources of teachers and education administrators pushed to the limit. For instance, school leadership staff such as principals are appointed without receiving sufficient training to manage and lead their schools, while various bodies in the educational system that are tasked to review and support education are facing much difficulty in fulfilling the ministerial expectations in the

face of growing workloads (Organisation for Economic Co-operation and Development [OECD], 2020). In short, the reforms have left a shortage of trained educators at all levels.

The future implementation of a new school evaluation framework presents a pathway for KSA to significantly re-examine how its school system is administrated for the purpose of raising the standards of education across the kingdom. In a report on KSA education, OECD suggests ways in which the education management system can be improved (OECD, 2020). One improvement could be found by reforming the school system governance. This would entail shifting the focus of educational bureaucratic bodies such as the MOE, Directorates and Education Offices from merely governing schools to supporting them in making improvements as effective and successful educational organisations. A second potential improvement lies in restructuring KSA's current system that informs its school evaluation and quality assurance. OECD found that KSA's educational initiatives are consistent with internationally sound practices. However, it suggests that these initiatives need to be informed and premised on an explicit vision for the new Saudi Arabian school that is accompanied by essential performance indicators of progress. The report's final recommendation involves school leadership, such as principals and teacher leaders, which is integral to the process of the improvement of schools. In KSA, it was found that such roles and capacities are not well developed and lack the essential systematic support (OECD, 2020).

Despite educational progress made in KSA, its extensive but inequitable system presents challenges in terms of the deployment of educational resources. For instance, a significant number of principals must be assigned around the kingdom and are tasked to run schools that are often small and isolated. Challenges arise with the deployment of a huge force of administrative staff who are tasked to manage governance at all levels of the educational system. The

requirement of recruitment of staff has raised a substantial challenge in view of the capabilities of the staff. The shortage of qualified staff to assume the various positions, which are often in remote areas, leads to the hiring of underprepared staff. This condition is additionally applicable to school administrators, who have the responsibility of assisting in the growth and improvement of their institutions.

It must be highlighted that the fragmented nature and extent of the KSA education system exacerbates the challenge of providing effective and efficient support and quality assurance. A total of 10,000 supervisors are appointed to oversee the evaluation and development of schools throughout the kingdom, but owing to a lack of preparation, their job is mostly focused on ensuring administrative adherence rather than leading and fostering school reform.

School system governance entails a vision for schooling that includes the why, where and how a country hopes to increase schooling achievement. This also includes how responsibilities for realising this vision are assigned at various government levels in an educational system. Effective governance depends on having clear objectives that are in line with the group efforts of many stakeholders, even when responsibility for school reform is distributed differently among nations (OECD, 2016b). According to more and more evidence, decentralised executive procedures for school-level decisions about pedagogy and curriculum, along with strengthened leadership capacity and unique accountability frameworks, are associated with better student outcomes (Burns et al., 2011). While the current educational system of KSA has its challenges and is traditionally centralised, such evidence certainly offers optimism for the improvement of its educational system if it is prepared to undertake reforms and provide appropriate capacity development.

2.2.2 Education Structure and Management in Saudi Arabia

The importance the government accords to education in KSA is exemplified by the rapid growth of the number of schools—from about 3,000 in 1970 to almost 35,000 in 2013. Moreover, in that period, the student population rose from about half-a-million to more than 5.2 million students (MOE, 2014; Oyaid, 2009). The public education system consists of three levels following pre-school. These are primary school (Years 1 to 6), intermediate school (Years 7 to 9) and secondary school (Years 10 to 12). Citizens begin their education at seven years old and complete the final stage at 18 years old. However, compulsory education is only up to the intermediate level in which citizens must complete the minimum of primary and intermediate schooling.

The main government agency responsible for directing and managing public schooling in the KSA is the MOE. Founded in 1954 when it took over from the Directorate of Education, it is also responsible for managing adult education, special education, junior college and teacher education. It is responsible for school building and the provision of school infrastructure (Alquraini, 2010) as well textbooks (Al-Otaibi, 2004; MOE, 2013; Oyaid, 2009). MOE manages the education system through a system of Directorates and Education Offices and works with educational offices found in each of the 64 educational districts across the country, in which they function as the intermediary between schools in the district and the MOE in educational matters (Alsharif, 2011). MOE has delegated its management duties among other Directorates nationwide according to each Directorate's size and responsibilities. Some of these may be further divided into separate Education Offices. All Directorates and Education Offices are structured in a similar fashion to the national ministry, consisting of departments, divisions and reporting units.

The MOE management structure can be described as decentralised, but executive decision-making seems to be centralised to a great extent. This is due, among other factors, to the fact that MOE oversees almost all national policy choices, which are subsequently passed on to the lower tiers of government. This is where such policies are seen as directives and mandates to be implemented. For instance, policies such as those around its teachers, school evaluation and learning are formulated at national level and implemented in uniformity throughout the kingdom. The vertical, top-down structure of the KSA educational system also means that collaboration across horizontal levels is minimal.

While the distinct vertical leadership among the Directorates and Education Offices facilitates the smooth cascading of policy directives, such a top-down structure is also a source of obstacles in terms of coordination at the local level. For example, the development department under the governing Directorate often receives reports from the teacher professional growth unit in an Education Office. Through the general manager of the Directorate, it in turn submits reports to the National Institute for Professional Educational Development. The Education Office or Directorate's other departments and units are not included in the information exchange, even though these departments and units' core responsibilities may involve important topics like curriculum and assessment, which are closely related to teacher professional development. As a result of this system, schools must liaise with the supervisors, teacher deployment and student affairs divisions of the Education Office. This is because schools do not communicate with their Education Office through a sole point of contact that connects them to all the relevant departments (OECD, 2020). As a result, structural inefficiencies such as duplication of efforts and the decreased coordination, both of which ultimately disadvantage schools, are created.

In terms of data management of the education system, there are two major databases that store and facilitate access to a range of administrative information relating to schools, staff and students. NOOR stores demographic information and performance at the different levels of student, teacher, school and the system; FARES stores human resources information such as staff remuneration. All system participants have access to both databases, and the data are often utilised for administration and monitoring.

Another part of the KSA education ecosystem is ETEC—an independent organisation, which stipulates the various benchmarks and protocols related to school evaluation. The primary aim of ETEC is to offer technical leadership that is linked to defining and measuring the quality of education in KSA (OECD, 2020). The autonomy accorded to ETEC along with its know-how means it can design various tools that are essential for tracking and enhancing the standards and quality of schooling. Next to the new approach to school evaluation, such tools also include professional standards for principals and teachers and student learning, together with various resources that aim to assess quality and develop dedicated evaluators to conduct such assessments.

However, an OECD (2020) report on KSA education indicates that while the various sub-units under ETEC are tasked with the development of such resources, close collaboration among them seems lacking. The report also suggests that there are clear avenues for the various units under ETEC to move more in concert in terms of coordination and collaboration. For instance, the way the quality of teaching that takes place in schools is assessed can be aligned with the new standards for teachers and principals that are set along with standards on school management and pedagogical leadership.

To do this, ETEC is creating a new school assessment system to replace the old ones. The new model includes academic standards that emphasise teaching and learning, results from learning, school management and the learning environment. All schools will be expected to complete self-evaluations in accordance with the new framework, using the recommendations made by ETEC. Schools will then be given the necessary support according to a rating on a scale of 1 to 4 that they assess themselves against school quality benchmarks. For instance, schools that rate themselves as either 1 or 2 will receive support in the development and implementation of an improvement plan, while schools that rate themselves as 3 or 4 would be eligible to undergo an external ETEC evaluation and seek official accreditation.

In addition, a battery of evaluation tools are being developed that are aimed at guiding both external and self-evaluation of schools in conjunction with the new evaluation framework benchmarks. The updated framework also intends to transfer oversight of school evaluation from supervisors to a panel of impartial evaluators who will be chosen and trained by ETEC.

Finally, next to ETEC, under MOE's Agency for Development and Planning is the Education Policy Research Centre. This Centre has a clear evaluation and researching aim and is tasked with assisting in performance monitoring as well as reviewing education policies and international best practices. It consists of staff who are suitably qualified and can be employed to review existing policies as well as project the impacts of planned educational reforms.

2.2.3 MOE and Vision 2030: The Role of Education in Remaking a Nation

In 2016, the KSA government launched a set of comprehensive reform initiatives meant to affect all key segments of society, with the aim of decreasing KSA's excessive dependence on oil and developing a diversified, vibrant and enduring economy instead. To accomplish these aspirations, several programs were put in place. One of these was the Human Capability

Development program (HCDP), which attempted to enhance the nation's educational system through the creation of a well-trained, capable and skilled labour force ready to meet the challenges of the information-based economy of the 21st century. Clearly, the Vision campaign has already stirred movement within all divisions of the educational system.

The MOE is the entity charged with the realisation of this Vision and achievement of the goals established by the National Transformation Program (NTP), which tracks the ministry's progress as it attempts to attain the goals. It is worth noting that the long-term goals of the Vision and the NTP are communicated to the different stakeholders involved in the education sector. Nonetheless, it ought to be emphasised that the KSA government has allocated significant funds to improve its educational system in the last two decades. According to the World Bank and ETEC (2016) and the United Nations Educational, Scientific and Cultural Organization (UNESCO) Institute for Statistics (2020), KSA spent 7.8% of its gross domestic product (GDP) on education in 2015, which was much more than the OECD average of 5.2%. Additionally, according to UNESCO, KSA's enormous investment in education has led to the tremendous advancement of its educational system. Access to education at the national level, where enrolment rates in basic and secondary school have virtually attained universal levels, is one noteworthy achievement (UNESCO Institute for Statistics, 2020).

Although progress has clearly been made in access to education, increased enrolment in education has not translated into corresponding progress in student learning and the quality of education. To measure outcomes linked to the quality of education, KSA has been taking part in several international testing programs. These programs to measure system-level outcomes include participation in the Trends in International Mathematics and Science Study (TIMSS) and

Progress in International Reading Literacy Study (PIRLS). More recently, KSA has also begun participating in the Program for International Student Assessment (PISA).

While there has been improvement in access to education in KSA, international assessment results such as TIMSS, PIRLS and PISA indicate that most of its students are lagging those in other countries in terms of literacy and numeracy. National assessment data corroborate this, as they show that many KSA students are relatively poor in literacy and numeracy skills (OECD, 2020). KSA students' results from these international testing endeavours indicate that while national educational investment in the country has been high, such expenditure has not materialised into high returns on investment in the shape of improved educational outcomes nationally. For instance, the results of PISA 2018 showed that KSA students continuously fared poorly in reading, mathematics and science literacy in comparison to students from other OECD countries. KSA students also obtained poor results in the subjects of science and mathematics literacy in comparison to the other countries of the region, such as the Middle East and North African countries (MENA), who took part in PISA. In terms of reading proficiency, more than 50% of KSA students were shown to not possess the minimal level of proficiency required to function fully in society, compared with less than 25% of students in the OECD countries (OECD, 2019). Such statistics unmistakably indicate the urgent need to enhance KSA students' learning outcomes to achieve the national economic goals established under Vision 2030.

An OECD (2020) report suggests that raising overall outcomes across KSA will entail addressing the existing inequities that beset its education system. In the PISA national results in 2018, KSA students from city schools fared better than those in schools in rural areas. A similar picture emerged when results were compared in terms of students' socio-economic status: the bottom 25% of KSA students fared 74 points poorer than the top 25%. This gap between the two

groups of students is thought to be equivalent to three years of schooling. Moreover, it is important to note that student socio-economic background in KSA seems to be more prominently linked to performance than is the case in other MENA countries. The OECD (2020) report indicates that in KSA, 12% of the variance in reading performance can be attributed to socio-economic status, while across MENA countries this number is 9%.

Such findings also indicate that demographic factors have an influence on KSA learning outcomes. In KSA, the increased access to education has been accompanied by increased population growth and urbanisation. For instance, in 2018, KSA had more than a quarter of its population under the age of 15, while OECD countries had 18% of their population under 15 years old. Moreover, three regions out of the country's 13 account for more than 60% of all general education students, which leads to an imbalance in the network of schools. As a result, more and more little schools are opening in poorer and rural regions, while schools in wealthier and metropolitan areas are becoming increasingly overcrowded.

What seems to have compounded the challenges of realising KSA's overriding reform mandate is the fact that there seems to be a lack of articulation of the vision of quality schooling (OECD, 2020). This is essential because an articulation of a common vision can provide a collective reference for the various agents and stakeholders in the KSA education system that are working towards the aim of enhancing the quality of schooling. In addition, a clear vision of schooling will be beneficial alongside a school evaluation model because it provides the larger aim that schools can strive towards while lending a sense of congruence to the other distinct aspects of education quality. According to OECD (2020), such a goal is frequently focused on student achievement and how a successful school develops students' intellectual, emotional and

social aspects. Thus, while the broad educational goals of Vision 2030 are purposeful and commendable, they have not been matched with an explicit articulation of quality schooling.

OECD (2020) also highlighted that the NTP has not assisted schools in comprehending and attaining student achievement goals, even though it is the entity that stipulates student attainment goals on a national level. More importantly, the NTP has also been unable to help schools in understanding and mobilising the aim of schooling under Vision 2030. Because of this, Directorates and Education Offices do not have a reference point from which to prioritise their actions when working with schools and, consequently, each Directorate lists its own priorities to define the key elements of effective schooling.

The OECD (2013) report further indicates that such a situation could mean that important national education objectives lack the attention needed. For example, the aims of consolidating basic student competencies and attaining fairer student outcomes have not been given the appropriate focus. A fundamental aim of all education systems is the development of student literacy and numeracy. Besides this, another aim of a national educational system is to realise equity in terms of educational outcomes for all students. However, OECD does not suggest that this means ensuring all students perform at matching standards. Instead, all students should have access to a high-quality education and equitable opportunity to meet the minimal requirements for education, regardless of their origins (OECD, 2013). In the case of KSA, this calls for clear policies and the formulation of specific targets that are aimed at narrowing the differences in student outcomes, along with safeguards such that people and remote areas that are at risk will obtain suitable help and support.

2.2.4 Teaching and Teacher Education Policies in Saudi Arabia

One of the distinct characteristics of the teaching profession in KSA is that the teaching workforce can be classified as large. Although a *prima facie* comparison indicates that KSA has a lower student-to-teacher ratio than OECD countries, the way teachers are classified in the civil service in KSA skews this ratio (OECD,2020). Teachers are classified as civil servants with permanent contracts. However, this classification also applies to practically all other staff hired in the education system, including principals and administrators. Because of such classifications, not all the half a million civil servants captured in the system are classroom teachers; a minimum of 10% of this number are school managers, administrators or supervisors. Therefore, the official student-to-teacher ratio in KSA was given as 11.7 at the primary level in 2016 and 11.0 at the secondary level in 2014, whereas other OECD countries had on average a ratio of 13.4 at the primary level and 12.1 at the secondary level (UNESCO Institute for Statistics, 2019). However, the actual student-to-teacher ratio in KSA is greater than the numbers that are officially given.

In addition, the seemingly low student-to-teacher ratios often do not mean small class sizes across KSA. This is because classroom teachers are not usually allocated in the most productive manner. For example, schools in rural regions, which are often small, usually have a far lower student–teacher ratio than schools in urban areas, where teachers frequently manage classrooms with very high student numbers. Additionally, the single-sex educational system causes its teacher demographics to differ from those of other countries. In 2015, this single-sex school system translated into a teacher gender breakdown of almost 50:50 in the KSA. However, in OECD countries there are more female than male teachers to the tune of almost 80:20 (World Bank & ETEC, 2016).

In relation to other OECD countries, which have a minority male teaching workforce, the comparatively high number of male teachers needed in the KSA education system creates systemic tensions in identifying and employing qualified male teaching candidates. One main reason for this difficulty is that male candidates usually have professional opportunities in other sectors of the economy. In addition, the Teaching and Learning International Survey (TALIS) 2018 data indicate that KSA teachers are on average younger than those in OECD nations. In contrast to OECD nations, where the average age of lower secondary teachers is over 44 years old, the KSA's typical teacher is under 38 years old (OECD, 2020). There is a great and urgent need to develop KSA instructors and the teaching profession in general since these instructors are frequently expected to remain in the profession for a large portion of their working lives. This is because KSA teachers are younger and are engaged on a permanent basis. Teacher management is centralised, like the other centralised parts of KSA education government. It does not, however, have deliberate coordination. For example, ETEC oversees creating national teacher standards, whereas MOE oversees and assists the teachers in the system. However, when it comes to the hiring of teachers, this role is undertaken not by MOE but by the Ministry of Civil Service, which was recently amalgamated with the Ministry of Human Resources and Social Development.

The entry requirements into initial teacher education in KSA are low (OECD, 2020). This seems to be consequently reflected in the low status of teaching and the small enrolment of students into Initial Teacher Preparation (ITP) courses (OECD, 2020). When compared with applicants of other degree programs, applicants of teaching degree programs often have lower scores in the General Achievement Test (GAT) and Scholastic Achievement Admission Test (SAAT). One reason why the entry standards are considered low is that most ITP providers do

not stipulate entry requirements that are fixed and explicit, or else these requirements are very broad—such as having a good behaviour record and health—and not directly linked to more specific ability and aptitude for teaching.

In addition, students who complete upper secondary education and pass the tertiary entrance test could be admitted into ITP programs in KSA. In many OECD nations, such as Australia, Korea and the Netherlands, candidate selection for entry into ITP programs is much stricter, and candidate test results are expected to be higher than the average scores on standardised tests or other entry tests required by the ITP providers (OECD, 2019). These practices are not like those of many of those nations. Another factor that compounds the challenges faced by KSA education is that the admission of applicants into its ITP programs is not balanced against the demand for teachers in the system, which consequently creates inefficiencies in staffing and resource deployment.

Some of these challenges contribute to the perception that teaching is not a demanding career in the KSA (Alnahdi, 2014). Furthermore, research suggests that teacher applicants are not driven to join the profession because of their love for teaching or passion for education, but rather by the job security it offers over one's lifetime. For instance, the TALIS 2018 data demonstrate that when ranked against countries across the world, teachers from KSA show the greatest correlation between opting for teaching as a first career and the motivation to join the profession because of its job security. Although the profession offers job security, in terms of monetary rewards it is considered low when compared with other professions in the private sector. This is because the compensation structure for KSA teachers, administrators and supervisors is flat and wage increases are only granted based on seniority, not on ability or performance.

According to the OECD's report on KSA education (OECD, 2020), the efficacy of its educational reform will be determined by the efficacy of its initiatives that are geared towards reforming teaching and the teacher profession in general. To put it another way, strengthening the teaching profession and enhancing the standard of learning are crucial for KSA's educational reforms to be effective. This observation is not surprising given that much research has previously established that one of the most significant factors in educational success is indeed teachers and their classroom practices. However, for KSA this may not be so straightforward.

Traditionally, ITP in KSA can be described as relatively easy to access, while its Pre-service teachers could not achieve high quality learning (OECD, 2020). Furthermore, the way teachers were evaluated compelled teachers to practise teacher-centred pedagogical practices that were largely based on student rote learning and memorisation, which comes at the expense of the development of essential student skills. Nonetheless, recent initiatives show that the KSA government is working towards raising the quality of teaching and the profession in general. Some of these reforms include the development of the new NPST and a new post-graduate ITP program as well as the redesigning of the role that teacher supervisors play. As OECD (2020) points out, such initiatives are more aligned with other teacher policies in the international context that are aimed at raising the quality of teaching and teachers, enabling the creation of a more competent and driven cohort of teachers in the country.

2.3 Curriculum and Assessment

MOE presides over the Curriculum Department, which is responsible for the design and implementation of the national curriculum, unifying the education system throughout all the districts of KSA (MOE, 2013). Additionally, the department oversees creating the many required textbooks that are utilised for the various levels of education and disciplines.

All levels of schooling in KSA follow a common academic year, which is broken into three semesters. Unlike most countries in the world, the school week in KSA is from Sunday to Thursday, with Friday and Saturday being the weekend. At primary school (Years 1–6), students have six 45-minute classes a day. In Years 1–3, students learn Arabic language, Islamic studies, mathematics, science and art. In Years 4–6, new subjects such as English and social studies are introduced. There are no final examinations at the end of Years 1 to 2. As a result, teachers employ continuous assessment to monitor students' progress at different levels (Alafaleq & Fan, 2014). However, students at Years 3 to 6 sit for final examinations at the end of each of these levels.

The MOE stipulates the set of assessment criteria and standards that the continuous assessments are based on. Each subject includes a set of standards based on the abilities and information that are required of students. For instance, teachers use rubrics that outline the numerous requirements that serve as evidence of students' proficiency in mathematics to evaluate students' performance. Students are generally graded according to their level of mastery: a distinction is awarded to students who have shown mastery of all the knowledge and skills required; a credit is awarded to students who show mastery of 66% of the knowledge and skills; a pass is awarded to students who show mastery of 33% of the content and those who fall below this fail the assessment and are not able to progress to the next year level (Alafaleq & Fan, 2014).

Intermediate students take final exams. Students must pass the tests that teachers assign to advance from the first intermediate level to the second intermediate level. These examinations typically assess students' understanding of what was taught in class and topics contained in the textbooks. Every subject has two different types of assessments: mid-term and final exams. The

final or end-term exam accounts for 60% of the final grade, whereas the mid-term assessment accounts for 40% of the grade and includes class questions and participation. To pass a unit, students must have a minimum result of 50%.

The last stage of schooling in KSA is secondary education, which is also the pre-university phase, lasting three years (Years 10–12). Upon the completion of the first year of secondary education, students are offered the choice of pursuing many pathways such as the arts or science pathway. Mathematics, chemistry, geology, physics and a few non-scientific courses including Arabic, English and Islamic studies are all included in the category of science. Subjects that comprise the arts pathway include Arabic studies such as Arabic literature and linguistics, Islamic studies that include learning the Quran and Islamic laws, social studies and English. As with the intermediate level of schooling, students sit for mid- and end-term assessments, which carry similar weightings.

KSA students currently have two national examinations at the end of their upper secondary schooling as a pathway to tertiary education, namely the GAT and SAAT. However, OECD found that these examinations are not well aligned with the curriculum (OECD, 2020). This lack of alignment means that the examination system is unable to effectively support the implementation of the curriculum. Moreover, it was found that the examination items assess a small range of skills and are, at times, internally incongruent. Possibly for this reason, the examinations see very high student success rates, which discourages students from putting in effort and challenging themselves.

To improve national examinations, OECD (2020) recommends that in the short term, KSA should develop and implement one examination that is not only better aligned with the curriculum but should also be used to mark the completion of upper secondary education as well

as being used to enter tertiary education. While such an examination undergoes development, the MOE could invite assessment experts to embark on a systematic and detailed review of the GAT and SAAT in which the relevant changes are affected as necessary.

While traditionally, the KSA educational system has been dependent on textbooks for content teaching and assessments, there have been recent endeavours to update the curriculum and assessment methods. For instance, ETEC recently designed a national curriculum framework that emphasises competencies and skills alongside a National Assessment Program (NAP), which functions to assess students using these revised learning outcomes. The NAP aims to track national student achievement. At the same time, the MOE conducted a large-scale assessment for the purpose of educational accountability. This assessment also operates at the level of the Directorate as it was used to monitor the performance of the students. Nevertheless, the OECD (2020) suggests consolidating the assessment procedures in the KSA by broadening the scope of NAP to provide instructors and school officials with valid data regarding students' performance and achievement of the learning outcomes specified in the curriculum. The findings of the students' performance on this exam should be used to raise their understanding of the standards set forth in the newly established curriculum. ETEC should actively participate in monitoring NAP in its role as the national assessment authority and access the aid it needs to fulfil its mandate.

The curriculum framework recently developed by ETEC is generally in compliance with teachers' and schools' standards (OECD, 2020). It is likely to orient the curricula to common objectives and play a key role in the improvement of material development. Nevertheless, according to the OECD's assessment of education in KSA (2020), the elements of the new curriculum framework are not all coherent, thus potentially making it difficult for teachers to

comprehend and to implement effectively. Additionally, Tatweer's educational resources might not be adequate in view of the significant variability in teachers' background and experience in assessment. To circumvent these shortcomings, OECD (2020) points to the need for an evaluation and further solidification of the framework prior to its launching. Furthermore, to ensure successful implementation, communication with teachers regarding the new framework and relevant instructional material production should be encouraged and sustained. The OECD (2020) further advised that a continuing review of the curriculum be conducted, so that it is in step with the latest advances in assessment and skill development.

Despite efforts to modernise curriculum and teaching strategies, instructors continue to be heavily reliant on the resources at their disposal, employing predominantly ministry-mandated textbooks without customising the contents to their students' requirements. This is mainly because of the present performance evaluation system, in which teachers are judged not by the learning outcomes achieved, but by the amount of the content that has been covered. As a result, OECD (2020) suggests a school-based appraisal system that is consistent with the intended learning outcomes and new expectations of the curriculum. In the same vein, OECD recommends that instructors' professional learning also reflects the goals of the new curriculum by emphasising modern practices in teaching and learning. Realising the lack of a unified framework to assess students' achievement, the MOE and ETEC created their own assessment system, yet this was hampered by the absence of coordination. To develop a more consistent and coherent student assessment framework, a national framework is necessary, outlining the goal, methods and relationship between assessments administered locally and nationally (OECD, 2020).

Concerning classroom assessment, is primarily summative, and to rely predominantly on rote learning and multiple-choice questions (OECD, 2020). Moreover, teacher development was not conducive to the improvement of assessment practices, while instructors' evaluative judgements were largely inaccurate and unreliable owing to the absence of moderation practices.

In addition, it was signalled by the OECD (OECD, 2020) that there should be less time spent testing, first by watching how long students spend on examinations and then making changes in response to what is learned. Teachers' diagnostic assessment should be developed, while diagnostic assessments given by supervisors should be removed (OECD, 2020). Additionally, teacher standards ought to make expectations clear at different stages of teachers' professional learning. These standards should facilitate the evaluation of advanced cognitive skills. It also recommended the development of moderation programs to enhance teachers' evaluative judgements and promote pre- and in-service training to expand teachers' assessment knowledge (OECD, 2020).

The KSA education landscape shows that it can make further improvements and to accomplish this, there should be a focus on learning outcomes (OECD, 2020). To this end, the government has shown its commitment by introducing significant education reforms to improve learning outcomes in the shape of Vision 2030. Furthermore, a revised school evaluation system that emphasises student learning has been implemented. A performance-based teacher career path has also been designed, while updates have been made to the curriculum and assessment practices. Such changes certainly support the aim of raising of the quality of education in KSA.

A focus on learning outcomes means more progressive use of assessments by teachers. This then evokes the question: how assessment-literate are the teachers of KSA? This is addressed in the next section.

2.4 Assessment Literacy: A Review of the Literature

Teacher assessment literacy has been identified as an essential competency of effective teachers (R. J. Stiggins, 1991a). The increasing interest in teacher assessment literacy can be attributed to the recognition of the importance of assessment as part of school and educational system reforms (Heitink et al., 2016). It is also seen as a vital part of teachers' work in fostering student learning through effective classroom assessments (DeLuca et al., 2013; DeLuca et al., 2015; M. F. Hill et al., 2010; Lam, 2015; Popham, 2011). It is crucial to define assessment literacy given its place in the overall setting of this study, and because the literature in the area is replete with various conceptions and perspectives on the nature and kinds of skills and knowledge that comprise assessment literacy.

Interest in teacher assessment literacy began about 70 years ago, dating back to the 1960s. That period saw a constant flow of publications on assessment, such endeavours epitomise the undergirding aim, that is, to improve teacher assessment literacy. While it could be said that the reasons behind enhancing teachers' assessment literacy have remained unchanged since interest in it began, what seems to have evolved are the conceptions of what sound assessment practices entail. In this regard, Davies (2008) identifies two factors: first, the professionalisation of assessment is escalating, and second, there is a rising understanding of the value of teachers' assessment knowledge and expertise in enhancing student learning outcomes.

Traditionally, the notion of assessment literacy has been employed to refer to the various aspects of assessment practices about which teachers are required to be knowledgeable to enable them to undertake the assessment of student learning (Inbar-Lourie, 2008; R. J. Stiggins, 1997; Stoyanoff & Chapelle, 2005). Some definitions of assessment literacy, on the other hand, emphasise teachers' comprehension of assessment knowledge and how to prepare students for

tests and examinations with high stakes. This includes understanding and being able to employ highly technical assessment principles such as validity, reliability, item difficulty and discrimination as well as test results interpretation. This technical conceptualisation of assessment literacy is, for instance, characterised by a definition by Fulcher (2012), who describes teacher assessment competencies that are linked to large-scale standardised testing. But such a technical conceptualisation tends to skew the focus on teacher professional learning towards the design and analysis of the tests themselves when this aspect should be one facet of teacher assessment literacy rather than its focus. Furthermore, such a technical understanding of teacher assessment literacy fails to capture the daily assessment practices of teachers in their efforts to obtain a sense of students' learning in the context of their pedagogical practices and classroom decisions—that is, while teaching.

Hattie (2008) identified in a study comprising 800 meta-analyses that significant improvements in student performance are shown when teachers and students take an active part in their own educational experiences. This implies that making gains on achievement requires that students learn to take ownership of their learning and can assess and monitor their own progress, rather independently of the teachers. This said, teachers must actively encourage students to develop such expertise because it does not come naturally to them. Therefore, the way to engender student abilities to self-regulate and self-assess (which some differentiate as 'assessment as learning') is dependent on the competency of the teachers to create such opportunities for students to do so. The importance of teacher assessment literacy in enhancing student learning is emphasised by Hattie (2008), who states:

The act of teaching requires deliberate interventions to ensure that there is cognitive change in the student: thus, the key ingredients are awareness of the learning intention,

knowing when a student is successful in attaining those intentions, having sufficient understanding of the student's understanding as he or she comes to the task, and knowing enough about the content to provide meaningful and challenging experience in some sort of progressive development. It involves an experienced teacher who knows a range of learning strategies to provide the student when they seem not to understand, to provide direction and re-direction in terms of the content being understood and thus maximize the power of feedback, and having the skill to get out the way when learning is progressing towards the success criteria. (p. 23)

While the fundamentals of assessment are emphasised in more conventional conceptions of assessment literacy, more contemporary perspectives have shifted toward the knowledge and abilities of assessment that instructors must acquire as part of their instructional practices. For instance, according to Fullan (2002), assessment literacy is the capacity of teachers and administrators to use achievement data from students to improve learning and to counsel policy makers on how best to use such student data. Similarly, Webb (2002) points to the significance of employing student assessment data to enhance teaching and the efficacy of initiatives that are aimed at supporting student learning, while R. J. Stiggins (2005) identifies the traits of teachers who are assessment literate as those teachers who are knowledgeable of the distinctions between effective and ineffective assessment.

Furthermore, such teachers, according to R. J. Stiggins (2005), are unfazed by the technical aspects of assessment and feel proficient to take on the relevant planning, designing, and implementing effective assessment that foster student learning. Prior to this, Stiggins (1999) suggested seven competencies that describe teacher classroom practices, and these competencies incorporate the various aspects listed in the Standards for Teacher Competence in the

Educational Assessment of Students (AFT et al., 1990). Stiggins emphasised, however, that the standards used for teacher competency were insufficient to account for the crucial role that teachers play in carrying out the requirements of their assessment responsibilities. He contends that his taxonomy of competencies is informed by research and is also wide-ranging. It should be emphasised that Stiggins's taxonomy of teacher assessment skills has not received empirical backing; there has been no study to show whether these abilities accurately reflect the many aspects of teacher assessment literacy in actual practice.

Popham (2011) believes that teachers' awareness of the fundamental concepts and procedures surrounding assessment that have an influence on educational decision-making constitutes their assessment. Like Stiggins, he stresses teachers' capacity to employ assessment to direct their decision-making processes on efficacious student learning and pedagogical approaches that enhance student learning.

Most definitions of teacher assessment literacy in the literature centre around the more practical dimensions of assessment, such as the strategies and skills needed to assess (Gotch & French, 2014; Schafer & Lissitz, 1987). Furthermore, much educational research in the area (Brown, 2008; DeLuca et al., 2013; R. Stiggins, 2009; R. J. Stiggins, 1995) as well as educational policies around the world have attempted to inform and support teacher practice according to what is thought to be essential to effective teaching practice. For example, in Australia, the Australian Professional Standards for Teachers (APST) includes the articulation of the competencies that teachers need to acquire in terms of assessment (AITSL, 2018).

Similarly, in KSA, the recent NPST also include the kinds of expertise that teachers at different levels of experience ought to possess in terms of assessment. Thus, on a broad level, teacher assessment literacy can be viewed as 'an individual's understandings of the fundamental

assessment concepts and procedures deemed likely to influence educational decisions' (Popham, 2011, p. 267). Using various tests that are in line with clearly established accomplishment objectives and extrapolating student performance from specific types of exams are two behaviours that demonstrate a teacher's assessment literacy.

Several other definitions that arise from notions of standards and competencies also appear in the literature. For instance, Stiggins (1991a, 1991b) asserts that teacher assessment literacy is an art that requires skills to measure the student learning. Some of the actions he classifies as being key to effective assessment include explaining and sharing with students the reasons behind assessing, what is to be assessed, designing precise assessment processes and communicating results of the assessment effectively. According to Fullan and Watson (2000), teacher assessment literacy refers to teachers' ability, either alone or in conjunction with other teachers, to examine and interpret student work and performance accurately, as well as teachers' capacity to develop remedial plans to alter the situation whenever necessary to achieve better outcomes.

On the other hand, Shepard (2000) focuses on teacher capacity in her definition of teacher assessment literacy. She warns that to bring about school improvements, assessment-literate teachers must have the capacity to design and administer more than just end-of-unit summative assessments. That is, a teacher's assessment tools must include other types of assessment, such as formative assessment. Adding to the significance of teacher ability, McMillan (2001) argues that teachers with a sound background in assessment can combine assessment with teaching such that they employ the relevant teaching approaches that can lead to improved student learning. Paterno (2001), like Shepard, views teacher assessment literacy as having the knowledge of the primary principles of effective assessment practices. This entails

understanding various terminologies, the design and employment of various assessment methodologies and techniques as well as knowledge of the standards of effective assessment and both conventional and alternative means of measuring student learning.

Consistent with the central notion of teacher capacity in teacher assessment literacy, Webb (2002) views it as the knowledge of the forms of assessing what students know and can do, how to read the assessment results and how to use such results to inform improvements in student learning as well as the efficacy of programs, reflecting the teachers' ability to conduct effective formative assessments. In his definition, Mertler (2003) extends the notion of teacher capacity and suggests that assessment-literate teachers are not only knowledgeable about assessment practices and the types of assessment to be employed to attain reliable data on student achievement; they are also able to convey assessment results successfully in various ways such as report cards, portfolios or conferences. They are further able to make use of assessments to optimise student motivation and learning by positioning students as genuine partners in assessment and its processes, making them active agents in the assessment process.

DeLuca and Klinger (2010) view assessment literacy as having the knowledge of the use of various assessment practices. They also underscore the importance of having the theoretical and philosophical knowledge that underpins the types of assessment of students' learning. In integrating both the micro and macro perspectives on educational and school assessment, Brookhart (2011) extends the notion of teacher assessment literacy and reconsiders the Seven Standards of Teacher Competencies in Educational Assessment of Students (AFT et al., 1990). Proceeding from these seven standards, the revised knowledge and skills teachers should possess, according to Brookhart (2011), are:

1. Teachers should understand learning in the content area they teach.

2. Teachers should be able to articulate clear learning intentions that are congruent with both the content and depth of thinking implied by standards and curriculum goals, in such a way that they are attainable and assessable.
3. Teachers should have a repertoire of strategies for communicating to students, what achievement of a learning intention looks like.
4. Teachers should understand the purposes and uses of the range of available assessment options and be skilled in using them.
5. Teachers should have the skills to analyse classroom questions, test items, and performance assessment tasks to ascertain the specific knowledge and thinking skills required for students to do them.
6. Teachers should have the skills to provide effective, useful feedback on student work.
7. Teachers should be able to construct scoring schemes that quantify student performance on classroom assessments into useful information for decisions about students, classrooms, schools, and districts. These decisions should lead to improved student learning, growth, or development.
8. Teachers should be able to administer external assessments and interpret their results for decisions about students, classrooms, schools, and districts.
9. Teachers should be able to articulate their interpretations of assessment results and their reasoning about the educational decisions based on assessment results to the educational populations they serve (student and his/her family, class, school, community).
10. Teachers should be able to help students use assessment information to make sound educational decisions.

11. Teachers should understand and carry out their legal and ethical responsibilities in assessment as they conduct their work. (pp. 6–10)

According to a survey of the literature on assessment literacy for teachers, the concept of assessment literacy evolved to include both the critical perspectives of assessment and its ‘social implications’ (McNamara & Roever, 2006). Such conceptions also entail that teachers play the role of information sharing to other education stakeholders in matters of assessment. Such changing perspectives of assessment literacy have meant a growing focus on ‘teacher capabilities to plan and implement quality assessment tasks, to interpret evidence and outcomes appropriate to the assessment purpose and type, and to engage students themselves as active participants in assessment of their own learning’ (Looney et al., 2018, p. 2).

These changes are apparent in current teacher standards across the world, which seem to now be more focused on formative assessments, support for teachers and the way assessment results are being reported and conveyed (DeLuca et al., 2016). For instance, in Australia, the AITSL teacher standards related to assessment is captured by the injunction to ‘assess, provide feedback and report on student learning’ in which the five Focus Areas are:

- ‘Focus Area 5.1 Assess student learning: Demonstrate understanding of assessment strategies, including informal and formal, diagnostic, formative and summative approaches to assess student learning.
- Focus Area 5.2 Provide feedback to students on their learning: Demonstrate an understanding of the purpose of providing timely and appropriate feedback to students about their learning.

- Focus Area 5.3 Make consistent and comparable judgements: Demonstrate understanding of assessment moderation and its application to support consistent and comparable judgements of student learning.
- Focus Area 5.4 Interpret student data: Demonstrate the capacity to interpret student assessment data to evaluate student learning and modify teaching practice.
- Focus Area 5.5 Report on student achievement: Demonstrate understanding of a range of strategies for reporting to students and parents/carers and the purpose of keeping accurate and reliable records of student achievement.’ (AITSL, 2018, pp. 18–19)

Newfields’s (2006) investigation of the many facets of teacher assessment literacy is one such study. However, the methodology employed in his study to postulate these aspects is theoretical, as is the work of Stiggins, and Mertler and Campbell. While these aspects were validated and underwent pilot testing, the drawback is that they were not based on empirical evidence that backs the existence of the proposed aspects of teacher assessment literacy in the first place.

DeLuca and Klinger (2010) and Fulcher (2012) conducted research on teacher assessment literacy that was grounded in empirical methods and provided preliminary research support for the presence of their suggested aspects of assessment literacy. De Luca and Klinger’s dimensions of teacher assessment literacy could be seen to link both practice and theory, but a few of their dimensions could be subsumed under a larger dimension. Examples of elements that could be combined into one dimension are summative assessment-related elements including assessment design, marking and knowledge of technical assessment practice. On the other hand, the dimensions proposed by Fulcher entail those aspects suggested by Newfields (2006), with an added dimension that is linked to classroom assessment practices. Although the research

described contributes to many aspects of teacher assessment literacy, it has certain limitations. The use of those characteristics to classify teacher assessment practices is instructive in developing an understanding of the skills and knowledge that teacher assessment literacy entails, such as those that are found in KSA's NPST.

It could be said that the notion of teacher assessment literacy has been gradually shifting over time. More conventional notions of teacher assessment literacy emphasise the dimensions of teacher knowledge and skills around assessment, that is, what teachers need to know and can-do regarding assessment. But more recent ideas on teacher assessment literacy also acknowledge the significance and function of context in the growth and use of assessment knowledge and skills. It is also viewed as a component of teachers' professional identities, in which they combine their assessment knowledge with that of content, pedagogy and a wider educational setting (Adie, 2013; Cowie et al., 2014; Looney et al., 2018; Scarino, 2013; Xu & Brown, 2016). This reconceptualisation of teacher assessment literacy is best captured by Willis et al. (2013), who define assessment literacy 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)

This latter definition of teacher assessment literacy considers that assessment practices are influenced by various contextual factors that can affect the nature and extent of assessment practices. Such factors include teacher experience, professional learning, classroom context and curriculum (Looney et al., 2018). It is such socio-cultural factors that can account for the diverse teacher experiences of assessment despite the uniformity in educational policies and professional

learning (Tierney, 2006). Past studies (for example, Coombs et al., 2018; DeLuca et al., 2016a, 2016b), indicate that these include teachers' understandings of the purposes of assessment and the assessment processes. Thus, while teacher assessment literacy continues to be important, more contemporary definitions of teacher assessment literacy consider assessment as a more complicated process that considers a variety of socio-cultural elements that affect what instructors think and do with relation to assessment practices.

Pastore and Andrade (2019) have more recently advanced this socio-cultural concept of evaluation literacy. They propose a three-dimensional model that considers the conceptual, praxeological and socio-emotional dimensions in which these dimensions are shaped by 'the nested cultural contexts of a classroom, school, system, and nation' (p. 135). Moreover, they hold the view that teacher 'assessment literacy is an adaptive competence that requires the ability to apply knowledge and skills in flexible, creative, and responsive ways in different educational contexts' (p. 134). Informed by the consensus of an international expert panel on assessment, Pastore and Andrade propose a revision of teacher assessment literacy as follows:

an interrelated set of knowledge, skills and dispositions that a teacher can use to design and implement a coherent and appropriate approach to assessment within the classroom context and the school system. An assessment literate teacher understands and differentiates the aims of assessment, and articulates a sound, cyclical process of collection, interpretation, use of evidence, and communication of feedback. In this way, assessment information can be used to make inferences about student learning, inform decisions about curriculum, and adjust instruction (pp. 134–135).

Assessment literacy is acknowledged to be culturally situated and contextually responsive in concepts of teacher assessment literacy, such as that above (Xu & Brown, 2016). This would

mean that it is subject to attendant contextual forces, just as teachers and teaching are. In terms of this proposed study, such definitions of assessment literacy are appropriate as they acknowledge the impact and the complexity of contextual social and cultural factors on teachers' assessment practices. While the dimensions of teacher knowledge and skills undoubtedly remain essential to the notion of teacher assessment literacy, the socio-cultural dimension that shapes the knowledge and practice must also be recognised, especially in a context as distinct as KSA.

2.5 Measuring Teacher Assessment Literacy

To date, there are no known studies measuring Saudi teachers' assessment literacy, and this is one gap that this study intends to fill. Most studies that have been carried out have been conducted in the Western context. These studies have sought to measure teacher assessment literacy quantitatively. Such measures were informed by the original 1990 *Standards for Teacher Competence in Educational Assessment of Students* (AFT et al., 1990).

The studies referred to in Chapter 1 Plake et al. (1993) and O'Sullivan and Johnson (1993) used the Teacher Competencies Assessment Questionnaire (TCAQ) to assess the seven competency standards and the study by O'Sullivan and Johnson (1993) to measure the extent of students' understanding of the teacher competency standards. Chapter 1 also referred to the Campbell et al. (2002) study which demonstrated that there were differences in competence in the seven assessment standards among student teachers and drew the conclusion that pre-service teachers suffered from a lack of competency in key aspects of assessment when they commenced their teaching career. Similar outcomes were found in research conducted by Mertler (2003) and Mertler and Campbell (2004, 2005). All these studies used the TCAQ as the measure of teacher assessment literacy.

Mertler and Campbell revised the TCAQ in 2004 to become the ALI (Mertler & Campbell, 2004, 2005) and then used the ALI with teachers as the measure to gauge teacher understanding of the standards. Moreover, it is important here to illustrate that researchers have frequently used two different names, TALQ and TCAQ, synonymously to describe this instrument. Initially, Plake and Impara (1992) and later Plake et al. (1993) developed and analysed this instrument but did not provide a specific short name for it. Additionally, O'Sullivan and Johnson (1993) referred to it as the 'Teacher Competences Assessment Questionnaire' or TCAQ.

In 2014, DeLuca introduced the acronym TALQ, which stands for 'Teacher Assessment Literacy Questionnaire'. However, in 2015, 2016, and 2018, DeLuca was once again referred to the TALQ as the TCAQ. Gotch and French (2014) mentioned the questionnaire without assigning a short name and simply called it the 'Teachers Assessment Literacy Questionnaire'.

The above studies demonstrate the TCAQ (or TALQ) as a valid tool to assess teachers' assessment literacy in the context of different education systems that have professional standards for teachers. Alkharusi (2011) further validated the tool and confirmed its reliability as a measurement tool. The results of his study of Omani teachers using the original TALQ 'support the utility of the Original for instructional and assessment purposes' (Alkharusi, 2011, p. 1615). Alkharusi had also translated the original TALQ into Arabic and subjected his translation to a critical review to ascertain its degree of validity. For the present study, the Arabic version of the original TALQ was instrumental as the basis for developing a research instrument that assesses teachers' assessment literacy in Saudi's educational context once it is aligned to Saudi's NPST. The NPST was used as the basis for the development of an assessment literacy framework and instrument to measure KSA teachers' assessment literacy (see further, Chapter3)

Given the objectives of this study, the name TALQ has been chosen, aligning this decision with the work of Alkharusi (2011). Henceforth, the term TALQ will consistently be utilised in this study.

2.6 Assessment Literacy Studies in the Context of Saudi Arabia

Little if any research has been conducted on assessment literacy in the context of KSA education. Despite this, there has been research on assessment more broadly and certain inferences can be drawn from these studies pertaining to the assessment literacy of teachers in the KSA. As mentioned earlier, there was a study on assessment that compared the Science Education Diploma Program at Jeddah in Saudi Arabia and a university in Iowa in their delivery of performance skills lessons (Al-Shikh, 2017). Findings from this research showed that students training to be science teachers in Saudi Arabia greatly differed from their US counterparts. It was found that US teachers were more knowledgeable in scientific content and had better assessment literacy, were more able to deliver skills-based lessons and employed more effective pedagogical approaches than Saudi teachers. The researchers recommended the introduction of, and the range of courses and course duration should be extended. Another inference that can be drawn from the same research is that KSA teachers' assessment literacy is inferior in comparison to that of US educators. They called for professional learning that could raise their assessment literacy levels.

More recently, Althobeti (2018) investigated assessment in tertiary education. His study involved reviewing the assessment methods and assessment culture in higher education among 171 faculty members at Shakra. Specifically, he examined the faculty's culture, assessment practices and their links to university education outcomes. Part of the research aimed to gauge student competencies and as well as the extent to which learning objectives and those associated with behaviour were attained. The findings indicated that teaching staff largely employed

outmoded assessment practices and the employment of more contemporary assessment practices was limited, while the variety of assessment practices was at a moderate level. A consequence of these findings saw the diploma program in KSA universities cease and the introduction of a new requirement in the training of candidates in the professional education master's program. As with the comparative study previously discussed, these findings suggest a lack of teacher assessment literacy, albeit at university level, and these educators may require further professional learning in assessment literacy.

Another study on assessment was undertaken by Yousuf (2018) and aimed at determining the obstacles faced in the implementation of authentic assessment in the learning of Islamic Education in the schools of the Tabuk Educational Region. It indicated that there were many obstacles to implementing authentic assessment. The size of the obstacles faced by teachers significantly differed from those faced by supervisors. Albalawi (2018) found that mathematics teachers in Riyadh had a positive attitude towards formative assessment and that they use formative assessment prominently in their assessment practices. However, the study also found that the teachers faced the following issues when implementing formative assessment in the classroom: insufficient class time, accommodating all students in a single large classroom and being able to use modern assessment procedures.

While the above studies relate to assessment more broadly, they nonetheless give an indication as to the state of teacher assessment literacy in the KSA. Overall, there seems to be a lack of focused research regarding Saudi teachers' assessment literacy. As mentioned, there is also an absence of a clear articulation of an assessment literacy framework and a tool to effectively measure teachers' assessment literacy against KSA's NPST.

There is presently no assessment system that specifies assessment activities at the national level or how various assessments function, according to an OECD (2020) research study regarding KSA education. It has been demonstrated that the examinations created by ETEC help students achieve greater learning outcomes. Furthermore, research on classroom evaluations has shown that these evaluations primarily measure factual knowledge and do not evaluate higher order abilities, particularly those that measure problem solving and critical thinking (Almuntasheri, 2016).

The OECD research (2020) also discovered that teacher evaluation is mostly summative at the classroom level. Additionally, evaluation materials heavily emphasise closed-ended item types and memorisation, which restrict critical thinking. According to the OECD analysis, more varied assessment methods, particularly formative ones, should be encouraged to boost classroom evaluations. These results provide further evidence of the need to increase the number and quality of professional learning programs within the KSA targeting assessment literacy for teachers.

2.7 Teacher Assessment Literacy Professional Learning

Recent initiatives in the KSA have clearly called for enhanced teacher assessment literacy to undertake the various kinds of assessment such as formative assessment. The new curricular framework emphasises formative evaluation as a crucial part of student development in this regard. A variety of evaluation tools have also been developed by ETEC to help teachers evaluate student learning properly while taking national requirements into consideration. The use of formative assessment in the classroom is also a subject of several initiatives by Education Offices within the KSA.

Several KSA education stakeholders are of the view that, in general, teachers have weak assessment literacy skills (OECD, 2020). Given that assessment effectively supports learning better than any other intervention (Black & Wiliam, 1998; Hattie, 2008), it would seem essential that KSA teacher assessment literacy programs should also be consistent with the key principles of effective assessment requirements and aims. To achieve this, some suggest that teacher assessment literacy reforms and professional learning programs should be organised around several of the key principles of effective assessment themselves. For instance, Davison (2013) suggests drawing on the very precepts regarding formative assessment or assessment for learning (AfL) by the Assessment Reform Group (1999, p. 7) to inform effective teacher assessment literacy reforms and professional learning. These are:

1. AfL is included in national curriculum for evaluation or assessment academically and institutionally.
2. Assessment reform goals are explicitly shared with all stakeholders and stakeholders together work to identify how to know and to recognise the standards they are aiming for.
3. Constructive qualitative feedback helps stakeholders to identify the subsequent steps requires for restructuring process and their utilisation.
4. All stakeholders regularly review and reflect on assessment data.
5. It is assumed every school, teacher and student can improve.

Applying the above principles of assessment to the context of KSA would mean the following. First, effective reforms in teacher assessment literacy and professional learning should consider where teachers are in terms of their assessment literacy competence. Therefore, effective assessment literacy reforms and professional learning that aim to develop KSA teachers' assessment literacy competence must be informed by understandings of their existing

competence before designing programs that meet their various needs. The findings of the present study can contribute to such understandings, although it is very difficult to generalise beyond the small purposive sample that was used in the study.

Second, the goals of assessment reforms must be clearly communicated to all stakeholders in the educational system so that there is a common understanding about assessment standards and goals among all stakeholders. It is such common understandings and shared aims around assessment that can effectively inform the collaboration among stakeholders. MOE or ETEC can take the lead in putting in place system-wide initiatives to establish common understanding regarding how to go about knowing and recognising the benchmarks to which they are aspiring. Given its current absence in the KSA educational context, it is suggested that this calls for the articulation of a clear assessment literacy framework - one that is based and connected to the NPST. Such an assessment literacy framework would be able to provide teachers, school leaders and key stakeholders with a shared understanding of and standards around assessment and assessment literacy, unifying all parties in both thought and action. Moreover, such an assessment literacy framework can inform a systematic and comprehensive teacher professional learning agenda that can be implemented country-wide in a concerted manner, rather than one which is more scattered in nature.

Third, the importance of recognising that qualitative feedback is instrumental to stakeholders' identification of what is required for effective changes and how to go about implementing what is needed to improve assessment literacy. Fourth, consistent review and analysis of assessment data. Fifth, as a tenet of formative assessment, there must be an explicit assumption that the key stakeholders such as schools, teachers and students can make progress.

In summary, the effectiveness of a teacher assessment literacy professional learning program starts with a clear understanding of the basic principles of effective assessment by stakeholders. In the case of KSA, this implies an important role for ETEC in coordinating such understandings, principles and implementation, so that assessment literacy framework can be executed in concert across all levels of education.

Difficulties in effectively implementing this should be anticipated. For instance, teachers' workload in the KSA is determined solely by classroom contact hours. Unfortunately, professional learning and other non-instructional tasks do not receive the necessary attention and are thus neither given due priority, nor rewarded. It follows that KSA teachers engage in a small amount of professional learning compared with international benchmarks. OECD's (2020) report on the education system of KSA indicates that the professional learning programs for teachers in the KSA suffer from a significant lack of assessment programs for teachers, and teachers use inaccurate and unreliable practices.

Thus, the teacher standards in KSA could set clear expectations for the assessment literacy of teachers at different levels of the professional pathways (OECD, 2020). The initial phase in the implementation of the new standards will necessitate that all Practitioner teachers master the essential competencies set by the new standards and become certified Practitioner teachers. Considering the prevailing KSA educational landscape of assessment, teacher literacy and the absence of an assessment literacy framework and tool, this proposed study will fill the gaps that have been identified in the literature review.

2.8 Summary

This review of the literature has identified several key themes. First, the educational context of KSA is one that is largely centralised where authority largely resides within the MOE,

even though current reforms such as Tatweer aim to reduce the centralisation of education in the country and grant schools more autonomy. More broadly, through Vision 2030, education in KSA is seen as a key tool to modernise the country and prepare it for a post-oil era, and the government continues to invest heavily in education.

Second, definitions of assessment literacy have evolved over time. While the notion of teacher competence in terms of skills and knowledge remains central to the construct, more recent reconceptualisation of teacher assessment literacy considers the socio-cultural dimension. Contemporary notions of teacher assessment literacy recognise that teachers' assessment practices and data use are shaped by the context in they take place, and that the learner is also a key agent in the assessment process. These notions were integral in the development of an assessment literacy framework and tool to measure KSA teachers' assessment literacy. This is the primary aim of this study.

Third, studies to measure teachers' assessment literacy have been quantitative in nature. Measures of assessment literacy are also often based on the professional teacher standards which are mandated in the respective educational context. While measures are advancing, an appropriate means to measure the assessment literacy of Saudi teachers as proposed in this study will be developed from the original TALQ (Alkharusi, 2011; Mertler & Campbell, 2004; Plake et al., 1993). The original TALQ is one of the more frequently used tools, albeit revised in some studies, to measure assessment literacy. Moreover, an Arabic version of it has been validated (Alkharusi, 2011). Such a questionnaire thus provides a solid basis for the development of a tool to measure KSA teachers' assessment literacy. This development of the assessment literacy framework and tool is also informed by KSA's NPST.

Fourth, teachers play a vital role in the practice of assessment, although they often face challenges in implementing assessment. Given that teacher assessment practices - and by extension, their assessment literacy level - affect student learning outcomes, gaining insights into their state of assessment literacy is vital. Such insights will inform the study's recommendations relating to teacher professional learning and support in assessment literacy.

Fifth, only a few studies related to assessment in the Saudi Arabian educational context have been undertaken and there has been no research into teachers' assessment literacy in the context of primary, intermediate, and secondary school levels in KSA.

Finally, the state of KSA teacher assessment literacy and the recent initiatives towards assessment reveal a troubling gap. This study aims to contribute to the key area of assessment literacy research in KSA schools and to develop an assessment literacy framework and tool to assess teachers' assessment literacy. The next chapter discusses the methodology and methods that will be employed to address the research questions.

Chapter 3: Methodology and Methods

3.1 Introduction

This chapter presents a detailed discussion of the methodology and methods used to investigate the research questions outlined in Chapter 1. It includes elaborations and justifications of the selection of participants, and participating research sites, ethical considerations, a presentation of KSA's professional standards for teachers (i.e., the NPST), data collection methods and data analysis - with data collection being done in three different stages. First, develop and validate a framework for teacher assessment literacy aligned to NPST. Second, evaluate existing measures of teacher assessment literacy (i.e., TALQ) to see how valid they are for measuring performance against the NPST. Finally, modify the TALQ and build and validate a new measurement instrument, called the MTALQ (the Modified Teacher Assessment Literacy Questionnaire), which is better aligned to the NPST and the assessment literacy framework.

One main purpose of this study was to solicit from a sample of primary, intermediate and secondary teachers what they considered to be the assessment knowledge and understanding that teachers need to have to demonstrate that they have met the Practitioner and Advanced standards described in the NPST, which are required for successful teaching in the KSA. This chapter describes the steps taken to produce a pool of elaborations collected to develop an assessment literacy framework regarding assessment knowledge and understanding perceived by KSA teachers as being essential in order meet the NPST. These understandings were then further explored and served as the basis for development of a new instrument, which is referred to as the Modified Teacher Assessment Literacy Questionnaire (MTALQ).

Practitioner teachers were of primary interest because it was important to understand whether this group of teachers had informed and balanced attitudes and beliefs about assessments in anticipation of possible professional development in support of reform efforts for assessment literacy. This study also investigated whether the teachers who had already met the Advanced level on the NPST had the knowledge and skills needed to be able to conduct assessment effectively so that they could use high-quality evidence to adapt and modify their teaching activities to improve student learning. Additionally, these teachers were asked if they would benefit from feedback received from assessing and supporting their colleagues.

3.2 Methodology

This design involved a single study containing qualitative and quantitative data collection, which were conducted simultaneously, following the concurrent triangulation research design proposed by Kroll and Neri (2009). The purpose of this type of design is to validate the findings generated by each method through evidence produced by the other. This study was primarily devoted to establishing a KSA teacher assessment literacy framework and a questionnaire, the MTALQ, that could be used to measure KSA teachers' assessment literacy. The study helps to provide insights into the extent to which Saudi teachers are assessment-literate. It also contributes to the development of a corpus of knowledge that can be used to build professional learning and development programs to support teachers in improving their knowledge and skills in assessment literacy, not only in the KSA, but perhaps elsewhere as well.

3.3 Delineating the Teacher Assessment Standards with the Professional Knowledge and Understanding to Develop the Assessment Literacy Framework

This study took each of the Practitioner and Advanced standards from KSA's NPST and identified the potential knowledge and understanding that teachers would need to demonstrate at the Practitioner and Advanced levels to show that they had met the standards. The NPST comprises various professional standards described in terms of professional values and responsibilities, professional knowledge and professional practice. These include commitment to Islamic values and professional ethics, continuous professional development, professional integration with communities and societies, understanding of linguistic, quantitative and digital skills, knowledge of critical and general teaching methods, planning of materials and activities, creating interactive and supportive materials, and evaluating students' performance.

However, this study focused solely on NPST standards related to assessment. The potential knowledge and understanding that teachers need to be able to demonstrate what is required by the Practitioner and Advanced levels of those standards were delineated and expressed in terms of a set of elaborations. This step was required because the NPST document does not articulate or detail the range of knowledge and skills that is connected to the relevant standards regarding assessment. These elaborations were produced based on document analyses of the NPST and a literature review of the areas around assessment literacy. Via this process, the 32 elaborations were developed at the Practitioner level and 33 at the Advanced level.

Given that schools may have their own assessment regimes, rules, or norms (and outcomes) expected of teachers practising at the Practitioner and Advanced levels, using the standards is an objective and unifying point of reference. To further develop this assessment

literacy framework (representing the collection of these derived elaborations) a Likert Survey (see Appendix B for The Likert Survey) was developed that presents each of the initial 65 elaborations derived and presented in the assessment literacy framework (see Appendix C). In this survey, teachers were asked to rate each elaboration as either ‘Essential’, ‘Desirable’ or ‘Not Necessary’ in terms of their alignment to the standards related to the assessment literacy required in the NPST. Each question was framed in a way that asks the respondent, ‘To what extent do Practitioner teachers (or teachers at the Advanced level) need to know “*the elaboration*” to demonstrate the specified standard?’ The advantage of this process is that viewpoints were collected from all the participants and their variations were captured. Subsequent focus group discussions over time allowed the participants to clarify and explain their ratings, allowing the researcher to better understand the motivations behind them. This provided a primary source of validity evidence regarding the alignment and relevance of the derived elaborations and the refined assessment literacy framework to the NPST.

The aim of this study was not only to explore the teachers’ understanding and their knowledge of the assessment standards as stipulated in KSA’s NPST; it was also to build a pool of more complete elaborations (ultimately collected into the assessment literacy framework) that could further be applicable for novice or beginning teachers. The data and findings from discussions about this assessment literacy framework (or collection of elaborations) could be useful for those involved in assessment policy, practice and research, specifically as it relates to assessment systems, teacher standards and educational reform.

3.4 Research Ethics

All research has certain norms or standards to be followed before, during and after the investigation begins. This study has not been an exception. All ethical norms and standards in

planning, conducting, and reporting this study have been followed. All the requirements of the University and ethical considerations included in Connelly's (2014) study have been followed. First, the research was proposed in collaboration with the researcher's supervisor and all other concerned faculty, staff and relevant responsible bodies at the university. Second, selected participants were volunteers. They were not under any kind of pressure to participate. Third, the tools were developed in collaboration with all the participants. They were asked to be involved in a series of focus group discussions so that they could freely express themselves. Fourth, during the data collection, participants were not influenced by the researcher and their responses were not manipulated in any way. One of the most important aspects of this collaboration was that the researcher worked together with his supervisors, who are assessment experts, but not under them. Fifth, the tools used for data collection were also prepared with participating teachers' feedback. The process for the development of the study tools included several iterative revisions and reformulations. After the development of the assessment literacy framework, the teachers and the experts involved in the study responded to the questions presented in a Likert survey. This helped in refining the assessment literacy framework and the MTALQ as needed. Sixth, the tools were implemented and used by the participants at different stages. They engaged in focus group discussions to identify the strengths and weaknesses in the assessment literacy framework and the original TALQ. As mentioned, before collecting the data, all the participants granted their consent to participate in the study voluntarily and their candid views in the research were indicative of this.

After gaining the University's consent to conduct the study, the KSA MOE, schools and teachers within schools were approached to gain their consent to participate in the study. Communication with the principals of the schools was paramount because they were the

‘gatekeepers’ to the study, and their approval was needed to invite teachers to participate in the study. Following appropriate research ethics, schools and staff were assured that their participation was voluntary and in participating, their identity would be kept confidential. It was made very clear to participants that their participation was voluntary and that their identities would be protected.

The teachers participated in focus group discussions as part of the validation of the assessment literacy framework and original TALQ, using the NPST as the reference. This was also used to establish a corpus of their understanding on the items provided. It was made known to the participants that the information collected was only for the purposes of the study and would not be used for any other purposes. The collected data were not manipulated, falsified or fabricated. Similarly, the findings and the conclusions have been drawn from fully documented and disclosed analysis and interpretation.

The generalisability (Best & Kahn, 2004) of the results of the study has some limits, mainly because of the KSA context and the way in which the sample was selected (though the process itself is quite generalisable and applicable in context beyond KSA). Sources referenced in the study are acknowledged and presented in the References.

3.5 The Sample

Using purposive sampling, the study included six KSA government schools. The participants in the focus groups in the study included three groups of teachers - primary, intermediate and secondary teachers from these government schools. In all, the researcher conducted four focus group interview sessions with these three groups of teachers as explained below.

- The first focus group was a group of primary teachers to discuss the Likert survey results regarding the elaborations collected relative to the assessment literacy framework.
- The second focus group was with intermediate and secondary school teachers to discuss the Likert survey results regarding the teacher ratings of the assessment literacy framework elaborations.
- The third focus group was with all levels of teachers (primary, intermediate and secondary) to access their views on the alignment of items in the original TALQ with the elaborations.
- The fourth focus group was with all levels of teachers to review the MTALQ created by the researcher (based on input from the previous focus groups) from the original TALQ and the elaborations or assessment literacy framework.

School principals were contacted and requested to support the study by creating opportunities for teachers to participate, and by managing their participation through to completion. Teachers who participated did so voluntarily and signed an informed consent form.

Once the schools were selected, all teachers in these schools were asked to participate. The primary unit of analysis was the responses of individual teachers. Two primary schools, two intermediate schools and two secondary schools were selected for the study. The schools were purposively selected to increase the likelihood that participating teachers would be familiar with the knowledge and skills related to assessments and have an ability to engage in an activity designed to share this knowledge with colleagues. Teachers who agreed to participate in the study were invited to complete the survey provided via their principals. The participating schools were selected to represent, as best as possible, the characteristics of all schools in the KSA and not be exceptional or special in any way other than to be familiar with knowledge and skills

relating to assessment. Such government schools represent a balanced range of schools in the context of KSA's education system. Because of this, it was expected that the development and validation of a teacher assessment literacy framework and the MTALQ to measure assessment literacy for teachers in the KSA could provide a 'rough' indicator of the majority of KSA schools when considering a more systematic sample in the future.

There were between 50 and 60 teachers from each school who were selected and volunteered to participate in the study (i.e., the final number of teachers who participated in each stage is given below). Participants were teachers who had been teaching for more than one year either in the primary, intermediate, or secondary level. They were teachers who had already graduated and learnt about KSA's NPST and the assessment system during their recent pre-service and in-service training. No students participated in the study.

As stated previously, only the 'Practitioner' and 'Advanced' standards in the NPST that related to assessment were selected to be included in this study:

Standard 10.1 Preparing evaluation tools;

Standard 10.2 Applying evaluation;

Standard 10.3 Engaging students in evaluation processes;

Standard 10.4 Employing evaluation results; and

Standard 5.5 Preparing Evaluation Reports.

The focus group discussions and responses to the questionnaires (i.e., Likert survey and MTALQ) were used for gathering original and first-hand data from the participants and these represented the primary sources of data collection. What follows is the description of the stages that were undertaken for this study.

3.6 Description of the Stages

3.6.1 Stage 1: Development and Validation of the Assessment Literacy Framework

This stage included the development and validation of the assessment literacy framework that was based on the NPST and elaborations that were produced from the NPST. The main purpose of this stage was to address the study question ‘What assessment literacy knowledge and skills do KSA teachers need to be able to demonstrate that they have met the assessment standards described in the KSA NPST?’ Elaborations were collected to build the assessment literacy framework (see Appendix C) and provided more detail and clarity regarding the knowledge and skills needed to demonstrate the standard. This stage entailed the following steps.

3.6.1.1 Step 1: Unpacking and Elaborating the Various NPST Assessment Standards

The various assessment standards in the NPST were delineated into elaborations that described more extensively the kinds of skills and knowledge that teachers need to be able to demonstrate that they have the knowledge and skills necessary for assessment as outlined in The Refined Assessment Framework (Appendix D). As mentioned, the NPST document does not articulate specifically or detail the range of knowledge and skills needed. Other documents such as the curriculum framework, the national assessment guidance and the national educational policy documents provide the larger context of assessment and assessment literacy in KSA. They were used as secondary sources of information to inform the unpacking of the goals of the NPST. The elaborations represent a ‘consensus’ of the application of all this information to make the NPST more operational for participating teachers. The elaborations were produced before conducting the focus group sessions.

3.6.1.2 Step 2: Refining the Assessment Literacy Framework

After the first step, a survey using a Likert design was constructed to present each of the elaborations to a small sub-sample of teachers - 44 teachers representing primary, intermediate and secondary schools. They were chosen from the original sample of schools and teachers who had agreed to be in the study and voluntarily agreed to fill out the Likert survey. The teachers were invited to indicate whether each elaboration was ‘Essential’, ‘Desirable’ or ‘Not Necessary’ relative to the nominated standard from the NPST. Each question in the survey was framed in a way that asked the respondents ‘To what extent do you (or teachers at the nominated level) have to know or understand the elaboration to demonstrate the specified standard?’ The information collected from the survey and the focus group sessions led to changes (refinements) in the assessment literacy framework. This is referred to as the refined assessment literacy framework to distinguish it from the original one (See Appendix D).

3.6.1.3 Step 3: Focus Group Sessions

Focus group interviews were then conducted to review the results of the survey. These focus group sessions were used as part of the validation process of the teacher assessment literacy framework. In a similar vein, focus group interviews were employed to enable teachers to provide their views on the issues identified, and provide the researcher with the means to improvise and probe further questions during the sessions, as relevant (Arksey & Knight, 1999).

One focus group discussion involved 26 primary teachers who had already completed the Likert survey. The second focus group comprised 18 secondary and intermediate schoolteachers who had also completed the survey. The focus group discussions were recorded with the consent of the participants. Information from these groups was used to further refine the elaborations which constitute the major component of the assessment literacy framework. Because of

COVID-19 restrictions, the focus group discussions were conducted online using the Zoom platform. This initially presented some unexpected challenges with connectivity and scheduling, but these were resolved, and the sessions proceeded as planned.

The elaborations were discussed with participants who had already completed the Likert survey, with a view to learning what teacher participants identified as being ‘problematic’ and to determine why certain statements were ranked so low (i.e., some elaborations were ranked high by the participants, while a few were ranked low). The participants were also invited to react to any commentary that the survey participants contributed through their responses to the open-ended questions.

The overall results from the Likert survey and the associated focus group discussions provided a comprehensive pool of content that formed the basis of the updated KSA teacher assessment literacy framework. The horizontal component of the assessment literacy framework is the career level progressions that define the KSA NPST (i.e., the Practitioner and Advanced teachers). The details within these components have been mentioned in the assessment literacy framework. There are two career levels for teachers (Practitioner and Advanced) in KSA. The NPST is included in Appendix E.

3.6.2 Stage 2: Development and Validation of the Assessment Literacy Questionnaire

This stage included the development and validation of the assessment literacy questionnaire (MTALQ) to assess the level of KSA teachers’ assessment literacy. This was achieved by aligning and evaluating the original TALQ against the NPST and the refined assessment literacy framework (in alignment with the second research question posed by this study), which was then modified to build an instrument that suited the KSA educational context. This stage aimed to address the research’s question ‘To what extent is the original TALQ a valid

measure of the assessment knowledge and skills required by KSA teachers to demonstrate that they have met the standards described in the NPST?’

Focus group interview was used in the process of building an evidential case for the use of the original TALQ to measure assessment literacy of a sample of teachers in the KSA. The meeting included 35 new teachers representing primary, intermediate and secondary schools from the original sample of schools and teachers who had agreed to be in the study and voluntarily agreed to participate in the discussion of aligning and evaluating the original TALQ against the NPST and the refined assessment literacy framework. This focus group discussion lasted approximately two hours. The session was audio recorded with the consent of the participants. The participants were given the items that comprised the original TALQ and the refined assessment literacy framework. They were then asked to identify which skill or content component of the refined assessment literacy framework was best assessed by the item in the original TALQ. If the item did not assess any of the skills or content in the refined assessment literacy framework, then they were asked to indicate the item as ‘not applicable’.

The original TALQ consisted of 35 multiple choice items, each with four options but only one correct answer. The items in TALQ were originally developed and designed by Plake and Impara (1992) to measure teachers’ assessment literacy. In this study, the Arabic version of the instrument was used. This version was translated by Alkharusi (2011) from the original English version for his research. The motivation for using this version was to make it easier for the participants to use, as it was presented in their native language of Arabic.

Through the processes outlined above, the participants shared what they deemed as relevant items in the original TALQ. At the end of this step, the shell of a KSA teacher

assessment literacy questionnaire emerged, and it comprised those items from within the original TALQ that were pertinent to the KSA NPST and were seen as relevant to the KSA context.

Qualtrics (a platform for creating and distributing online/web-surveys) was used as the survey platform for all stages of this study. This is because it affords users online survey design and the ability to allow multiple people edit or read-only access rights, which helped facilitate efficient and effective data gathering. This was also done to accommodate the challenges posed by COVID-19.

3.6.3 Stage 3: The Final Review of the Instrument (MTALQ) and Implementation

At this stage, a new questionnaire (MTALQ) was constructed, which retained those items from the original TALQ in Stage 2 that were judged by the focus group as having high alignment with the refined assessment literacy framework and the NPST (this is aligned to the third research question posed by this study). The revised version also contained newly constructed ‘replacement items’ as well as modifications of the original TALQ items. The total set of items was then reviewed once again by the same focus group that evaluated the items from the original TALQ (i.e., one focus group that comprised primary, intermediate and secondary teachers). The resulting modified version of the original Arabic TALQ was called ‘the MTALQ’ and was then administered to the entire sample for the study. The following steps elaborate the details of the processes that comprised Stage 3.

3.6.3.1 Step 1: The Use of Teacher Assessment Literacy Framework

In this step, the refined assessment literacy framework was used as the ‘Table of Specifications’ for developing the MTALQ. The items from the original TALQ that had been deemed to be content appropriate (Stage 2) were identified in the refined assessment literacy framework. Furthermore, new items were written and reviewed by assessment experts to provide

evidence against those cells in the refined assessment literacy framework that did not have the items already allocated. This, then, was the process that resulted in a new questionnaire that is tailored to fit the KSA NPST - that is, the MTALQ.

3.6.3.2 Step 2: The Feedback Sessions from the Participants

The new items were classified according to their location in the refined assessment literacy framework collaboratively by the researcher and the participants. The members of the focus group discussion who were involved in Stage 2 were invited to attend a second focus group in which they reviewed the new items in the MTALQ from a subject matter point of view, a technical point of view and an editorial point of view. All the members of the focus group took part in this review process. The items were then edited according to the feedback received from those focus group sessions.

3.6.3.3 Step 3: The Finalisation of MTALQ

As the last step, the full set of items (those from the original TALQ that were appropriate, those from the original TALQ that were modified and those new items that were reviewed and accepted by the focus groups as part of Stage 3, Step 2) was collated into a questionnaire to become the MTALQ. To ensure the quality of translation of the MTALQ from English to Arabic language, three PhD students, with Arabic language and cultural backgrounds are selected. They graduated from the schools of Education, at Australian universities and are fluent in Arabic and English. They agreed that the MTALQ is an accurate translation and did not point to any issues regarding the translation. The MTALQ was later administered online to all the teachers selected in the sample.

3.7 Analysis using Classical Test Theory

Classical test theory (CTT) was used for the analysis of the questionnaire items in MTALQ as well as for the evaluation of the reliability of the questionnaire. The following outlines the sequential steps that were taken using the CTT:

- Once the responses from the questionnaire MTALQ were returned, each item response was keyed for each participant into a data matrix suitable for transmission into SPSS.
- Then, each entry was inspected to ensure that it was a correct representation of the original data.
- After that, a scoring key was constructed to ensure there was one and only one correct answer with which to mark the MTALQ.
- Then, SPSS was used to score all the items in the MTALQ for all participants using the subroutine in SPSS called 'Reliability'.
- Next, the frequency of total score responses, mean, standard deviation and Coefficient Alpha (Reliability) for the entire questionnaire were calculated. Item-level responses, item means and standard deviations and item-to-total correlations for each item were also calculated.
- Finally, the summary results were tabulated for presentation.

The final version of the questionnaire (i.e., the MTALQ) was administered to all the teachers participating in the study. From these responses, CTT item and distractor analysis (including traditional measures of internal consistency reliability) was used to identify whether the items were functioning appropriately and whether the MTALQ was functioning as anticipated. This latter step formed an additional part of the collection of validity evidence (Best

& Kahn, 2004). The same responses were also used later to measure the baseline performance of KSA teachers in teacher assessment literacy.

3.8 Administration of the Survey

The administration of the Likert survey and the MTALQ was performed via the specific and systematic procedures mentioned earlier. The assessment literacy framework was primarily developed to better understand the knowledge of the teachers related to assessment standards and to make the NPST more operational and pragmatic for participating teachers. Most of the details about assessment knowledge and skills were gathered through the discussion of the assessment literacy framework with the focus groups. The MTALQ was then used. It comprised 36 items. After obtaining their consent to respond to the MTALQ, teachers participating in the survey responded to all 36 items.

3.9 Collection of Validity Evidence

The collection of validity evidence, especially for the refined assessment literacy framework and the MTALQ (referred to collectively from here as ‘instruments’), is inherently important. These instruments were validated through several different efforts across different stages. The participants (i.e., the focus group participants who volunteered from the original sample) provided a lot of input, helping to align the instruments to the knowledge and skills required of the NPST (content-related validity evidence). In addition, focus group participants were able to review the original TALQ and provide direct alignment of each item with its value in measuring the NPST. If an item did not fit (i.e., was not aligned), it was either modified or eliminated, with a new one constructed to replace it based on participant input. This suggests strong evidence for the alignment of the instrument with its intended purpose (content and construct and content-related evidence).

3.10 Summary

This chapter discussed the methodology, procedures and research design adopted to conduct this research. This entailed the explication and elaboration of the crucial phases of the study that were undertaken to address the study questions so that the key steps contained within each phase of the study is made clear and transparent. This discussion included the way the NPST elaborations were developed and collected into an assessment literacy framework and how the questionnaire (MTALQ) use in this study were developed. During this process, the focus group participants were involved at the various steps of development and were key to the validation processes that were used to help ensure the validity and reliability of various tools. The next chapter discusses the findings from the development and validation of the assessment literacy framework and MTALQ.

Chapter 4: Results

4.1 Introduction

The first aim of this study was to produce a pool of elaborations informed by a group of KSA primary teachers, secondary teachers, and intermediate teachers regarding what they considered to be the essential assessment knowledge, skills, and understanding required to be able to meet the assessment standards captured in the Practitioner and Advanced Standards for Teachers of the KSA (NPST). This chapter presents the Likert survey results and information obtained from focus group discussions and how these data and information are used in validating the assessment literacy framework and developing the MTALQ. Chapter 5 discusses the results of the application of MTALQ to participants in this study.

4.2 Survey Responses

The previous chapter has illustrated the process in which the standards associated with assessment in the KSA NPST were unpacked in a way that made explicit the skills and knowledge teachers might need to have to be able to demonstrate the assessment standards described within the NPST, referred to in this study as ‘elaborations’.

In gathering survey responses using a traditional Likert design, each of the elaborations developed by the researcher were presented to samples of teachers from the primary, intermediate and secondary groups. The survey asks them to rate each of the elaborations as either ‘Essential’, ‘Desirable’ or ‘Not Necessary’ according to the extent to which the elaboration is needed for the teacher to demonstrate the assessment standard. Recall, the elaborations were intended to clarify, operationalise, and make more understandable the assessment literacy standards of the NPST. Each statement in the survey had the following format: ‘To what extent

do you (or teachers at the nominated level) have to know or understand the elaboration to demonstrate the specified standard?’

The summary results from an analysis of the responses to these Likert questions are presented in Tables 4.1 and 4.2. The mean for each elaboration was generated by multiplying the number of ‘Essential’ responses by 2, the number of ‘Desirable’ responses by 1 and the number of ‘Not Necessary’ responses by 0. These results were then aggregated and divided by the number of respondents. Tables 4.1 and 4.2 summarise these results in intervals. For example, it can be seen from Table 4.1 that no elaboration had a mean score below 0.49 for either the Practitioner or Advanced career levels’ responses from the primary teachers. One elaboration (number 23) had a mean of between 0.5 and 0.99 at the Practitioner career level; Seventeen of the elaborations were rated between 1.0 and 1.49 for the Practitioner career level and 22 were rated between 1.0 and 1.49 for the Advanced career level. The remaining elaborations laid between 1.50 and 2.00 (14 at the Practitioner level and 11 at the Advanced level). Hence, most of the elaborations (31 for the Practitioner level and all 33 for the Advanced level) were rated as at least ‘desirable’ if not ‘essential’ according to the participants giving the review. These item numbers have not been included in the table because they were effectively identified as being essential and were not discussed by the focus groups.

Table 4.1

Range of Means for Elaborations Obtained from the Judgement of Primary Teachers on the Relative Importance of the Elaborations of the Practitioner and Advanced Career Levels of Standard 10 of the Saudi Arabian Professional Standards for Teachers

Means range	Practitioner career level		Advanced career level	
	Number of items	Elaboration item numbers	Number of items	Elaboration item numbers
0.0–0.49	0		0	
0.5–0.99	1	23	0	0
1.0–1.49	17	5, 13, 14, 16, 17, 19, 20, 21, 22, 24, 25, 27, 28, 29, 30, 31, 32.	22	33, 34, 36, 37, 38, 42, 43, 44, 45, 46, 47, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61
1.5–2.00	14		11	
Total number of elaborations	32		33	

Table 4.2 shows the results obtained from the survey of the secondary and intermediate teachers. The results were like those for the teachers at the Practitioner level. For example, teachers at the Practitioner level also rated the elaborations as ‘Desirable’ and ‘Essential’, while all 33 elaborations were rated as such for the Advanced level teachers. It can be seen from a comparison of the results in Tables 4.1 and 4.2 that the secondary and intermediate teachers rated more of the elaborations at a higher level than the primary teachers.

Table 4.2

Range of Means for Elaborations Obtained from the Judgement of Secondary and Intermediate Teachers on the Relative Importance of the Elaborations of the Practitioner and Advanced Levels of Standard 10 of the Saudi Arabian Professional Standards for Teachers

Means range	Practitioner career level elaborations		Advanced career level elaborations	
	Number of items	Elaboration numbers	Number of items	Elaboration numbers
0.00–0.49	0		0	
0.50–0.99	1	23	0	
1.00–1.49	14	9, 10, 12, 14, 15, 16, 21, 22, 24, 25, 29, 30, 31, 32	6	36, 40, 44, 51, 56, 57
1.50–2.00	17		27	
Total number of elaborations	32		33	

Following the analysis of the Likert survey, the school principals were contacted to identify participants for focus groups, who had already completed the (Likert) survey with the intention of discussing the results. Participants were to be asked their points of view regarding the rating of the elaborations in the survey; particularly those that were rated less highly by survey respondents. This exercise was carried out to refine the assessment literacy framework and make it a more valid reflection of the standards. The focus group discussions reviewed those elaborations where the mean score on the results from the survey was less than 1.5 (mean < 1.50) or right at the cut-off between ‘Essential’ (a rating of 2) and ‘Desirable’ (a rating of 1). The results from the various teacher focus groups are discussed in the next sections.

4.3 Primary Teachers' Focus Group Responses – Elaborations requiring changes

Fourteen primary school teachers voluntarily accepted the invitation to give their views about the results of the survey. Focus group sessions were conducted with the teachers via Zoom, as in-person meetings were impossible to carry out because of global COVID-19 health protocols that were in place at the time. It can be seen from Table 4.1 that there were 17 elaborations rated between 0.00 and 1.50. All but one of these elaborations were considered by respondents in the survey to be at least 'Desirable'. The participants were asked to provide reasons why they thought these elaborations were rated as less important than the elaborations that were deemed 'Essential'.

Primary teachers rated one elaboration (i.e., elaboration 23) as problematic and considered most of the elaborations as at least 'Desirable'. However, primary teachers argued that some elaborations needed to be reviewed before developing the questionnaire (MTALQ) as some wording and concepts were a concern.

The next sections consider some of the elaborations that were discussed in the first focus group.

To enhance organisation, the most interesting elaborations have been presented selectively. In addition, elaborations related to validity, reliability, feedback, and reporting students' achievement, which are regarded in the literature as essential skills and knowledge were discussed with a view to making amendments if necessary. The first set of elaborations 4.3.1 to 4.3.7 are elaborations that were considered and changes made on the basis of the discussion.

4.3.1 Elaboration 29

The primary teachers in the survey rated Elaboration 29, which was a response to the question ‘To what extent do Practitioner teachers have to “know the main outcomes from research on feedback” to demonstrate Practitioner Standard 10.4?’ This elaboration was rated as 1.11. The participants in the focus group explained that while they believed that feedback has the capacity to provide evidence for the teacher to engage students in their learning, the way the elaboration was currently worded put the focus on the outcomes from research on feedback. They felt that this was not necessary. For instance, one of the teachers said:

From my standpoint I think that teachers must know the requirements of feedback which take place in the practical instruction and are applicable to use but not those outcomes from research and its related definitions or details.

In a similar vein, another participant suggested that the elaboration should be eliminated from the assessment literacy framework altogether by arguing that ‘there is no need to make a teacher a research centre!’ This response obtained consensus among the other participants in the focus group. Upon reflection, the wording of the elaboration may be the cause of some of the confusion that led to this elaboration being rated lower than expected by primary teachers. The intention was to focus on the importance of feedback, which the focus group participants thought was essential; but the way the elaboration was worded prompted these participants to focus on the use of research. While it could not be removed at that point in the survey, the elaboration will be revised for any future administration to ‘To what extent do Practitioner teachers have to “know and use feedback to inform their teaching”?’ The focus group was unanimous in their belief that this would remove the current ambiguities.

The participants consider feedback a critical skill in assessment. However, the Advanced and Practitioner level teachers indicated that their workload might hinder their ability to follow the latest research on feedback, especially in the context that this assessment skill is considered relatively low in importance in comparison to all the other aspects of assessment literacy. These discussions assisted in the development of a valid questionnaire (MTALQ) to measure KSA teachers' assessment literacy. This was because these discussions with the teacher participants helped to enhance clarity regarding the alignment and value of many aspects of the elaborations in anticipation of developing a measurement questionnaire aligned to the NPST. Also, the discussion helped to interpret the test results and collect more validity evidence on the questionnaire (MTALQ) and establish connections between the data gathered from such discussion and the results.

4.3.2 Elaboration 57

The mean of this elaboration was 1.00. It represented the results of the survey response to the question 'To what extent do Advanced teachers have to "know and understand how to interpret the results of large-scale testing programs"?' This elaboration relates to Advanced Standard 10.4. The mean of 1.00 rates it as 'Desirable'. The participants in the focus group considered that the elaboration was more essential for teachers at the Expert standard rather than the Practitioner and Advanced standards. One participant stated that 'According to my way of thinking, this skill is a key aim of the expert teacher, so I suggest removing it from the assessment literacy framework.' During the discussion, a second participant echoed this response when he stated that 'Our education system doesn't require Advanced teachers to interpret the test results as it seems that it is some specialist's job.'

What is interesting in this response is that while these participants argued for this elaboration to be removed from the assessment literacy framework for Practitioner and Advanced standards, it was rated on average as being ‘Desirable’ by the primary teacher respondents to the survey. When this was pointed out to the participants, they decided that it should remain in the assessment literacy framework, but there would be an expectation that any items that related to this elaboration would prove relatively difficult for teachers to demonstrate at the Practitioner or Advanced Levels.

In brief, this elaboration was considered by participants to be very close to being what can be described as a problematic elaboration. Teachers were of the view that having the skill to interpret results of large-scale testing programs might be useful, but because of the limitations placed on gaining access to such information, these types of data literacy skills were more appropriate for someone who was a specialist in this area. Nonetheless, these discussions help provide views on the usefulness of these skills and in providing evidence to support the inclusion of the elaboration in the assessment literacy framework and its assessment through items in the MTALQ.

4.3.3 Elaboration 23

The participants in the primary focus group responded to the question related to elaboration 23 (i.e., ‘To what extent do Practitioner teachers have to “know and understand the meaning of some of the more commonly used statistical terms used for analysing and summarising student performance data”?’). The mean from the survey was 0.89, which indicates that this was the only elaboration that had an overall rating less than ‘Desirable’.

Even though this was the elaboration that was rated as least important by the survey respondents, the focus group participants agreed that having this basic knowledge of statistical

terms was beneficial for all teachers at all levels. In this regard, one of the participants commented on this to his colleagues: ‘Is it necessary for Practitioner teachers to know all the statistical terms or just the basic statistical terms around assessment?’ Another participant responded by saying: ‘In our education system, there is not a significant difference between Advanced and Practitioner teachers in terms of their knowledge and understanding of statistical terms.’ Another participant concurred: ‘I agree with my colleagues that the Practitioner teacher must know basic knowledge in relation to the statistical terms.’

Such views are consistent with KSA NPST Standard 10.4, which is associated with professional knowledge and under the sub-standard, which places emphasis on the understanding of the basic knowledge for collecting, analysis, and interpreting the data for Practitioner teachers (see Appendix E). In response to this discussion, the decision was made to revise Elaboration 23 to ‘Practitioner teachers must know and understand the meaning of basic statistical terms used for analysing and summarising student performance data.’

This elaboration was the only one that was rated as definitely problematic. The teachers who responded to the survey did not all agree that they should know and understand the more commonly used statistical terms used for analysing and summarising student performance data, but the feeling in the focus group was that basic knowledge in relation to statistics was useful. This knowledge can be used to ensure that any items that target this skill are relatively easy or less cognitively demanding.

4.3.4 Elaboration 43

This elaboration was related to teachers’ knowing and understanding of the content, purpose and structure of major state, national and international testing programs. It was rated as 1.17. which located it in the ‘Desirable’ classification. The participants reasoned that knowing

and understanding various components of the major testing programs would be problematic and burdensome for teachers. The benefits of knowing about these programs were also considered to be of little value to teachers. The participants suggested that the elaboration be changed to ‘knowing the content and purpose of national and international testing programs’.

From discussion of the focus group, the reason that the elaboration expressed as ‘teachers should know and understand the content, purpose and structure of major state, national and international testing programs’ was considered problematic by the respondents to the survey was attributable to one word: ‘structure’. The participants in the focus group believed teachers should know the content and the purpose of such testing programs but should not necessarily have to know the ‘the structure’ of such programs. The focus group participants were satisfied that with the change the elaboration should be retained in the literacy assessment framework with such a change. Hence, this adjusted elaboration was included in the refined assessment literacy framework as ‘Teachers should know and understand the content, and purpose of major state, national and international testing programs.’

4.3.5 Elaboration 14

Focus group participants discussed Elaboration 14, which was ‘Practitioner teachers have to know and understand the requirements, structure, purpose and action of conducting formative assessment appropriately’ to demonstrate Practitioner Standard 10.2. Primary teachers in the survey rated this elaboration as 1.33. This elaboration was included because the literature (Black & Wiliam, 2009; R. Stiggins, 2009) indicates that formative assessment is an essential skill for teachers. However, in the survey, the teachers did not consider it to be essential. Instead, their responses indicated that the knowledge of and skill to conduct formative assessment was considered merely desirable by the survey respondents. The focus group discussion revealed that

teachers were not familiar with the descriptors associated with the term ‘formative assessment’ in the elaboration. That is, they were confused by prefacing ‘formative assessment’ with ‘requirements, structure, purpose and action of conducting (formative assessment)’. As the use of the words describing the type of formative assessment evidently caused some confusion for teachers, the decision was made by the participants in the focus group to only focus on teachers needing to know and understand the meaning of formative assessment.

In a few words, the focus group discussions illustrated that KSA teachers have limited confidence in their understanding of the meaning of formative assessment. Teachers mixed up the definitions of continuous assessment, summative assessment, formal assessment, informal assessment and formative assessment. Considering their lack of knowledge on various concepts and definitions of the types of assessment, it is not surprising that primary teachers have difficulty when having to implement them. This is one reason why the participants in the focus group suggested the incorporation of more items related to formative assessment being included in the MTALQ.

The recommendation that was accepted by the focus group was to include the following elaboration in the refined assessment literacy framework: ‘Practitioner teachers have to know and understand the meaning of conducting formative assessment appropriately.’

4.3.6 Elaboration 16

The mean of responses to the question related to Elaboration 16, ‘Practitioner teachers have to know and understand the main purpose of self- and peer-assessment’, was 1.22. It was considered ‘Desirable’ by respondents to the survey. One of the focus group participants said: ‘In my judgement, to tighten up the assessment literacy framework, we have to combine the definition of self- and peer-assessment based on their purpose.’

Since definitions of self- and peer-assessment and their purposes are not separate skills, the decision was made to incorporate their definitions and purposes into a single category in the refined assessment literacy framework. It emerged from the discussion in the focus group that teachers were not clear on the meaning of self- and peer-assessment. Given that they did not know the definition, it was understandable that they could not comprehend the function and purpose of these kinds of assessments.

The ability to engage students in their learning is seen as a vital skill. Under the NPST, the standard linked to assessment requires teachers to be able to engage students in their own assessment of learning. In this regard, Black and Wiliam (2010) highlighted that the self- and peer-assessment used as formative assessment methods are useful strategies to enhance student learning outcomes. This is why the elaboration was included in the assessment literacy framework. The response to the survey by teachers, followed by the focus group discussion, suggests that the elaboration should be rewritten as ‘Practitioner teachers must “Know and understand the definition and the main purpose of self- and peer-assessment”.’

4.3.7 Elaboration 51

The question in the survey related to Elaboration 51 was ‘To what extent do Advanced teachers have to know and understand how to provide opportunity for students to construct success criteria’? This was designed to ascertain the extent to which respondents to the survey felt that the knowledge and understanding of generating and conveying success criteria to students was essential, desirable or not important in relation to advanced Standard 10.3. It elicited a mean rating of 1.11 on the survey, which was classified as ‘Desirable’. This item was focused on identifying the extent to which the standard requires teachers to know and understand how to create and convey the assessment criteria for success to students to provide opportunities

for students to co-construct the assessment criteria. In response to the results from the survey on this elaboration, one of the participants in the focus group expressed: ‘I think teachers do not have enough time to provide the opportunity for students to construct the criteria although it is important to provide the opportunity for students consistently to know about the criteria.’

This comment prompted the rest of the teachers to share more about some of the obstacles that teachers in primary schools face in relation to providing opportunities for students to construct and convey success criteria. Participants in the focus group held strong views that students in primary schools have insufficient knowledge on how to co-construct successful assessment criteria and teachers’ lack of time within a crowded curriculum inhibits them in developing this knowledge. This said, participants agreed that Advanced teachers should be aware of the significance of providing students with these opportunities and sharing the assessment success criteria to give them autonomy in their learning and to engage them in the learning process. In response to the teachers’ views, the decision was then made to revise this elaboration from ‘know and understand how to provide opportunity ...’ to ‘know and understand the importance of providing the opportunity ...’.

Participants in the focus group considered this elaboration to be a desirable skill. However, the discussions with teachers also shed light on the other issues that were discussed previously in Chapter 2; issues that might hinder the ability of teachers to provide students with opportunities to construct and share success criteria. These issues included time constraints, teachers’ workloads and the largely traditional structure of the KSA education system.

4.4 Primary Teachers' Focus Group Responses - Elaborations not Requiring Amendments

The focus group participants also considered certain items that were scored relatively low in the Likert survey but after consideration, they did not require changes according to the conclusions drawn by the participants. This section discusses these elaborations and why the participants felt that they did not need to be amended or modified.

4.4.1 Elaboration 20

The question in the survey related to elaboration 20 was 'To what extent do Practitioner teachers have to "know and understand of some feedback techniques to engage students in the instruction"?' The elaboration related to Practitioner Standard 10.3 and the results showed a mean of 1.22 on the survey. Participants in the focus group were generally of the view that this elaboration should remain unchanged as it aptly captured the kinds of skills and knowledge required of teachers. For instance, one participant expressed that it was indeed important that 'teachers need to know the requirements of feedback', especially when feedback plays such a crucial role in enhancing student learning. Another participant reasoned that Elaboration 20 should remain unchanged because it was essential that teachers were knowledgeable of the contextual use of the relevant techniques in assessment.

In essence, the ability to know and understand feedback techniques to engage students during classroom instruction was considered 'Desirable'. However, all participants in the focus group agreed that this elaboration did not require any revision or review because they all felt that feedback and its related strategies are essential in enhancing student achievement.

4.4.2 Elaborations 21, 22, 24 and 25

Participants also were of the view that elaborations 21, 22, 24 and 25 (with means of 1.17, 1.17, 1.17 and 1.39, respectively), did not require changes. The questions in the survey relating to these elaborations were:

- Elaboration 21. To what extent do Practitioner teachers have to ‘know the definition of data literacy’?
- Elaboration 22. To what extent do Practitioner teachers have to ‘know and understand that there are a range of skills needed to transform data into information, evidence and strategies for teaching and learning’?
- Elaboration 24. To what extent do Practitioner teachers have to ‘know the meaning of data, information and evidence, as well understanding that evidence is used to inform educational practice and reporting’?
- Elaboration 25. To what extent do Practitioner teachers have to ‘know and understand how to collect and organise data for classroom use’?

All elaborations related to Practitioner Standard 10.4.

The consensus among focus group participants was that data literacy and the skills to convert data to information and evidence are critical skills for teachers to have. This was strongly endorsed by the participants in relation to the context of educational assessment, in which data are integral in informing teaching that fosters more effective learning.

All these elaborations, which were related to data literacy, were seen as critical, as all participants accepted that these aspects of data use in assessment literacy are important in classroom assessment practices.

4.4.3 Elaborations 30 and 31

Elaborations 30 (mean of 1.33) and 31 (mean of 1.22) were also considered to be very important by the focus group teachers. These elaborations were: ‘To what extent do Practitioner teachers have to “know and understand the purposes of reporting”?’ and ‘To what extent do Practitioner teachers have to “know and understand some of the more common reporting strategies used in schools for disseminating assessment information to parents”?’ , respectively. Both elaborations relate to Practitioner Standard 10.5.

The participants were of the view that there were no significant differences between Advanced and Practitioner teachers in relation to the reporting of student achievement and the importance of reporting against standards. Given this, participants were unanimous in asserting that there was no need to change these two elaborations.

In brief, the discussions on reporting and common reporting strategies, skills and knowledge provided more insights regarding the communication of assessment results. What emerged was a sound understanding of the degree of the knowledge of KSA teachers in this area. This piqued the researcher’s interest to see whether the MTALQ results on questions relating to reporting also showed the kind of knowledge and understanding of reporting and communication of results that emerged during the discussions.

4.4.4 Elaboration 32

The survey question related to Elaboration 32 was ‘To what extent do Practitioner teachers have to “know and understand the meaning of the grades that are used to summarise performance”?’ It referred to Practitioner Standard 10.5 and elicited a mean of 1.44. The focus group believed that this elaboration required no change. Focus group participants stated that it was fundamentally essential for teachers to recognise the aim behind grading.

Another reason why participants felt that this elaboration required no change was because it underscored the role of the teacher ‘to communicate students’ achievement with stakeholders’, which all participants believed is an important aspect of assessment. Furthermore, the need to communicate student assessment to stakeholders requires these stakeholders to understand the meaning of grading practices. As one participant stressed, ‘Differences in grades in different subjects need to be understood by the counsellor and teachers.’

The participants considered this elaboration to be a basic skill required by all teachers and the outcome was that the elaboration was retained, unchanged.

4.4.5 Elaboration 36

The question in the survey related to Elaboration 36 was ‘To what extent do Advanced teachers have to “Know and understand how to develop items and tasks based on analyses of outcomes in terms of the factors that may affect these items”?’ The elaboration was aligned to Standard 10.1 and is an Advanced standard. It had a mean rating from the survey of 1.33. It is another elaboration that participants indicated did not need amendment. The participants felt that teachers at both Advanced and Practitioner levels should possess this aspect of assessment literacy. This sentiment was captured by a participant who stated: ‘The Advanced teacher and even Practitioner teacher should know about how to develop items and tasks based on analyses of outcomes in terms of the factors that may affect these items.’

The outcome of the focus group was that Elaboration 36 should be retained in its current state, so it was included in the refined assessment literacy framework unchanged.

4.4.6 Elaboration 37

The question related to Elaboration 37 was ‘To what extent do Advanced teachers have to “know and understand how to develop assessment items and tasks to ensure high quality in their

validity and reliability”?’ The elaboration is aligned to Standard 10.1 and is an advanced Standard. The elaboration had a mean of 1.28. The participants in the focus group strongly agreed that this knowledge should be possessed by Advanced teachers. They expressed the view that these teachers should ‘be capable of developing assessment tasks’ that are of high quality and contribute to the consistency and fairness of the assessment.

The participants in the focus group considered the ability to produce high-quality assessments that have strong reliability and validity as critical, even if teachers in the survey rated it as ‘Desirable’. In addition, the discussion of this elaboration in the focus group proved beneficial to identifying potential gaps between teacher knowledge and actual teacher assessment practices when it came to reliability and validity.

4.4.7 Elaboration 44

The question related to Elaboration 44 was ‘To what extent do Advanced teachers have to “Know various ways to record results to capture an image of student performance that can be compared against performance standards”?’ It related to Advanced Standard 10.2 and resulted in a mean rating of 1.28 on the survey. Like the previous elaboration, participants in the focus group responded by indicating that this was another essential aspect of assessment literacy and thus required no change. Participants stressed the ‘importance of knowing how to use assessment rubrics and checklists.’

The importance of assessing student performance against standards (Tognolini & Stanley, 2007) is a critical skill for all contemporary teachers to have. This is primarily because of the current movement towards standards-referenced assessment systems. It is important to know the degree to which the teachers in the KSA are ready to transition to such a system and achieve the NPST.

4.4.8 Elaboration 47

The question related to Elaboration 47 was ‘To what extent do Advanced teachers have to “know and understand the definition of self-regulation”?’ It aligned with Advanced Standard 10.3 and returned a mean of 1.22 on the survey. Participants in the focus group all agreed that this was another aspect of assessments skills and knowledge that Advanced teachers should possess. One participant echoed the sentiment of the group when he volunteered that ‘Advanced teachers should know about self-regulation and support his/her colleagues’ as was required in this elaboration.

4.4.9 Elaboration 56

The question related to Elaboration 56 was ‘To what extent do Advanced teachers have to “Know and understand that different kinds of data exist and have a basic understanding of which kind of data is useful for particular decisions”?’ The elaboration was designed to address Advanced Standard 10.4 and had a mean of 1.39 on the survey. The focus group participants were unanimously of the view that it did not need to be changed. One participant stated that all teachers should have the ability to ‘make comparisons with performance standards’ and know how to employ student data in making pedagogical decisions. This was also seen as important because participants felt that assessment should be designed in such a manner that it is ‘consistent with students’ capacity’.

Simply put, ‘knowing and understanding that different kinds of data exist and have a basic understanding of which kind of data is useful for particular decisions’ is critical for those teachers aspiring to the Advanced Standard 10.4. In addition, the participants were strongly of the view that these teachers should support Practitioner teachers.

The next section considers comments from participants involved in the second focus group. While the first focus group used primary school teachers only, the second focus group comprised secondary and intermediate teachers. The secondary and intermediate teachers were combined into a “non-primary group” to maximise the number of participants in each group and maximise the power of the results drawn from these groups.

4.5 Secondary and Intermediate Teachers’ Focus Group Responses - Elaborations Requiring Amendments

4.5.1 Elaborations Requiring Amendments

The secondary and intermediate teachers in the focus group (i.e., comprised 18 secondary and intermediate schoolteachers) proposed making some changes to Elaborations 14 and 16. Like the participants in the primary focus group, they suggested changing the concept of ‘constructive evaluation’ to ‘formative evaluation’. Their reasoning behind this was that formative assessment was more widely used and understood in the KSA education system. In addition, the participants recommended that both the definition and main purpose of self- and peer-assessment be combined, in much the same way as suggested by the participants in the primary focus group.

4.5.2 Elaborations not Requiring Amendments

As described earlier when reviewing the results from the primary focus group, it was decided to not only discuss the problematic elaborations with the participants but also provide them with the opportunity to shed light on the reasons behind the ratings of some elaborations that were considered less than essential, even though these skills and knowledge were deemed essential in the body of research on assessment literacy. The outcomes from these discussions would potentially be helpful when interpreting the patterns of responses to the questions

targeting elaborations assessed through the developed questionnaire, the MTALQ. This section discusses these elaborations and provides some reasons as to why the participants felt that the elaborations did not need to be amended.

4.5.3 Elaboration 21

The question relating to Elaboration 21 was ‘To what extent do Practitioner teachers have to “know the definition of data literacy”?’ The elaboration related to Practitioner Standard 10.4 and it had a rating of 1.25. Secondary teachers who responded to the survey indicated that this elaboration related to knowledge and was primarily considered essential. One of the participants reflected the group’s sentiment, stating that ‘knowing the definition of data literacy is crucial.’

Like primary teachers, secondary teachers value the assessment and data literacy skills that foster student learning. They can see a clear link between the knowledge and skill required and the Standard.

4.5.4 Elaboration 22

The question related to Elaboration 22 was ‘To what extent do Practitioner teachers have to “know and understand that there are a range of skills needed to transform data into information, evidence and strategies for teaching and learning”?’ The elaboration enables the teacher to demonstrate Practitioner Standard 10.4 and was rated as 1.25 on the survey. Secondary, Intermediate and Primary school participants in the focus groups agreed that this elaboration required an aspect of assessment literacy that was essential, and therefore felt that the elaboration required no change. One of the participants made the point that teachers ‘need to know the difference between data, information and evidence as articulated in the elaboration.’

Secondary and intermediate participants agreed that it was essential to know and understand that there are a range of skills needed to transform data into information, evidence

and strategies for teaching and learning to achieve NPST. This discussion was very important as it included many skills in assessment that teachers would need to conduct assessment effectively.

4.5.5 Elaboration 30

The question related to Elaboration 30 was ‘To what extent do Practitioner teachers have to “know and understand the purposes of reporting”?’ It addressed Practitioner Standard 10.5 and returned a mean of 1.46 on the survey. The secondary and intermediate participants in the focus group felt that this elaboration was another of those aspects of assessment literacy that are crucial for Practitioner teachers. One participant’s remarks captured this sentiment: ‘It is essential for teachers to know how to communicate students’ achievements.’ The prevailing view of focus group participants was that this elaboration did not require any amendments. Both Secondary and Intermediate teachers who participated in the study consider most skills related to reporting as essential and they also consider the teachers’ knowledge related to the purposes of reporting, as at least desirable—if not more probably essential.

4.5.6 Elaboration 31

The question related to Elaboration 31 was ‘To what extent do Practitioner teachers have to “know and understand some of the more common reporting strategies used in schools for disseminating assessment information to parents”?’ The elaboration addressed Practitioner Standard 10.5 and the mean rating from the survey was 1.38. The focus group participants believed that teachers at the Practitioner standard must know about the reporting strategies employed by schools to disseminate assessment information to parents. As one participant stressed: ‘it is indeed essential that teachers know the strengths and weaknesses of the reporting strategies, so they know how best to communicate the assessment information to students’ parents.’ The participants of this focus group were consistently of the view that knowledge of

reporting strategies was important, and so they recommended that the elaboration be retained in the refined assessment literacy framework, unchanged.

4.5.7 Elaboration 32

The question related to Elaboration 32 was ‘To what extent do Practitioner teachers have to “know and understand the meaning of the grades that are used to summarise performance”?’ It related to Practitioner Standard 10.5 was rated with a mean of 1.23 by secondary and intermediate teachers completing the survey. Judging from the sentiments expressed by the teachers, this is yet another aspect of assessment literacy that is essential and therefore required no modifications in terms of the elaboration. As one teacher stressed, ‘I think it is highly important for any Practitioner teacher to understand how grades are used to reflect student performance and learning.’ Like primary teachers, secondary and intermediate teachers shared views as to why grading was one of the critical skills related to assessment literacy.

4.5.8 Elaboration 36

The question in the survey related to Elaboration 36 was ‘To what extent do Advanced teachers have to “know and understand how to develop items and tasks based on analyses of outcomes in terms of the factors that may affect these items”?’ It was linked directly to Advanced Standard 10.1 and had a mean rating of 1.46 on the survey. In terms of the knowledge of assessment that is expected for teachers at the Advanced standard, the focus group of secondary and intermediate teachers shared the view that it is essential for advanced teachers to have more advanced knowledge of designing assessment tasks that is informed by analyses of student assessment outcomes using such tasks. This group view was aptly captured by the remarks of one participant, who said:

Certainly, this knowledge should be possessed by Advanced teachers because it is important to know the factors that threaten the validity of the assessments. So, these could include the item design in terms of its clarity, appropriateness, quality, and the duration given to students to complete the assessment tasks satisfactorily.

Secondary and intermediate participants in the focus group were of the view that Advanced teachers must ‘know and understand how to develop items and tasks based on analyses of outcomes in terms of the factors that may affect these items.’

4.5.9 Elaboration 44

Elaboration 44 in the survey was expressed as: ‘To what extent do Advanced teachers have to “know various ways to record results to capture an image of student performance that can be compared against performance standards”?’ It was designed to evoke a response in relation to Advanced Standard 10.2 and the resulting mean from the survey was 1.42. The focus group of secondary and intermediate participants made it clear that this was another elaboration that should not be changed. They felt it essential for Advanced teachers to have the knowledge to be able to document their students’ performance and be able to benchmark this against the broader performance standards. As one participant expressed:

It is of course important to record our students’ performance but it is also important for advanced level teachers to be able to also compare their students’ performance with the set performance standards so they can gauge how the student is faring against these standards.

This focus group was united in agreeing on the importance for Advanced teachers to ‘know various ways to record results to capture an image of student performance that can referenced to against performance standards.’

4.5.10 Elaboration 56

Elaboration 56 was expressed in the survey as: ‘To what extent do Advanced teachers have to “know and understand that different kinds of data exist and have a basic understanding of which kind of data is useful for particular decisions”?’ The elaboration addressed Advanced Standard 10.4 and its mean rating on the survey was 1.38. As with the previous elaborations, the focus group participants were unanimously of the view that this was yet another aspect of assessment literacy that Advanced teachers should be well-versed in. While knowing the different kinds of assessment data that may exist is important for all teachers, Advanced teachers should have a level of understanding of the kinds of data that would be useful in informing different decisions. However, even Advanced teachers need support to be able to perform this task. As one participant expressed: ‘Teachers need support from the school management as it is hard to track the different kinds of assessment data of many students.’

In conclusion, comparing secondary, intermediate and primary results from the survey and from the focus groups teachers’ results indicated that primary, secondary and intermediate teachers rated all elaborations generally at the same level and both cohorts rated just one elaboration—that is, the elaboration related to statistical terminology—as ‘problematic’.

All sectors (primary, intermediate and secondary) were of the view that the elaborations relating to formative assessment and self- and peer-assessment needed to be altered to align them with the standards as contextualised for the KSA.

The changes to the elaborations agreed upon by the focus groups were incorporated into the assessment literacy framework and a refined assessment literacy framework was developed that is more aligned with the KSA NPST and hence is more valid than the original assessment literacy framework. This refined framework is provided in Appendix D.

The next section aligns the items from the original TALQ with the elaborations captured in the refined assessment literacy framework.

4.6 Aligning Items of the Original TALQ with the Elaborations that Comprise the Refined Assessment Literacy Framework

This section presents information about the alignment of the original TALQ with the elaborations presented in the refined assessment literacy framework (see Appendix D). The process is part of the collection of the validation evidence needed to support the decisions resulting from the responses of candidates to the original TALQ (see Chapter 6 for content validation) and or their modifications for a more specific MTALQ. This validation stage of the project involved contacting the principals of the schools that had agreed to participate in the study and seeking permission to send invitations to staff to become involved in this stage of the project. Involvement meant reviewing each of the items comprising the original TALQ against the elaborations comprising the refined assessment literacy framework. This alignment activity would be a primary source of content and process validity evidence.

Before this stage commenced, the participants were provided with the refined assessment literacy framework and original TALQ. This was to enable the participants to examine the items closely so that they had an idea of what each of the original TALQ items was assessing. In total, 35 secondary, intermediate and primary teachers had volunteered to participate in this third focus group also agreed to attend (by Zoom) a single focus group meeting where they could review the items. Table 4.3 summarises the key justifications provided by participants.

Table 4.3

Aligning a Sample of the Original TALQ Items with Elaborations of the Refined Assessment Literacy Framework and the NPST

Standards

Item numbers in the TALQ	Standards in the TALQ	Elaborations in the refined assessment literacy framework	NPST Standard
1, 2, 3,	Standard 1: Choosing assessment methods	• Know and understand some of the more common purposes of assessment, including how the assessment is likely to be used and what are likely to be the intended and unintended consequences of using these assessments • Know and understand the meaning of reliability and validity in relation of selecting assessment tasks	Standard 10.1 Preparing evaluation tools
6, 7, 8,	Standard 2: Developing assessment methods	• know and understand how to select assessment tasks which clearly related to learning objectives • Know and understand how to develop assessment items and tasks in order to ensure high quality in their validity and reliability • Know the relationship between item difficulty and the level of cognition based and can use classification systems and taxonomies in building assessments	Standard 10.1 Preparing evaluation tools
12, 15	Standard 3: Administering, scoring and interpreting assessment results	• Know the process on how to accurately mark assessments and collect, record student results • Know and understand the illegal, inappropriate, and unethical assessment practices	Standard 10.2 Applying evaluation
21	Standard 5: Using assessment in grading	• Know and understand how to use a range of assessment items that are fit-for-purpose.	Standard 10.2 Applying evaluation

Item numbers in the TALQ	Standards in the TALQ	Elaborations in the refined assessment literacy framework	NPST Standard
20	Standard 4: Using assessment results for decision making	Know and understand the assumptions that underpin educational assessment. such as: assessment must be fair and professional judgment	Standard 10.4 Employing evaluation results
4	Standard 1: Choosing assessment methods	Know and understand some of the more common assessment dichotomies used in education and the meaning of relevant terms associated with them. such as (assessment for and of learning,	Standard 10.1 Preparing evaluation tools
9, 10	Standard 2: Developing assessment methods	- Know how to write, develop and use a range of item types and assessment tasks that are technically sound, editorially correct and that appropriately assess the outcomes and performance standard levels being assessed - Know and understand how to design summative assessments tools that will enable the results to be used to improve teaching and learning	Standard 10.1 Preparing evaluation tools
5	Standard 1: Choosing assessment methods	Know how to administer a range of assessments and understand the need for consistent instructions to ensure a consistent testing environment	Standard 10.1 Preparing evaluation tools
11, 13	Standard 3: Administering, scoring and interpreting assessment results	- Know and understand how to collect and organise data for classroom use - Know and understand that there are a range of skills needed to transform data into information, evidence and strategies for teaching and learning	Standard 10.2 Applying evaluation
22	Standard 5: Using assessment in grading	know and understand the meaning of basic statistical terms used for analysing and summarising student performance data	Standard 10.2 Applying evaluation

Item numbers in the TALQ	Standards in the TALQ	Elaborations in the refined assessment literacy framework	NPST Standard
17	Standard 4: Using assessment results for decision making	• Know and understand that different kinds of data exist and have a basic understanding of which kind of data is useful for particular decisions. • Know and understand how to generate teaching and learning interventions from an analysis of assessment data.	Standard 10.4 Employing evaluation results
23, 24, 25	Standard 5: Using assessment in grading	• Know and understands the meaning of the grades that are used to summarise performance. • Demonstrate ways that results of student assessments are recorded and used to arrive at an overall grade representing a performance standard	Standard 10.2 Applying evaluation
18	Standard 4: Using assessment results for decision making	• Know various ways to record results to capture an image of student performance that can be compared against performance standards	Standard 10.4 Employing evaluation results
27, 29	Standard 6: Communicating assessment results	• Knows and understands the importance of collecting reliable and valid assessment data about the achievement of students and report it in a manner which is consistent with the reporting strategy(ies) used within a school • Know and understand some of the ways that assessment methods can be misused and misinterpreted	Standard 10.5 Preparing Evaluation Reports
32, 34	Standard 7: Recognizing unethical practices	• Know and understand and carry out their legal and ethical responsibilities in assessment at the school- and classroom levels	

Table 4.3 clearly articulates the alignment of the selected original TALQ items to the elaborations of the refined assessment framework and the NPTS. Such strong agreement is both a central argument to the content validity of the refined assessment literacy framework and the selected TALQ items but also indicative of a way to measure attainment of the NPTS standards going forward.

The meeting started with participants being made aware of the goals of the focus group meeting and then moved on to clarify the focus group meeting procedures, including describing how the participants could provide their comments on the items of the original TALQ. Out of 35 items in the original TALQ, the participant teachers identified 26 items as aligning with the elaborations of the refined assessment literacy framework to measure teacher assessment literacy and related standards (NPST). This is presented in Table 4.3.

The processes of eliminating or modifying items followed specific instructions that considered both the refined assessment literacy framework and the NPST. The items in the original TALQ must assess elaborations in the refined assessment literacy framework and be directly linked to at least one standard of the NPST as follows:

- Standard 10.1 Preparing evaluation tools
- Standard 10.2 Applying evaluation
- Standard 10.3 Engaging students in evaluation processes
- Standard 10.4 Employing evaluation results
- Standard 10.5 Preparing Evaluation Reports

The following are some of the comments made by the participants regarding how they made the link between the items in the original TALQ and those related elaborations in the

refined assessment literacy framework as well as NPST. These are designed to clarify the validation process in this current stage.

The participants indicated that item 21 (see Figure 4.1) was explicitly aligned to NPST Standard 10.4 and addressed teachers' ability to construct test items that were highly likely to produce reliable results.

Figure 4.1

Item 21 of the Original TALQ

- 21- Of the following, which choice typically provides the most reliable student-performance information a teacher might consider when assigning a unit grade?
- a. Scores from a teacher-made test containing two or three essay questions related directly to instructional objectives of the unit.
 - b. Scores from a teacher-made 20 item multiple-choice test designed to measure the specific instructional objectives of the unit.
 - c. Oral responses to questions asked in class of each student over the course of the unit.
 - d. Daily grades designed to indicate the quality of in-class participation during regular instruction.

On the contrary, some of the participant teachers strongly recommended deleting nine items from the original TALQ. The reasons they provided included that the original TALQ items were neither contextual nor related to NPST, and that they had found no direct relationship with the refined assessment literacy framework. One example was Item 19 in the original TALQ (Figure 4.2).

Figure 4.2

Item 19 of the Original TALQ

19. Students' scores on standardized tests are sometimes inconsistent with their performances on classroom assessments, e.g. teacher tests or other in-class activities. Which of the following is **NOT** a reasonable explanation for such discrepancies?
- a. Some students freeze up on standardized tests, but they do fine on classroom assessments.
 - b. Students often take standardized test less seriously than they take classroom assessment.
 - c. Standardized tests measure only recall of information while classroom assessments measure more complex thinking.
 - d. Standardized tests may have less curriculum validity than classroom assessment

Here are two sample comments from participants regarding the appropriateness of this item:

Comment 1: There doesn't seem to be any relevance to our educational context.

Comment 2: There is no relation between this item and the skills in the refined assessment literacy framework.

The participants asserted that using standardised test data was not consistent with the KSA educational context because of the limited access of teachers to these types of tests and related data, despite the importance of data literacy. As a result, the participants voted to remove Item 19 and replace it with other items that used assessment data to assess student achievement. The discussion above may be an indication of the robustness and explicitness of the deletion process, which considered the alignment and coherence of three tools (the original TALQ, the refined assessment literacy framework and NPST) as well as their potential overlap and interdependence.

Table 4.3 shows the alignment between the original TALQ items, the refined assessment literacy framework and NPST. For example, it shows that most of the participants argued that items 1, 2, 3, 6, 7, 8, 12, 15 and 21 (see Appendix F for original TALQ) were aligned with the ‘assessment methods and the lesson objectives’ elaborations on the refined assessment literacy framework; also, that Item 17 assessed the ‘skills and knowledge of validity’. Participants expressed their agreement that Item 20 in the original TALQ assessed one aspect of educational assessment. This item assessed teachers’ ability to use assessment results as an element for decision-making to support their professional judgement about student achievement.

Participants considered these items to be very important because teachers must not only know how to choose and develop assessment tasks that assess outcomes but, in addition, must provide student responses that can be used by the teacher to improve student learning through feedback. Participants argued that Item 5 was also very important because it aimed to measure teachers’ knowledge of students’ higher-order thinking.

Participants considered Item 10 (Figure 4.3) as an appropriate item to assess teachers’ ability to use assessment data to group students who have the same challenges.

Figure 4.3

Item 10 of the Original TALQ

10. Several students in Ms. Atwell's class got low scores on her end-of-unit test in doing multi-step story problems in mathematics. She wanted to know which students were having similar problems so she could group them for instruction. Which assessment strategy would be best for her to use for grouping students?
- a. Use the test provided in the "teacher's guide."
 - b. Have the students take a test that has separate items for each step of the process.
 - c. Look at the student's records and standardized test scores to see which topics the students had not performed well on previously.
 - d. Give students story problems to complete and have them show their work.

Participants were united in their opinion that items 21, 22 and 24 (Figures 4.4, 4.5 and 4.6) respectively are all necessary items to assess teachers' knowledge and skills against elaborations in the refined assessment literacy framework that are related to the vital standards in NPST (Standard 10.4: Employing Evaluation Results and Standard 10.2 Applying evaluation).

Figure 4.4

Item 21 of the Original TALQ

21- Of the following, which choice typically provides the most reliable student-performance information a teacher might consider when assigning a unit grade?

- a. Scores from a teacher-made test containing two or three essay questions related directly to instructional objectives of the unit.
- b. Scores from a teacher-made 20 item multiple-choice test designed to measure the specific instructional objectives of the unit.
- c. Oral responses to questions asked in class of each student over the course of the unit.
- d. Daily grades designed to indicate the quality of in-class participation during regular instruction.

Figure 4.5

Item 22 of the Original TALQ

22. A teacher gave three tests during a grading period and she wants to weight them all equally when assigning grades. The goal of the grading program is to rank order students on achievement. In order to achieve this goal, which of the following should be closest to equal?

- a. Number of items.
- b. Number of students taking each test.
- c. Average scores.
- d. Variation (range) of scores.

Figure 4.6

Item 24 of the Original TALQ

24. Which of the following grading practices results in a grade that **least** reflects students' achievement?
- a. Mr. Jones requires students to turn in homework; however, he only grades the odd numbered items.
 - b. Mrs. Brown uses weekly quizzes and three major examinations to assign final grades in her class.
 - c. Ms. Smith permits students to redo their assignments several times if they need more opportunities to meet her standards for grades.
 - d. Miss Engle deducts 5 points from a student's test grade for disruptive behavior.

In relation to knowing the differences between norm- and criterion- or standards-referenced information, results show that assessment features and types are best assessed by items 18, 27, and 29. Participants were also firmly of the view that Items 32 and 34 have a significant role to play in assessing the capacity of teachers to understand the ethical aspects that are relevant to some elaborations in the refined assessment literacy framework as well as KSA national professional standards. Chapter 6 will provide much greater detail on these items and their related results.

Table 4.4 shows the nine items in TALQ that most participants in the focus group believed added no—or at least very little—value to the construct represented by the NPTS. They suggested that all nine of these items be deleted because they did not align with the refined assessment literacy framework.

Table 4.4*Items to be Deleted in Response to the Refined Assessment Literacy Framework*

Statement/Reason	Item number in original TALQ
Not contextual	14, 28, 31, 16
No direct link with the refined assessment literacy framework	19, 33, 26, 30
No need for more items to measure assessment ethical aspects	35

Table 4.4 shows the items that were identified as being problematic in TALQ for assessing the standards associated with the NPST and refined assessment literacy framework. For example, Items 19, 33, 26 and 30 were identified for elimination because there was consensus among the participants that they did not align with the standards and associated elaborations in the refined assessment literacy framework. As shown earlier in this section, Item 19 was considered by 30 of the 35 participants to be assessing their knowledge about utilising standardised tests data, rather than classroom assessment practices and the standards associated with the refined assessment literacy framework. Similarly, there was consensus amongst participants that Item 26 (Figure 4.7) might assess teachers' skills in how to communicate assessment, but there is no direct link to the refined assessment literacy framework. However, participants also felt that the skill being assessed in Item 30 (Figure 4.8) is rarely used in the Saudi context. As a result, participants called for the deletion of these items.

Figure 4.7

Item 26 of the Original TALQ

26. In a routine conference with Mary's parents, Mrs. Estes observed that Mary's scores on the state assessment program's quantitative reasoning tests indicate Mary is performing better in mathematics concepts than in mathematics computation. This probably means that:
- a. Mary's score on the computation test was below average.
 - b. Mary is an excellent student in mathematics concepts.
 - c. The percentile bands for the mathematics concepts and computation tests do not overlap.
 - d. The mathematics concepts test is a more valid measure of Mary's quantitative reasoning ability.

Figure 4.8

Item 30 of the Original TALQ

30. John scored at the 60th percentile on a mathematics concepts test and scored at the 57th percentile on a test of reading comprehension. If the percentile bands for each test are five percentile ranks wide, what should John's teacher do in light of these test results?
- a. Ignore this difference.
 - b. Provide John with individual help in reading.
 - c. Motivate John to read more extensively outside of school.
 - d. Provide enrichment experiences for John in mathematics, his better performance area.

Participants unanimously agreed that Item 33's Options (Figure 4.9) were too vague (i.e., the expressions 'every cloud has its silver lining', 'into each life some rain must fall', 'the rich get richer; the poor get poorer' and 'a bird in the hand is worth two in the bush') and were loaded against the KSA context. As a result, they suggested that the item be deleted.

Figure 4.9

Item 33 of the Original TALQ

33. A state uses its statewide testing program as a basis for distributing resources to school systems. To establish an equitable distribution plan, the criterion set by the State Board of Education provides additional resources to every school system with student achievement test scores above the state average. Which cliché best describes the likely outcome of this regulation?
- a. Every cloud has its silver lining.
 - b. Into each life some rain must fall.
 - c. The rich get richer and the poor get poorer.
 - d. A bird in the hand is worth two in the bush.

Similarly, participants thought that items 14, 28, 31 and 16 were contextually inappropriate for KSA teachers. For example, most participants were of the view that it would be challenging for Practitioner and Advanced teachers to interpret the scores from a standardised test as referenced in Item 14 (Figure 4.10).

Figure 4.10

Item 14 of the Original TALQ

14. Frank, a beginning fifth grader, received a G. E. (grade equivalent score) of 8.0 on the Reading Comprehension subtest of a standardized test. This score should be interpreted to mean that Frank:
- a. can read and understand 8th grade reading level material.
 - b. scored as well as a typical beginning 8th grader scored on this test.
 - c. is performing in Reading Comprehension at the 8th grade level.
 - d. will probably reach maximum performance in Reading Comprehension at the beginning of the 8th grade.

The main reason for deleting Item 16 (Figure 4.11) was related to the length and the complexity of the item, even though participants recognised the importance of measuring diagnostic assessment skills in the refined assessment literacy framework. In lieu of the suggested deletion, participants proposed writing a new item that aimed to measure the knowledge and skills of using diagnostic assessment and how teachers can use it to check their students' prior knowledge before designing a new lesson.

Figure 4.11

Item 16 of the Original TALQ

16. Ms. Camp is starting a new semester with a factoring unit in her Algebra I class. Before beginning the unit, she gives her students a test on the commutative, associative, and distributive properties of addition and multiplication. Which of the following is the most likely reason she gives this test to her students?
- a. The principal needs to report the results of this assessment to the state testing director.
 - b. Ms. Camp wants to give the students practice in taking tests early in the semester.
 - c. Ms. Camp wants to check for prerequisite knowledge in her students before she begins the unit on factoring.
 - d. Ms. Camp wants to measure growth in student achievement of these concepts, and scores on this test will serve as the students' knowledge baseline.

Finally, participants also proposed deleting Item 35 (Figure 4.12). They felt that this item should be deleted because there were already two other items that related to the ethical aspects of assessment. According to the participants, retaining an item which seemed to relate to the same aspect lengthened the test and yielded little relevant information.

Figure 4.12

Item 35 of the Original TALQ

35. Mrs. Overton was concerned that her students would not do well on the State Assessment Program to be administered in the Spring. She got a copy of the standardized test form that was going to be used. She did each of the following activities to help increase scores. Which activity was unethical?
- a. Instructed students in strategies on taking multiple choice tests, including how to use answer sheets.
 - b. Gave students the items from an alternate form of the test.
 - c. Planned instruction to focus on the concepts covered in the test.
 - d. None of these actions are unethical.

The next section describes the process used to modify or replace those items that were removed from the TALQ.

4.7 Building the Modified TALQ-MTALQ

Before discussing this process, it is important to note that the sixth standard in the TALQ, that is ‘Recognising unethical practice’, was considered by all focus groups as an essential standard in the TALQ, but they were firmly of the view that it would be best assessed in each standard separately in the MTALQ. This meant that items addressing unethical practice should be embedded across the whole MTALQ.

Table 4.5 shows how the items from the TALQ (unmodified or deleted) and the new items written for the MTALQ were allocated to the standards.

Table 4.5*The Number of Items Assigned to the Five Standards in the NPST*

Standard in the MTALQ	NPST Standard	Items <i>Not</i> Modified from TALQ	Items Deleted from TALQ	New Items	Total Items
Standard 1 (Choosing assessment methods)	Standard 10.1 Preparing evaluation tools	5	0	1	6
Standard 2 (Developing assessment methods)	Standard 10.1 Preparing evaluation tools	5	0	1	6
Standard 3 (Administering, scoring and interpreting assessment results)	Standard 10.2 Applying evaluation	4	1	1	5
Standard 4 (Using assessment results for decision- making)	Standard 10.4 Employing evaluation results	3	2	2	5
Standard 5 (Using assessment in grading)	Standard 10.2 Applying evaluation	5	0	1	6
Standard 6 (Communicating assessment results)	Standard 10.5 Preparing Evaluation Reports	2	3	2	4
Standard 7 (Modern assessment, or relevant knowledge and skills for assessment)	Standard 10.3 Engaging students in evaluation processes	0	0	4	4

This section outlines the process for modification and presents the process used for writing and reviewing items. It has been done considering all the ideas and suggestions from the earlier stages. After removing the items identified as inappropriate by the focus group from the TALQ, the resulting assessment is the modified TALQ or MTALQ. In the following presentations, the items are represented by their item number as it appears in the new MTALQ.

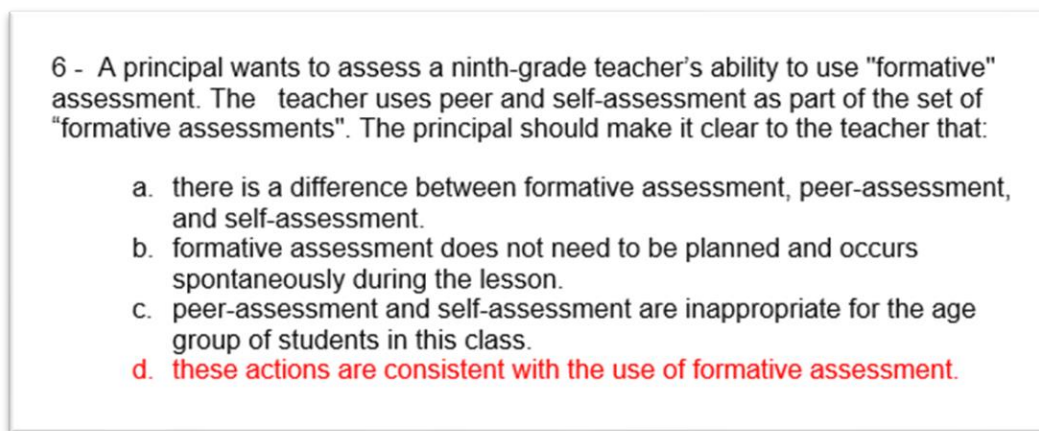
Any replacement items were constructed based on the suggestions of the participants focus group; reviewed by the supervisors; and, accepted by the focus groups as part of Stage 3, Step 2.

4.7.1 TALQ Standard 1: Choosing Assessment Methods (Standard 10.1 in NPST)

The TALQ included five items assessing Standard 1 (choosing assessment methods). A further item (Item 6) was written and reviewed. It assesses Standard 10.1 in NPST related to teachers' knowledge of formative assessment by asking whether peer- and self-assessment are formative assessment strategies. Item 6 is shown in Figure 4.13.

Figure 4.13

Item 6 of the MTALQ



4.7.2 TALQ Standard 2: Developing Assessment Methods (Standard 10.1 in NPST)

There were five items for this standard in the TALQ. The focus group suggested adding another item to assess Standard 10.1 in NPST. The item assesses teachers' basic knowledge about the types of assessment. Figure 4.14 shows this item as item 12.

Figure 4.14

Item 12 of the MTALQ

12. When a teacher administers an assessment at the end of an educational unit or semester for reporting home to parents, this test is considered to be:

- a. Assessment for learning.
- b. **Assessment of learning.**
- c. Formative assessment.
- d. None of the above.

4.7.3 TALQ Standard 4: Using Assessment Results for Decision-making (Standard 10.4 in NPST)

Two items were added to the TALQ to assess Standard 10.4 in NPST more effectively, following the suggestions from the focus groups. Figure 4.15 shows the first of the additional items (Item 18 in the MTALQ). It is designed to assess teachers' knowledge and skills regarding the use of assessment results for decision-making after using diagnostic assessment tools.

Figure 4.15

Item 18 of the MTALQ

18. Before planning a lesson, a teacher wants to know what knowledge and skills the students already have in the unit of work. Which of the following would be the best way to find out?

- a. Do nothing as the teacher has already given instruction on the pre-requisite skills.
- b. Review the past lessons homework to understand students depth of learning.
- c. Conduct structured interviews with each student.
- d. **Administer a short quiz.**

The second item included to assess Standard 10.4 is Item 21 (Figure 4.16). It assesses data literacy and teachers' ability to use assessment results.

Figure 4.16

Item 21 of the MTALQ

21. The best ways to find out the variation of students' grades is to calculate the

- a. mean.
- b. median.
- c. total score for each student.
- d. **standard deviation.**

4.7.4 TALQ Standard 5: Using Assessment in Grading (Standard 10.2 in NPST).

Standard Five consisted of five items in the original TALQ (i.e., No items were removed from the original TALQ for this standard). Following the suggestions of the focus group, one more item (Item 28, shown in Figure 4.17) was added to assess the extent to which teachers can

distinguish between quantitative and qualitative data when using assessment in the process of grading.

Figure 4.17

Item 28 of the MTALQ

28. What kind of data can be used to carry out statistical analysis? the MOST correct answer is:
- a. Qualitative and quantitative.
 - b. Quantitative.**
 - c. Qualitative.
 - d. None of the above.

4.7.5 TALQ Standard 6: Communicating Assessment Results (Standard 10.5 in NPST).

There were five items for this standard in the TALQ to assess Standard 10.5 in NPST and related skills in communicating assessment results. The focus group had removed three items from the TALQ. One of the items (Item number 26) was removed because it was not aligned with the refined assessment literacy framework. A second item (Item number 28) was removed because it was contextually inappropriate for the KSA context, while the third item (Item number 30) was removed because it used specific terms (e.g., ‘percentile ranks’) as mentioned in the previous stage. Figure 4.18 shows Item 28, which was deleted from the original TALQ because it was considered contextually inappropriate. One item (Item number 30 in the MTALQ was written to replace the deleted items aligned to this standard.

Figure 4.18

Deleted Item No. 28 of the Original TALQ

28. The following standardized test data are reported for John.

<u>Subject</u>	<u>Stanine Score</u>
Vocabulary	7
Mathematics Computation	7
Social Studies	7

Which of the following is a valid interpretation of this score report?

- a. John answered correctly the same number of items on each of the three tests.
- b. John's test scores are equivalent to a typical seventh grader's test performance.
- c. John had the same percentile rank on the three tests.
- d. John scored above average on each of the three tests.

Figure 4.19 shows the item that was added to replace the three items that were removed by the original TALQ for this standard. The new item was added to assess the extent to which the teachers know how to communicate assessment results from national and international tests to KSA assessment stakeholders. The item is designed to test teachers' knowledge of the purpose of PISA. While not every student takes the PISA test, it is high profile in the press and has a level of political gravitas, which could account for why parents and the educational community in general refer to its results. It is essential that teachers know the purpose of using such international measures, such as PISA, and be able to communicate this purpose to their students' parents and community members to try and make sure that the results are correctly interpreted.

Figure 4.19

Item 30 of the MTALQ

30. The purpose of administering PISA in Saudi Arabia is:
- a. for the teacher to get a more complete image of the students' achievement.
 - b. to collect information on the efficacy of educational policy.
 - c. to generate competition for acclaim internationally.
 - d. to diagnose teaching and learning deficiencies.

4.7.6 TALQ Standard 7: Modern Assessment, or Relevant Knowledge and Skills for Assessment (Standard 10.3 in NPST).

Following the focus group discussions, some major changes were made to the way that the NPST standards can be fairly assessed. The skills and knowledge included in the refined assessment literacy framework can be better assessed (i.e., in a more valid manner) by adding new items and deleting others that the focus group considered not to be appropriate for the KSA context, or well aligned to the NPST, by the focus groups. All four items in Standard 7, shown below, have been added to yield more insights in terms of teachers' skills in applying modern assessment theory.

Standard 7: modern assessment (or relevant knowledge and skills for assessment).

33. A teacher uses rubrics to assess students' performance in developing and delivering a speech in front of their classmates.

- a. There is nothing wrong with this approach.*

- b. *The teacher should have used direct observation rather than recording in order to fully observe the student.*
- c. *The student can assess himself by asking his classmates about his speaking competence and informing the teacher.*
- d. *Rubrics are only used in summative assessment.*

34. A seventh-grade teacher develops students' assessment skills to support self-regulation. We can clearly state that:

- a. *the student's self-regulation is concerned with emotional aspects of learning only.*
- b. *developing the student's assessment skills increases self-regulation.*
- c. *developing assessment skills is for the teacher, not for the student.*
- d. *self-regulation of the student contradicts his independence in learning.*

35. After the end of the unit on the history of ancient civilisations, the history teacher assigns each student a project that simulates these civilisations. The teacher should:

- a. *assign all students the same project.*
- b. *realise that project work is not an assessment tool.*
- c. *ask each student to design a project while adhering to unified criteria for grading.*
- d. *none of the above.*

36. A teacher allows students to view and participate in building assessment criteria. Which of the following is MOST correct?

- a. *Only the teacher should be aware of the assessment criteria, not the students.*
- b. *The teacher and parents should be aware of and participate in building criteria.*

- c. *The students' participation in building the assessment criteria could improve their learning outcomes.*
- d. *Assessment criteria are NOT usually used in the classroom.*

4.8 Finalising the MTALQ

This process aimed to collect validity evidence before administrating the MTALQ to the targeted teachers in the study sample. The school principals were contacted to encourage the teachers to volunteer to participate in the final focus group discussion, However, certain of the participants indicated their preference to review the MTALQ individually and give their respective feedback by email. This was not viewed this as a shortcoming as it not only suited the preferences of most of the participants and their larger context, but it also seems to have given these participants more opportunity to devote undivided attention and provide more in-depth feedback.

After making all the recommended and necessary changes to the TALQ, consent was sought to email the participants the MTALQ (N = 16). Five of the participants responded via email, confirming that the questionnaire was ready for administration. One of the participants gave his feedback by email, clarifying that two of the sentences still needed some editing (referring to the mechanics of writing like grammatical errors). Meanwhile, ten participants wanted to participate in a focus group discussion as they had questions or concerns about some of the technical aspects of the items in the MTALQ. In response to this, the researcher organised an additional focus group discussion with these participants on Zoom to discuss the items in question.

When the discussion started, participants asked about the 'similarity of distractors'. They felt that in some items in the MTALQ, the distractors were very similar. They asked whether this

was intended. In response to this query, the researcher answered: ‘Yes, because in some items a number of distractors were correct, but one BEST (or MOST correct) answer required.’ It was expected that such items would more effectively facilitate discrimination of those respondents who had a general idea of the knowledge or skill being assessed and those that had a deeper understanding of the knowledge or skill being assessed.

To encourage more opinions to be shared it was made clear that the aim of developing MTALQ was to measure the required skills by contextualising the skills and knowledge, to ensure that answers were purposefully chosen, while also preventing overlap between assessment terms in some items. Reacting to the researcher’s explanation, one of the teachers said: ‘If that is the case, then in my opinion, the questionnaire is now publishable.’

As mentioned earlier, the aim of the focus group currently under discussion was to seek more evidence about the validity of the MTALQ from a subject matter point of view, a technical point of view and an editorial point of view.

The participants emphasised that the new items should not be located randomly in the matched Standard, but the new location in the questionnaire should consider the ‘elaboration’ disposition in the refined assessment literacy framework. This refined assessment literacy framework is designed to define a continuum from basic skills to more advanced knowledge in assessment. Consider Item 6 of the MTALQ (Figure 4.20).

Figure 4.20

Item 6 of the MTALQ

6 - A principal wants to assess a ninth-grade teacher's ability to use "formative" assessment. The teacher uses peer and self-assessment as part of the set of "formative assessments". The principal should make it clear to the teacher that:

- a. there is a difference between formative assessment, peer-assessment, and self-assessment.
- b. formative assessment does not need to be planned and occurs spontaneously during the lesson.
- c. peer-assessment and self-assessment are inappropriate for the age group of students in this class.
- d. these actions are consistent with the use of formative assessment.

The participants in the latest focus group moved Item 6 to be position of last under Standard 1 in MTALQ (choosing assessment methods) as the knowledge related to formative assessment (i.e., self- and peer-assessment) was given in the refined assessment literacy framework after the purpose of assessment, reliability, assessment strategy, feedback and validity. Furthermore, the participants considered self- and peer-assessment to be more complex in the KSA context.

In addition, participants suggested changing the place names in the Arabic version of the questionnaire to town, city, county, region and province, so that these were aligned to the KSA context without prejudice to the meaning. More importantly, the participants devoted more effort to Standard 7 in MTALQ: modern assessment (or relevant knowledge and skills for assessment) as this was a new standard in the MTALQ that encompassed four items. For example, Item 33 addressed Standard 7 (Figure 4.21).

Figure 4.21

Item 33 of the MTALQ

33. A teacher uses rubrics to assess students' performance in developing and delivering a speech in front of their classmates.
- a. There is nothing wrong with this approach.
 - b. The teacher should have used direct observation rather than recording in order to fully observe the student.
 - c. The student can assess himself by asking his classmates about his speaking competence and informing the teacher.
 - d. Rubrics are only used in summative assessment.

The participants asserted that Item 33 was a valid item to assess teachers' ability to collect, record and report students' achievement using rubrics as acceptable tools to achieve these aims; however, participants argued that Option B might be too easy of a distractor for the targeted sample to ignore, because using rubrics is an essential skill for teachers. Results showed that 27.7% of teachers chose Option B, which might reflect that Saudi teachers do not use rubrics consistently in the classroom.

Consider item 35 (Figure 4.22):

Figure 4.22

Item 35 of the MTALQ

35. After the end of the unit on the history of ancient civilizations, the history teacher assigns each student a project that simulates these civilizations. The teacher should:
- a. assign all students the same project.
 - b. realise that project work is not an assessment tool.
 - c. ask each student to design a project while adhering to unified criteria for grading.
 - d. None of the above

Participants suggested changing one word in this item, from ‘describe’ to ‘simulate’. This was because there is a slight difference in meaning between the two words in Arabic, which could affect the meaning of the item and thus the result. The suggested change was made.

Item 36 (Figure 4.23) was then considered by the focus group.

Figure 4.23

Item 36 of the MTALQ

36. A teacher allows students to view and participate in building assessment criteria. Which of the following is MOST correct?
- a. Only the teacher should be aware of the assessment criteria, not the students.
 - b. The teacher and parents should be aware of and participate in building criteria.
 - c. The students’ participation in building the assessment criteria could improve their learning outcomes.
 - d. Assessment criteria are NOT usually used in the classroom.

The participants argued that allowing students to participate in building assessment criteria is a complex issue in the KSA context. This was consistent with the points of view of participants in Stage 1, Elaboration 51 (i.e., the validation process to refine the assessment literacy framework). The participants in the focus group session recommended that this item be replaced. However, this skill and knowledge being assessed by the item were considered desirable in the survey. Hence, the researcher consulted the assessment experts (supervisors) to gain their opinion. According to the experts, part of assessment for learning is allowing students to create criteria or share them with other students. Hence, the item's results might shed insights into the extent to which teachers in KSA understand modern assessment or assessment for learning. As such, the item was retained without change.

By the end of this stage of discussions with participants, the MTALQ was ready to be administered to the teachers targeted in the next and final stage (i.e., Stage 3, Step 3 as described in Chapter 3).

4.9 Chapter Summary

This chapter has discussed how the results (obtained from Likert Survey and focus group discussion) were used in generating validity evidence that the assessment literacy framework and the development of the MTALQ, is strongly aligned with the NPST standards and, as such, provide a base for measuring attainment of those standards in the KSA educational context. NPST did not provide actionable details regarding the skills and knowledges needed for teachers to master the NPST standards. As such, the development and refinement of the assessment literacy framework (with its elaborations) and its strong alignment to the NPST standards will play a vital role in providing in Saudi a tool that can be used to show attainment of these

standards over time, possibly leading to enhanced professional learning programs with better teacher preparation in assessment literacy and ultimately improved student performance.

The results obtained from examining primary, intermediate, and secondary teachers' responses of the focus groups to the survey, led to the conclusion that most teachers rated elaborations as desirable and essential, which is consistent with the procedures in developing the assessment literacy framework focusing on the essential skills in assessment in the context of Saudi Arabia.

However, the participants argued that there were some problems in the elaborations, which were related to the use of statistical terms in the assessments as well as the problems raised with the self-peer-assessment and engaging the students in the evaluation which are considered by the participants are obvious obstacles the teachers face in KSA. Other issues such as providing opportunity for student to develop the success criteria or standards were also discussed—both of which led to the refinement of the assessment literacy framework.

Furthermore, the participants in the focus groups argued that most of the elaborations did not require amendments or adjusting. Aligning items of the original TALQ with the elaborations that comprised the refined assessment literacy framework resulted in the deletion of some items and the suggestion to add new ones. Even though the participants agreed that some items in the original TALQ might be helpful to assess some knowledge and skills assessment in general, but they argued that it did not align with KSA context and did not assess any skills in the refined assessment literacy framework.

These focus group discussions also led to the modification of some items in the original TALQ to be consistent and applicable in the context of KSA. The discussion was valuable for the writing of new items that matched NPST and assess the elaborations in the refined

assessment literacy framework. Moreover, the elaborations in the refined assessment literacy framework paved the way for the development of the MTALQ in the context of KSA. The final stage to collect more validity evidence about the instrument, MTALQ, was necessary to ensure that the instrument is ready to be administered as a target sample. In the following chapter, the results from the questionnaire will be discussed and findings will be summarised to show the validity of the instrument.

Chapter 5: Results from Analysis of the Modified Teacher Assessment Literacy Questionnaire (MTALQ)

5.1 Introduction

This Chapter presents the quantitative results obtained from the analysis of data collected via the MTALQ that contained questions relating to the standards of the NPST and consisted of varying numbers of items assessing each of the standards. Classical Measurement Theory, using item response analyses, were conducted. The most salient statistic described in the process is the percentage of correct responses to each question, indicating the difficulty of the items in the questionnaire. In instances in which most teachers responded to a particular item correctly, then that item is easy. On the other hand, if only a small percentage of the teachers responded correctly, that item is considered difficult. The results are described in this chapter, beginning with an overview of respondents' demographics and profiles.

5.2 Demographics of the Respondents to the MTALQ

Demographic and career profiles of the respondents (teachers targeted in the sample) are presented in Table 5.1. The survey was administered to 480 teachers. Most (76%) survey respondents were male, with females representing only 24% of the survey respondents. Most (42.3%) of the respondents taught at primary level, then came the secondary level (33%) and the remainder were from the intermediate school (24.7%). About 96% of the respondents possessed at least a bachelor's degree. Practitioner teachers made up 82.5% of the respondents, with the rest (i.e., 17.5%) being teachers at the Advanced career level. Approximately 80% of the respondents had more than 10 years of teaching experience and about 14% had 6–10 years of

teaching experience. Despite this experience only 76% of the respondents considered themselves to be trained teachers.

Table 5.1

Demographic and Career Profiles of KSA Teacher Respondents

Item		Frequency	Percentage
Gender	Male	221	75.9
	Female	70	24.1
School level taught	Primary	123	42.3
	Intermediate	72	24.7
	Secondary	96	33.0
Academic qualifications	Diploma	10	3.4
	Bachelor	215	73.9
	Master	56	19.2
	Doctoral	10	3.4
Career level	Practising teacher	240	82.5
	Advanced teacher	51	17.5
Years of experience	1–5 Years	16	5.5
	6–10 Years	41	14.1
	More than 10 Years	232	79.7
	Do not remember	2	0.7
Training	Yes	220	75.6
	No	71	24.4

5.3 The MTALQ Results

Table 5.2 shows that there were 36 items in the MTALQ. Standards 1, 2 and 5 in the MTALQ were assessed by six items each; Standards 3, 4 in the MTALQ were assessed by five items each and Standard 6, 7 in the MTALQ were assessed by four items each. Responses were received from 291 participants. The limited sample of responses was attributable to circumstances created by the COVID-19 pandemic restrictions. The mean score across the MTALQ was 17.31 (out of a possible 36, which represents a mean of 46.8%). The mean scores for Standards 1 to 7, respectively, were 2.51 out of 6 (42%), 2.54 out of 6 (42.3%), 2.87 out of 5 (57.4%), 2.60 out of 5 (52%), 2.65 out of 6 (44.2%), 1.29 out of 4 (32.2%) and 2.85 out of 4 (57.4%). Cronbach's alpha for the entire scale (all items considered together) was 0.562. which is considered low when benchmarked against one rule of thumb which suggests that an Alpha reliability of 0.7 would be acceptable (Taber, 2018). What this means is that much of the variation in respondent responses was internally inconsistent. This might be more to do with the survey being very difficult for most of the respondents (more discussion about reliability and validity of the MTALQ will be discussed in the next chapter).

It can also be seen from Table 5.2 that there were 13 items that had relatively low mean corrected item-to-total score correlations, with those less than or equal to 0.01 explicitly identified. These items will be discussed further in the next section. They were clearly not working for some reason or another, and this contributed to the relatively low reliability of the MTALQ.

Table 5.2*Summarised Highlights of Results*

	All items	MTALQ Standard 1: (Standard 10.1 in NPST)	MTALQ Standard 2: (Standard 10.1in NPST)	MTALQ Standard 3: (Standard 10.2in NPST)	MTALQ Standard 4: (Standard 10.4 in NPST)	MTALQ Standard 5: (Standard 10.2 in NPST)	MTALQ Standard 6: (Standard 10.5 in NPST)	MTALQ Standard 7: (Standard 10.3 in NPST)
Number of items	36	6	6	5	5	6	4	4
Number of participants	291	291	291	291	291	291	291	291
Mean	17.31	2.51	2.54	2.87	2.60	2.65	1.29	2.85
Standard deviation	4.07	1.02	1.05	1.23	1.14	1.14	0.87	1.33
Coefficient Alpha (reliability)	.562							
Number of items with item-to-total-score correlation less than or equal to 0.01	13	3	2	1	1	3	3	0

5.4 Summary of Item Statistics by Standards

Table 5.3 shows the individual item statistics for all the items comprising the MTALQ. For example, it shows that the easiest item (the one with the highest percent correct score) was Item 1, which 92.8% of the respondents answered correctly. The most challenging item was Item 8, which only 8.20% of the respondents answered correctly. Table 5.3 also shows that 13 out of the total 36 items had a mean corrected item-to-total score correlation less than or equal to 0.01. It is interesting to note that for Standard 7, all the items had mean corrected item-to-total score correlations above 0.01, which was a new standard in the MTALQ introduced to measure Standard 10.3 in NPST (see Chapter 4 for more details regarding Standard 7 in the MTALQ). For Standards 3 and 4 in the MTALQ, all the items apart from 1 in each standard (Item 15 in Standard 3 and Item 20 in Standard 4) had mean corrected item-to-total score correlations above 0.01. Standards 1, 2 and 5 had 3, 4 and 3 items respectively that assessed the standards and yielded a corrected item-to-total score correlation above 0.01. In Standard 6, only one out of the four items had a corrected item-to-total score correlation above 0.01. Items with very low corrected item-to-total score correlation did not discriminate among the respondents. This could be a function of the items being too difficult or it could indicate that the scoring key may have been incorrectly applied. A check of the scoring key indicated that the same was correct.

Table 5.3*Summary of Statistical Results by Item*

The MTALQ Standard	Item	NPST Standard	Percentage of correct responses	Mean	Standard deviation	Corrected item-to- total score correlation	Cronbach's alpha if item deleted
1	1	Standard 10.1	92.8%	0.928	0.259	0.057	0.554
	2		52.9%	0.529	0.500	0.033	0.540
	3		28.9%	0.289	0.454	-0.091	0.570
	4		14.8%	0.148	0.355	-0.055	0.578
	5		34.7%	0.347	0.477	-0.010	0.572
	6		26.8%	0.268	0.444	0.028	0.550
2	7	Standard 10.2	74.9%	0.749	0.434	0.046	0.561
	8		8.2%	0.082	0.276	0.001	0.571
	9		43.6%	0.436	0.497	0.041	0.561
	10		77.3%	0.773	0.419	0.047	0.545
	11		19.2%	0.192	0.395	0.062	0.565
	12		30.2%	0.302	0.460	-0.081	0.575
3	13	Standard 10.2	64.6%	0.646	0.479	0.144	0.552
	14		74.2%	0.742	0.438	0.189	0.528
	15		40.9%	0.409	0.492	-0.007	0.580
	16		53.6%	0.536	0.500	0.147	0.536
	17		54%	0.540	0.499	0.236	0.534
4	18	Standard 10.4	78.4%	0.784	0.413	0.163	0.536
	19		51.9%	0.519	0.501	0.219	0.539
	20		18.9%	0.189	0.392	-0.092	0.578
	21		66.7%	0.667	0.472	0.106	0.538
	22		44%	0.440	0.497	0.144	0.546
5	23	Standard 10.2	43%	0.430	0.496	-0.062	0.567
	24		14.8%	0.148	0.355	0.013	0.564
	25		69.1%	0.691	0.463	0.031	0.546
	26		44%	0.440	0.497	0.108	0.545
	27		67.4%	0.674	0.470	0.125	0.552
	28		26.8%	0.268	0.444	-0.105	0.567
6	29	Standard 10.5	26.8%	0.268	0.444	-0.075	0.580
	30		27.1%	0.271	0.445	-0.054	0.572
	31		56.7%	0.567	0.496	-0.020	0.545
	32		18.2%	0.182	0.387	0.063	0.566
7	33	Standard 10.3	40.5%	0.405	0.492	0.078	0.552
	34		72.9%	0.729	0.445	0.294	0.543
	35		68.4%	0.684	0.466	0.289	0.547
	36		67%	0.670	0.471	0.346	0.535

The last column in Table 5.3 shows the relationship between the mean corrected item-to-total score correlation and the reliability of the MTALQ. For example, if Item 28, with the corrected item-to-total score correlation of -0.105 , is removed from the MTALQ, the reliability increases to a value of 0.567 . While removing these items from the MTALQ could improve the reliability of the assessment, it could also detract from the validity of the MTALQ. The validation process conducted in the previous chapter would suggest that the items are assessing what they are supposed to be assessing. Removing them on the grounds of this statistic could lead to a diminution in the overall validity of the MTALQ.

There needs to be further assessment of why these items did not function as anticipated. It appears that the items with the smallest corrected item-to-total score correlations were the ones that were too difficult for the respondents. This could imply that the respondents were guessing. If the teachers thought these were important items to represent the standards, but don't know the answers then these could indicate what to focus upon for future professional development activities to improve knowledge and skills about assessment literacy. Another area for future research could involve more forensic analysis of the functioning of the more difficult and less discriminating items using a 'Think Aloud Protocol' (Boren & Ramey, 2000).

The next sections consider each of the seven standards in the MTALQ and examine how the items assessing the NPST standards have functioned.

5.4.1 MTALQ Standard 1: Choosing Assessment Methods (Standard 10.1 in NPST).

This standard relates to teachers' ability to select valid and reliable methods to assess student learning and monitor their progress. The teachers who responded to the MTALQ would appear to have low assessment literacy when it comes to selecting assessment methods. Due to this, a deeper examination was required to identify patterns of responses that would indicate why

this was the case. This may then lead to improvements in the MTALQ. These patterns could also provide robust, comprehensive and meaningful information that may be utilised to develop appropriate professional learning programs that are driven by evidence of teachers' current level of assessment literacy and evidence from the literature.

Table 5.3 shows that there were six items in Standard 1 to measure Standard 10.1 in the NPST. Item 1 was the easiest item in Standard 1 and it was answered correctly by 92.8% of the respondents. The most difficult item in this standard was Item 4, which only 14.8% answered correctly. None of the six items discriminated at a relatively high level, most likely because most of these items are either too difficult or too easy for the respondents. Figure 5.1 shows Item 1. The correct answer is C. As shown in Table 5.3, 92.8% of the participants answered this item correctly.

Figure 5.1

Item 1 of the MTALQ

What is the most important consideration in choosing a method for assessing student achievement?

- a. Ease of scoring the assessment.
- b. Ease of preparing the method of assessment.
- c. Accuracy of assessing attainment of instructional objectives.
- d. Acceptance by the school administration.

The responses to item 1 show that 92.8 percent of respondents chose the correct answer, ‘Accuracy of assessing attainment of instructional objective’, showing they can demonstrate their knowledge concerning one critical aspect of assessment literacy.

The next item considered in this Standard is Item 4.

Figure 5.2

Item 4 of the MTALQ

4. What is the most effective use a teacher can make of an assessment strategy that requires students to show their work , e.g., the way they arrived at a solution to a problem or the logic used to arrive at a conclusion?
 - a. Assigning grades for a unit of instruction on problem solving.
 - b. Providing instructional feedback to individual students.
 - c. Motivating students to attempt innovative ways to solve problems.
 - d. None of the above.

Figure 5.2 shows Item 4, which was the most difficult item assessing Standard 1 in the MTALQ (measuring Standard 10.1 in NPST). Only 14.8% of the respondents answered this item correctly. The percentage of respondents choosing each of the options was as follows: 5.88% chose Option A, 14.8% chose Option B, 62.6% chose Option C and 17.3% chose Option D. Most respondents selected an incorrect response, Option C. This item also had a negative corrected item-to-total score correlation, which meant more of the weaker respondents chose Option B than the stronger participants. Therefore, if the item was removed from the questionnaire, Cronbach’s alpha would increase from 0.562 to 0.578. The pattern of responses to this item suggested that either respondents tended to guess their selected answer or that they

were used to using assessment methods as a motivator for students. However, the key point is that the focus groups identified this item as assessing essential knowledge for the teachers at the Practitioner and Advanced levels. These results provide some evidence regarding the most effective assessment strategies, which include feedback and motivation.

These results could also be explained by the conceptual overlap between the concepts of feedback and motivation from these respondents' perspective. For instance, in the refined assessment literacy framework, teachers should know and understand the features of feedback that make it an effective tool to assist student learning. Such feedback should be clear, purposeful, meaningful, timely and compatible. For instance, some examples of feedback that illustrate the most effective delivery modes should be offered. Effective feedback should be:

- delivered in an appropriate setting;
- focused on performance and not the individual;
- delivered using neutral, non-judgemental language;
- focused on highlighting areas for improvement and suggesting strategies to improve; and
- aimed to empower and encourage learning.

On the other hand, unlike Practitioner teachers, Advanced teachers could be equipped with the knowledge to describe ways to evaluate whether the feedback given has been effectively understood by their students. Some of the ways that this can be done include interacting with students to ensure there is a shared understanding of terminology and language, and reflecting on how the feedback that has been given may assist them in improving their learning. The importance of this is exemplified by Hattie and Timperley (2007), who argue that effective feedback might use descriptive comments in students' work to make the link between learning outcomes and targeted standards clear.

Furthermore, to illustrate the difference between the two concepts in the assessment process, teachers could choose assessment methods that can be used as a motivational tool, as research suggests that both external and internal motivation improve student performance assessment (Brookhart & Durkin, 2003). For example, students sometime tend to perform well in performance assessment rather than tests because the former allows them to show their work to classmates (Brookhart & Durkin, 2003). Thus, Advanced teachers in the current refined assessment literacy framework should use and develop assessment approaches to engage prior students' knowledge and skills in lesson plans that consider students' prior characteristics. Doing so raises students' self-efficacy in relation to the classroom environment and motivation, as found by Brookhart et al. (2006). For instance, Advanced teachers could use an authentic assessment approach to create a new learning environment and enhance student motivation towards learning and evaluate these approaches to support Practitioner teachers.

5.4.2 MTALQ Standard 2: Developing Assessment Methods (Standard 10.1 in NPST).

This standard measures teachers' skills and knowledge in relation to developing assessment methods to measure Standard 10.1 in NPST. According to OECD (2020), such knowledge—which is vital to teaching and learning in KSA—is lacking. Having been used to a top-down policy approach, which has characterised the KSA education system, teachers are not keen to stray from the well-trodden path and create innovative assessments that meet the individual needs of their students more specifically. What could be attributed to this as well is that such endeavours entail customising assessment methods, a skill that KSA teachers may also be lacking. This finding is echoed by the previous research done by Volante and Fazio (2007) that found that some aspects of teachers' assessment literacy-such as being able to develop assessment tools, administer and score assessment tasks-need more development.

Table 5.3 shows that there were six items assessing Standard 2 in the MTALQ. Of these, the easiest was Item 10 with 77.3% correct responses, closely followed by Item 7 with 74.9% producing the correct response.

Item 7 (Figure 5.3) measures teachers' ability in relation to being able to establish the validity of the scores from a classroom assessment strategy to use for assigning grades on a class unit. Most respondents—74.9% of the respondents—selected the correct answer, which read 'match an outline of the instructional content to the content of the assessment strategy.' This indicates that most teacher respondents know and understand how assessment should be used in the classroom. However, teachers might need to strengthen their knowledge by knowing how to write, develop and use a range of item types and assessment tasks that are technically sound, editorially correct and that appropriately assess the outcomes and performance standard levels being assessed.

Figure 5.3

Item 7 of the MTALQ

7. A teacher wants to document the validity of the scores from a classroom assessment strategy she plans to use for assigning grades on a class unit. What kind of information would provide the best evidence for this purpose?
 - a. Have other teachers judge whether the assessment strategy covers what was taught.
 - b. Match an outline of the instructional content to the content of the assessment strategy.
 - c. Let students in the class indicate if they thought the assessment was valid.
 - d. Ask parents if the assessment reflects important learning outcomes.

Item 8 was the most difficult, as only 8.2% of respondents provided the correct answer. This item is shown in Figure 5.4. The item assessed knowledge about the reliability and the

required actions that would most likely increase the reliability in a multiple-choice examination. Most of the respondents (68.9%) selected an incorrect option, Option A. This might have been because of a lack of basic knowledge about reliability, or because they mixed up validity and reliability. Aligning the items better with the Standards by using a test blueprint would tend to increase the validity of the test (which, if it was done well, would also increase the reliability), but the surest way of increasing the reliability would be to increase the number of items in the test.

Figure 5.4

Item 8 of the MTALQ

8. Which of the following actions would most likely increase the reliability of Mrs. Lockwood's multiple choice end-of-unit examination in physical science?
- a. Use a blueprint to develop the test questions.
 - b. Change the test format to true-false questions.
 - c. Add more items like those already in the test.
 - d. Add an essay component.

5.4.3 MTALQ Standard 3: Administering, Scoring and Interpreting Assessment Results (Standard 10.2 in NPST).

This Standard consisted of five items. Item 14 was the easiest, with the highest percentage (74.2%) of correct responses. The corrected item-to-total score correlation was highest for item 17 (i.e., 0.236). The corrected item-to-total score correlation value for Item 14 was 0.189. Item 14 assessed the skills and knowledge associated with scoring procedures that

prepare a model or blueprint of the critical features and assigns scoring weights to these features. Moreover, all items in this Standard had an acceptable corrected item-to-total score correlation except Item 15 (i.e. $-.007$), which assessed the teachers' knowledge and skills in terms of interpreting assessment results. Both Items 14 and 15 are shown in Figure 5.5.

Figure 5.5

Items 14 and 15 of the MTALQ

14. Students in Mr. Jakman's science class are required to develop a model of the solar system as part of their end of unit grade. Which scoring procedure below will maximize the objectivity of assessing these student projects?
- a. When the models are turned in, Mr Jakman identifies the most attractive models and gives them the highest grades, the next most attractive get a lower grade and so on.
 - b. Mr. Jakman asks other teachers in the building to rate each project on a 5 point scale based on their quality.
 - c. Before the projects are turned in, Mr. Jakman constructs a scoring key based on the critical features of the projects as identified by the highest performing students in the class.
 - d. Before the projects are turned in, Mr. Jakman prepares a model or blueprint of the critical features of the product and assigns scoring weights to these features. The models with the highest scores receive the highest grade.
15. At the close of the first month of school, Mrs. Friend gives her fifth grade students a test she developed in social studies. Her test is modeled after a standardized social studies test. It presents passages and then asks questions related to understanding and problem definition. When the test was scored, she noticed that two of her students who had been performing well in their class assignments, scored a lot lower than other students. Of the following types of additional information which would be most helpful in interpreting the results of this test?
- a. The gender of the students.
 - b. The age of the students.
 - c. Reliability data for the standardized social studies test she used as the model.
 - d. Reading comprehension scores for the students.

Responses to Item 14 (see Figure 5.5) indicated that 74.3% of respondents chose the correct answer. These results suggest that most teachers seem to possess the basic knowledge of scoring procedures to assess student projects, although it must be noted that according to the OECD (2020) report, student projects are not commonly used as a main means of assessment in the KSA. As discussed previously, this can be attributed to the prevalence of using multiple-choice questions as the main means of assessment in schooling. Nevertheless, the results on this item can be interpreted as encouraging because they infer that KSA teachers have the knowledge to assess students in ways other than multiple-choice questions if schools ensured that other forms of assessment such as student projects were part of the assessment regime.

Considering the foregoing discussion, assessment literacy professional learning programs should have as one of their aims the improvement of teachers' knowledge and skills in scoring so that more advanced and vital skills are developed. This includes developing teachers' knowledge of the various ways to score and record results to capture a snapshot of student performance that can be compared against performance standards. Furthermore, this can help demonstrate ways that results of student assessments are scored, recorded and used to arrive at an overall grade representing a performance standard.

Moreover, Advanced teachers should have the knowledge to be able to process how to accurately score assessments and collect and record student results. Advanced teachers can share exemplars with less experienced teachers on how to collect, score and record students' results in a systematic way and identify the potential problems with marking a range of assessment types to avoid bias in scoring students' work (Gipps, 1999). In this regard, Tognolini (2020) and Popham (2009) view assessment as a means of collecting evidence. Therefore, to increase the validity and reliability of scoring, Advanced teachers should have knowledge of some of the

more common ways of achieving comparability at the school level in a standards-referenced assessment system, as well as knowledge of its strengths and weaknesses. They should also be able to describe the ways that consensus works in moderation and how to employ moderation using common tasks that can be applied within schools to obtain comparability across classes. Along with their knowledge on the strengths and weaknesses of consensus and common-task moderation, Advanced teachers can assist Practitioner teachers in these aspects of assessment literacy.

Item 15 measured teachers' skills and knowledge in relation to interpreting the results of tests. Teacher responses to Item 15 (see Figure 5.5) showed that only 40.9% of respondents could provide the correct answer, while more than half of respondents, that is, 55.7 percent, selected the incorrect response of Option C. This item also had a negative corrected item-total correlation (i.e. $-.007$). What this means is that respondents with weaker assessment literacy as measured by the MTALQ, chose Option D and those respondents who are stronger chose incorrect options more often. These results suggest that these teachers lack the knowledge of how to distinguish between data (responses), information (responses as marks) and evidence (information considered through the lens of outcomes or other educational aims) to interpret assessment results and identify where and how evidence is used in schools to inform teaching. This underscores the importance of developing teachers' knowledge and skills to be able to align information needs with assessment content and structure. Such compatibility is crucial (Brookhart & Durkin, 2003).

The results of Item 15 corroborate the OECD report (2020). According to this report, the assessment methods of teachers in Saudi Arabia primarily measure factual knowledge and pay little attention to measurement skills such as problem solving and critical thinking measuring.

These results provide further evidence of the necessity to provide teachers in the KSA with a structured professional learning program that is aimed to enhance their ability to use assessment methods. This will ultimately help measure such skills and improve student achievement.

5.4.4 MTALQ Standard 4: Using Assessment Results for Decision-making (Standard 10.4 in NPST).

Standard 4 in the MTALQ was related to teachers' ability to transform data into actionable information and evidence to support student learning to measure Standard 10.4 in the NPST. Data-driven decision-making (DDDM), also known as 'evidence-based decision-making', consists of multiple processes that enable teachers to make professional judgements about student work or monitor student progress. Commonly, the process focuses on the teachers' capacity to collect and organise data, and then analyse, summarise and synthesise the information for use as evidence to support student learning. Some knowledge and skills related to measurement and statistics are required for teachers to engage in DDDM.

There were five items in Standard 4 in the MTALQ to measure Standard 10.4 in NPST. The easiest was Item 18 (Figure 5.6), as indicated by 78.4% correct responses. This item assessed the teachers' knowledge and skills regarding diagnostic assessment. These were considered by the focus group in Stage 2 of this thesis (the process to refine the assessment literacy framework) as essential skills for both career levels of teachers, Practitioner and Advanced. The correct Option was Option D. Option A was selected by 4.2%, B was selected by 14.5% and C was the least commonly selected, chosen by just 2.4% of respondents. The results of Item 18 might indicate that teachers in the KSA possess some knowledge regarding use of diagnostic assessment to know what knowledge and skills the students already have before planning a new lesson. Nonetheless, professional learning programs and ITP should develop

teachers' ability to demonstrate some of the ways to establish students' prior knowledge and performance. In this regard, Advanced level teachers should be able to support Practitioner level teachers to demonstrate some of the ways to ascertain students' prior knowledge and performance. Item 18 was developed collaboratively with the participants. Hence, the results might reflect the degree of validity of the item, as it is considered by the focus groups to be consistent with the KSA context.

Figure 5.6

Item 18 of the MTALQ

18. Before planning a lesson, a teacher wants to know what knowledge and skills the students already have in the unit of work. Which of the following would be the best way to find out?
- a. Do nothing as the teacher has already given instruction on the pre-requisite skills.
 - b. Review the past lessons homework to understand students depth of learning.
 - c. Conduct structured interviews with each student.
 - d. **Administer a short quiz.**

The most difficult item was Item 20 (Figure 5.7), with the lowest frequency of 18.9% correct responses. Fifty-six percent of respondents selected Option C, which gave the answer that 'both norm- and criterion-referenced information are equally helpful for instructional planning'. The response may indicate a lack of macro- and micro-level knowledge about standards-based assessment in the KSA context. The finding here is not surprising as there is little encouragement for teachers to use standards-referenced rather than norm-referenced assessments in the KSA education system. Sadler (1989) argues that without clear aims or referenced standards, students are unable to judge the quality of their work. This perspective is supported by Tognolini and

Davidson (2012), who view standards-referenced assessment regimes as a practical approach in assessment. Practitioner level teachers might encounter some challenges when dealing with the standards, but engaging in a professional learning community, especially with Advanced or Expert level teachers, can provide them with sufficient knowledge on how to use implicit and explicit standards as referenced systems to gauge students' work alongside their experience (Sadler, 1989). Moreover, teachers should have the knowledge of when and how to assess and report student progress against standards so that they are able to complete and interpret reports for stakeholders in a standards-referenced assessment system. This is exhibited in Figure 5.7.

Figure 5.7

Item 20 of the MTALQ

20. When planning classroom instruction for a unit on arithmetic operations with fractions, which of these types of information have more potential to be helpful?

norm-referenced information: describes each student's performance relative to a other students in a group (e.g. percentile ranks, stanines),
or

criterion-referenced information: describes each student's performance in terms of status on specific learning outcomes (e.g., number of items correctly answered for each specific objective)

- a. Norm-referenced information.
- b. Criterion-referenced information.
- c. Both types of information are equally useful in helping to plan for instruction.
- d. Neither, test information is not useful in helping to plan instruction.

The highest corrected item-to-total score correlation of 0.219 was recorded for Item 19, in Figure 5.8. Item 18 had the second highest item-to-total score correlation (0.163).

Figure 5.8

Item 19 of the MTALQ

19. To evaluate the effectiveness of the mathematics program for her gifted first graders, Ms. Allen gave them a standardized mathematics test normed on third graders. To decide how well her students performed, Ms. Allen compared her students' scores to those of the third-grade norm group. Why is this an incorrect application of standardized test norms?
- a. The norms are not reliable for first graders.
 - b. The norms are not valid for first graders.
 - c. Third grade mathematics items are too difficult for first graders.
 - d. The time limits are too short for first graders.

The responses to Item 19 (Figure 5.8) show that slightly more than half of the respondents, 51.2%, selected the correct answer of Option B. This item assessed teachers' knowledge and ability to use various assessment methods to evaluate student performance professionally. While a concern, these results seem to be consistent with the findings of other items discussed, as well as the OECD report, which indicate that KSA teachers are lacking in this area of assessment literacy. As discussed previously, assessment methods used by KSA teachers are known to be limited and this may reflect the traditional culture of rote learning that continues to prevail in the KSA educational system.

In view of the above, professional learning programs and Initial Teacher Preparation (ITP) should equip KSA teachers with the knowledge of assessment data that includes the ability to understand and use assessment data effectively alongside describing the importance of validity and reliability procedures in classroom assessment. Such knowledge would allow them to develop a range of assessments that consider potential sources of bias, such as those caused by differences in gender and race. Furthermore, referring to the NPST, one of the essential aspects of assessment literacy for Practitioner teachers is to have the knowledge of the range of

assessment forms along with their strengths and weaknesses to make an accurate judgement about students' work. On the other hand, Advanced teachers should be equipped with the knowledge and understanding of the assumptions that underpin educational assessment by being able to describe the major assumptions and produce evidence to support the claim that the assessments are fair and measuring what they are supposed to be measuring.

5.4.5 MTALQ Standard 5: Using Assessment in Grading (Standard 10.2 in NPST).

Standard 5 in the MTALQ pertained to using grading in assessment, is another crucial aspect of assessment literacy. Various aspects of knowledge in the area should be addressed to support teachers to conduct assessment effectively, such as having the knowledge of reliability, statistics as well as the strengths and weaknesses of multiple methods used in grading.

Standard 5 in the MTALQ consisted of six items to measure Standard 10.2 in the NPST, as shown in Table 5.3. The easiest item was Item 25, with 69.1% of teachers responding correctly. As stated previously, this standard assessed teachers' knowledge regarding the use of assessment in grading; therefore, the focus group deemed these skills essential to determine the actual level of student achievement and performance. This may have prompted teachers to respond appropriately by choosing the Option that involved explaining to the parent how the grade was determined and displaying samples of the student's work to effectively communicate student progress. In addition, responses to Item 25 (see Figure 5.9) are consistent with Newfields's (2006) ideal strategies for promoting teacher assessment literacy. Teachers should be able to know how to, 'when deciding how to grade a course, make sure it is both educationally valid and sufficiently clear to all stakeholders' However, in KSA, many stakeholders, such as parents, do not properly understand school assessment methods (OECD, 2020) nor how to interpret the reports.

Figure 5.9

Item 25 of the MTALQ

25. When a parent asks a teacher to explain the basis for his or her child's grade, the teacher should:
- a. explain that the grades are assigned fairly, based on the student's performance and other related factors.
 - b. ask the parents what they think should be the basis for the child's grade.
 - c. explain exactly how the grade was determined and show the parent samples of the student's work.
 - d. indicate that the grading scale is imposed by the school board and the teachers have no control over grades.

The most difficult item was Item 24 (Figure 5.10), with 14.8% correct responses. The percentage of respondents choosing each of the options under Item 24 was as follows: 12.5% chose Option A; 11.4% chose Option B and 61.6% chose Option C. The results of item 24 might indicate that either the participants lacked sufficient knowledge about using assessment in grading or they lacked adequate data literacy, which would have enabled them to distinguish between using average scores and variation for this purpose.

Clearly, the teachers in this sample need more support in the professional learning program to understand the many ways to use basic statistical terms such as those used for analysing and summarising student performance data. This result is consistent with other items that were associated with statistics in the MTALQ. Therefore, professional learning programs or ITP should equip teachers with the skills to interpret summary data that is key in representing student performance. In terms of the refined assessment literacy framework, Practitioner teachers should be equipped with the knowledge of basic statistical terms such as mean, median and mode; variance and standard deviation; correlation; range; percentiles; and norms. Advanced

teachers should additionally have knowledge of the strengths and weaknesses of some of the more common statistics used to summarise student performance, while also having the ability to support Practitioner teachers in this area.

Figure 5.10

Item 24 of the MTALQ

24. A teacher gave three tests during a grading period and she wants to weight them all equally when assigning grades. The goal of the grading program is to rank order students on achievement. In order to achieve this goal, which of the following should be closest to equal?
- a. Number of items.
 - b. Number of students taking each test.
 - c. Average scores.
 - d. Variation (range) of scores.

Generally, the values of corrected item-to-total score correlation were in the lower range, the highest being only 0.125 for Item 27. Responses to Item 27 (Figure 5.11) indicate that 67% of respondents, or more than two-thirds of teachers, already possess the knowledge that decisions like grades on assignments should be based on more than one source of information. Such results might point to professional learning programs that aim to provide teachers with more advanced knowledge on assessment bias or how to approach making decisions about student performance or grading their works.

Figure 5.11

Item 27 of the MTALQ

27. During the most recent grading period Ms. Johnson graded no homework and gave only one end-of-unit test. Grades were assigned only on the basis of the test. Which of the following is the major criticism of how she assigned the grades?
- a. The grades probably reflect a bias against minority students that exists in most tests.
 - b. Decisions like grade assignment should be based on more than one piece of information.
 - c. The test was too narrow in curriculum focus.
 - d. There is no significant criticism of this method providing the test covered the unit's content.

The item with the lowest corrected item-to-total score correlation across all items in the MTALQ was Item 28 (i.e. -0.105), shown in Figure 5.12. Responses to Item 28 indicated that only 26.8% of respondents chose the right response. While this might be a low value, the result is not unexpected. Other items in the MTALQ that are associated with assessment data and data literacy received similarly low scores, which further supports the finding that teachers in this sample are lacking in assessment and data literacy. This indicates that more emphasis needs to be placed on assessment data as well as data literacy in the in-service professional learning or ITP programs. According to the NPST, Practitioner level teachers should know and understand how to collect and organise data for classroom use. In this way, they can effectively collect raw response data through different means (using technology, where appropriate) so that it can be used to evaluate student learning. For Advanced level teachers, professional learning programs should equip them with the knowledge and understanding that different kinds of data exist along with the knowledge to discern what kind of data is useful to inform decisions.

Figure 5.12

Item 28 of the MTALQ

28. What kind of data can be used to carry out statistical analysis? the MOST correct answer is:

- a. Qualitative and quantitative.
- b. Quantitative.**
- c. Qualitative.
- d. None of the above.

In this regard, the respondents selected distractor A for item 28 (Figure 5.12) as the most correct Option as both qualitative and quantitative data can be used to carry out a statistical analysis. This could indicate that the respondents may have misconceptions in terms of the appropriate uses of qualitative and quantitative data.

The results of item 28 demonstrate a lack of assessment data knowledge among teachers in KSA based on this limited sample. Therefore, it is suggested that a professional learning program focused on the elements of data literacy that are related to showing teachers how to collect, analyse, and interpret data to help students learn more effectively and make the appropriate teaching adjustments, might be a logical addition to the professional learning of teachers in the KSA.

In summary, using grading in assessment is considered an important aspect of assessment literacy. Various aspects of knowledge in the area should be addressed to support teachers to conduct assessment effectively, such as having the knowledge of reliability, statistics, as well as

the strengths and weaknesses of multi methods used in grading. However, conducting tests, grading and trusting the outcomes of assessments to make decisions regarding students' academic performance is still fraught in KSA because of problems associated with the reliability of the assessments and their interpretation.

5.4.6 TALQ Standard 6: Communicating Assessment Results (Standard 10.5 in NPST).

Standard 6 in the MTALQ is related to communicating assessment results. This skill is viewed as a crucial skill in assessment practice, particularly when it comes to reporting student achievement and performance or providing stakeholders with accurate data that can inform them of assessment strategies used in the classroom and school.

There were only four items relating to Standard 6 in the MTALQ. The easiest item was Item 31 (Figure 5.13), answered correctly by 56.7% of the respondents. As stated previously, participants in Stage 2 of this thesis agreed that assessment ethics were crucial for all aspects of the assessment process, including in the selection, creation, scoring and communication of assessment results and the administering of assessments. This item assesses teachers' understanding of the potential ethical issues that arise when communicating results to students. However, 56.7% of the respondents selected the correct answer, which was Option A.

Figure 5.13

Item 31 of the MTALQ

31. Mrs. Brown wants to let her students know how they did on their test as quickly as possible. She tells her students that their scored tests will be on a chair outside of her room immediately after school. The students may come by and pick out their graded test from among the other tests for their class. What is wrong with Mrs. Brown's action?
- a. The students can see the other students' graded tests, making it a violation of the students' right of privacy.
 - b. The students have to wait until after school, so the action is unfair to students who have to leave immediately after school.
 - c. Mrs. Brown will have to rush to get the tests graded by the end of the school day, hence, the action prevents her from using the test to identify students who need special help.
 - d. The students who were absent will have an unfair advantage, because her action allows the possibility for these students to cheat

Item 32 (Figure 5.14) was more difficult. Only 18.2% of the respondents answered it correctly. More significantly, however, the results of Item 32 reveal that teachers in the KSA do not seem to have the knowledge to be able to differentiate between terms related to assessment practices such as formal, informal, formative and summative assessment when communicating the meaning of their marks (or results). The results are like those of Standard 2 in the MTALQ, Item 6, which further supports the finding that teachers are lacking in this area of assessment literacy. Therefore, the suggested professional learning programs would aim to improve teachers' skills and understanding of how to give meaning to results by comparing them to standards, normative or ipsative referenced groups. They would also focus upon enhancing teachers' understandings of the primary distinction between formal and informal assessment as connected to the structure of the assessment tools. Hence, teachers must know that formal

assessment is often more structured than informal assessment. For example, formal assessment may include classroom tests or the most structured tools, such as examinations. In contrast, informal assessment tools include methods of assessment like observations and conversations, which are generally relatively less structured. As 41.9% of respondents chose Option D, which was ‘formative and summative assessment’, this may suggest that these teachers may need more knowledge concerning formative assessment, summative assessment and feedback. The response options for Item 32 (Figure 5.14) also suggested that the respondents used formal, informal, formative and summative assessment interchangeably as Option A was correctly chosen by only 18.2% of respondents, and Option D was selected by 41.9%.

Figure 5.14

Item 32 of the MTALQ

32. Mr. Klein bases his students' grades mostly on graded homework and tests. Mr. Kaplan bases his students' grades mostly on his observation of the students during class. A major difference in these two assessment strategies for assigning grades can best be summarized as a difference in:
- a. formal and informal assessment.
 - b. performance and applied assessment.
 - c. customized and tailored assessment.
 - d. formative and summative assessment.

The results of Item 32 should be considered fundamental to the creation of professional learning programs because they provide insightful information about the assessment literacy

level of teachers in relation to the assessment terms. Consequently, the interplay between these terms may have some significant impact on the efficacy of assessment procedures.

A vast majority of participants devoted their discussions in the validation process of the refined assessment literacy framework and the original TALQ on the importance of having more items in the test that focused on assessing teachers' knowledge and skills about communicating assessment results. As a result, there were four items to assess this knowledge. However, three items had corrected item-to-total score correlations under 0.01 (i.e., Items 29, 30 and 31). These findings highlight some issues with communicating assessment results, including understanding the purpose of administering national and international large-scale assessments and utilising data to enhance teaching and learning. For instance, teachers should have a greater understanding of communicating students' assessment results to colleagues in the professional learning community, as well as enhancing their ability to reflect on their assessment process and share standards to promote moderation and comparability strategies. While these problems could be addressed by providing teachers with a structured professional learning program that is aimed at helping them to communicate results in a more effective manner, mustering the required political will to provide and implement such programs for teachers could be a challenge due to various administrative, logistical and financial constraints.

5.4.7 MTALQ Standard 7: Modern Assessment, or Relevant Knowledge and Skills for Assessment (Standard 10.3 in NPST).

Standard 7 is a new standard in the MTALQ to measure Standard 10.3 in the NPST that contains items that measure the most needed skills and knowledge related to modern assessment, updated assessment practices and engaging students in the evaluation processes. As such, assessment literacy is essential to the process of developing the skills that students require to be

ready for a fast-evolving world. The following discussion illustrates the level of assessment literacy possessed by this KSA sample of teachers in terms of using rubrics, self-regulation, using criteria for grading and allowing students to view and participate in building the assessment criteria.

Standard 7 consisted of four items. Item 34 (Figure 5.15), with 72.9%, of the respondents answering correctly, was the easiest item. The frequencies of responses for Options A, C and D were 7.3%, 14.2% and 5.2%, respectively. The results might suggest that respondents to some extent know one or more of the learning and assessment principles that underlie student-centred learning by improving their assessment skills and metacognitive skills to reflect on their learning, which influences self-regulation and deeper thinking. Brookhart (2011) argues that formative assessment is a well-known strategy in learning theories and another strength behind formative assessment (or assessment for learning) is that it enables student to self-regulate and make decision about their learning. She argues for the development of more skills and knowledge in educational assessment related to self-regulation. Moreover, Sadler (1989) articulated that formative and summative assessment and issues regarding self-regulation are metacognitive skills that allow students to evaluate their performance.

Definitions of self-regulated learning (SRL) have changed over time. SRL researchers have moved away from ‘decontextualized SRL measurement’ techniques towards ‘domain-specific’ and subsequently ‘context-sensitive SRL measurement’ techniques because of shifting definitions of SRL (Boekaerts & Corno, 2005). Hence, based on the current results, Advanced level teachers should assume the responsibility for supporting Practitioner level teachers’ learning of the fundamental principles of SRL as a critical element that allows students to reflect on their own learning. Moreover, Advanced level teachers should also be able to develop

instruments to measure students' SRL and ascertain the issues around validity and reliability of these instruments at school level to develop a school environment that fosters and enhances SRL.

Figure 5.15

Item 34 of the MTALQ

34. A seventh-grade teacher develops students' assessment skills to support self-regulation. We can clearly state that:
- a. the student's self-regulation is concerned with emotional aspects of learning only.
 - b. developing the student's assessment skills increases self-regulation.
 - c. developing assessment skills is for the teacher, not for the student.
 - d. self-regulation of the student contradicts his independence in learning.

The corrected item-to-total score correlation was above 0.01 for all items assessing this Standard, with the highest value of 0.346 for Item 36, followed by 0.294 for Item 34 and 0.289 for Item 35.

5.5 Chapter Summary

In this chapter, the survey results from the respondents have been presented and thoroughly discussed. The focus has been on the responses to the MTALQ (Modified Teacher Assessment Literacy Questionnaire). A significant finding from these results is the alignment with the conclusions drawn in the OECD's 2020 report on the Saudi educational system. The OECD report had pointed out that Saudi teachers exhibit a low level of assessment literacy, and the results of the MTALQ survey appear to support and corroborate this observation.

In the subsequent chapter, the discussion will shift towards the refinement of the assessment literacy framework and the validity of the assessment instruments. The primary objective of this next chapter is to provide a deeper understanding of these aspects. Furthermore, it aims to address the core research questions that underpin the study. By delving into the refined assessment literacy framework and instrument validation discussion, the study aspires to shed more light on the issues related to assessment literacy within the Saudi educational context and potentially offer solutions or recommendations to enhance it.

Chapter 6: Discussion

6.1 Introduction

This research focused on how aspects of NPST in the KSA are related to assessment literacy knowledge and skills. One aim of the study was to develop an assessment literacy framework that included skills and knowledge related to assessment literacy to better allow teachers to demonstrate that they have met the requirements of NPST as it relates to assessment literacy. An additional aim of this research was to gauge the validity of the MTALQ as constructed for use in KSA, as a tool for collecting information regarding teachers' knowledge and skills about assessment literacy. Following on from a presentation of the results in the previous chapter, this chapter discusses the results considering the three research questions posed for this study.

By way of introduction, the three research questions guiding this thesis are reproduced here.

- RQ1. What assessment literacy knowledge and skills do KSA teachers need to demonstrate that they have met the assessment standards described in the KSA NPST?
- RQ2. To what extent is the Teacher Assessment Literacy Questionnaire (the original TALQ) a valid measure of the assessment knowledge and skills required by KSA teachers to demonstrate that they have met the standards described in the NPST?
- RQ3: What evidence is available to support the validity of a new questionnaire (MTALQ) based upon the original TALQ, but augmented by items that better align it with the requirements of the KSA assessment literacy framework?

The first section of the discussion primarily addresses Research Question 1: What assessment literacy knowledge and skills do KSA teachers need to demonstrate that they have

met the assessment standards described in the KSA National Professional Standards for Teachers (NPST)?

6.2 Research Question 1: The Assessment Literacy Framework

RQ1. What assessment literacy knowledge and skills do KSA teachers need to demonstrate that they have met the assessment standards described in the KSA NPST?

The construction of the assessment literacy framework drew on the context of KSA, as mentioned above, to align and maintain consistency with the KSA teacher standards (NPST) through the MTALQ. The way the refined assessment literacy framework was used in the study was illustrated and discussed in the previous chapters. It was developed using document analysis, particularly addressing the questions of why, how and what skills and knowledge (expressed as elaborations) are important for teachers in KSA if they are to demonstrate that they have met assessment standards articulated in the NPST. The processes and strategies used to develop the refined assessment literacy framework were discussed as key aspects of the validity evidence for the efficacy of the refined assessment framework. Also discussed were the kinds of skills and knowledge that were or were not included in the refined assessment literacy framework, and how participant feedback guided its development was also discussed. In all, the results of these discussions with participants shed light on the knowledge and skills KSA teachers need to demonstrate that they have met the assessment standards described in the NPST. They are a key part of the validity evidence for the refined assessment literacy framework and the MTALQ.

Secondly the refined assessment literacy framework's associated knowledge and skills and the MTALQ are closely linked to the what, why and how to assess advanced by R. Stiggins (2009), as well as the process of communicating assessment results and the ways to engage students in self-assessment. Specifically, topics raised during the focus group discussions

regarding such issues as the importance of engaging students in evaluation processes through self- and peer-assessment and sharing the success criteria with students have impacted on the nature and structure of the assessment literacy framework.

Unlike standards in many other countries, such as Australia and the United Kingdom, the NPST specifies only one standard - Standard 10.3 - that concerns engaging students in the evaluation process. As the standard states and as the refined assessment literacy framework recounts, the Practitioner teacher ‘applies self-assessment strategies effectively and supports students to participate and make use of its results.’ The standard encourages Practitioner teachers to use self-assessment methods and the resulting data to inform teaching and learning. Taking this into account, decisions were made to include in the assessment literacy framework more details in terms of self- and peer-assessment to support teachers in demonstrating this standard. Other additions included ‘knowledge and understanding of the definition and purpose of self- and peer-assessment’; also, that the Practitioner teacher should also know how to describe some of the techniques that improve students’ assessment knowledge and ‘the Practitioner teacher should be able to describe how to apply self- and peer-assessment during instruction in a positive manner.’ Incorporating these elaborations should make teachers aware of the need to allow students to develop their own assessment skills which is deployed as a sub-standard in the refined assessment literacy framework. All these discussions, explicitly or implicitly, provide significant evidence that teacher assessment literacy knowledge and skills have been captured in the refined assessment literacy framework, allowing instruments like the original TALQ and the MTALQ to capture evidence of the assessment literacy of teachers in the KSA.

It is stressed that in this thesis the focus has been on using a small purposive sample of teachers to demonstrate the whole process of how results can be linked to judgements about

teacher assessment literacy and how to build an evidential argument to support the validity of the assessment instrument rather than draw conclusions across the sector of the assessment literacy of KSA teachers as a population. At the same time, assertions are made based on the small sample to demonstrate how the results from the MTALQ might be interpreted, in concert with evidence from major reports (e.g., 'OECD 2020' report), to infer meaning from the results.

One of the key outcomes from the focus group discussions was that enabling students to use self- and peer-assessment is one of the greatest challenges faced by KSA teachers. Participants in the focus group expressed a variety of concerns regarding involving students in the assessment process. Concerns advanced included a lack of time and a heavy workload. Participants agreed that the current obstacles would inhibit teachers in the KSA fulfilling the NPST standard, and that the refined assessment literacy framework might go some way in encouraging teachers to include such assessments in their assessment practices.

One of the questions in the MTALQ (item 36) is dedicated to assessing the competency of teachers in promoting the importance of students' skills in assessment:

A teacher allows students to view and participate in building assessment criteria. Which of the following is MOST correct?

(The most correct answer is that the students' participation in building the assessment criteria could improve their learning outcomes.)

This item was written to supplement the items in the original TALQ because it was seen as very important, and it was not assessed in the original TALQ. It is quite clear in the NPST that Advanced level teachers should be able to teach self-assessment strategies to students and develop effective strategies that enhance students' participation in the assessment process. In addition, there is a subliminal requirement that the Advanced level teachers should also be able

to support their Practitioner colleagues. Hence, the inclusion of this elaboration into the assessment literacy framework provides opportunity for Advanced level teachers to introduce self- and peer-assessment strategies to their students. What this inclusion does, in effect, is further enhance the validity of the resulting MTALQ. It is also important to acknowledge here the cumulative nature of the NPST and associated elaborations. Advanced level teachers should be able to demonstrate what is required by Practitioner level teachers, plus some more of the construct. Hence it is understood that these Advanced level teachers have an obligation to provide advice, where necessary, to teachers who are practicing at the Practitioner level. This is a characteristic of the professional teaching standards of most countries. It is also a requirement of measurement scales built using modern measurement theory.

The refined assessment literacy framework also includes the knowledge and understanding of the definition of self-regulation, which emerged in the focus group discussions. Self-regulation is not only essential in terms of engaging students in the assessment process, but also in developing the metacognitive skills which help learners use assessment methods to reflect on their performance (Glaser, 1981). Focus group participants stressed that self-regulation improves students' metacognitive learning and fosters lifelong learning through assessment as learning. Another idea derived from the discussion with the focus group participants reinforces the fact that Advanced level teachers can support Practitioner level teachers because it relies on the skill to design assessment tasks that provide students with feedback and the opportunities to implement such advice that contributes directly to their learning.

Item 34 was developed and added to the MTALQ to collect some evidence on self-regulation:

A seventh-grade teacher develops students' assessment skills to support self-regulation.

We can clearly state that:

- a. the student's self-regulation is concerned with emotional aspects of learning only.*
- b. developing the student's assessment skills increases self-regulation.*
- c. developing assessment skills is for the teacher, not for the student.*
- d. self-regulation of the student contradicts his independence in learning.*

Participants in the focus groups provided invaluable information for building the refined assessment literacy framework and questionnaire as outlined in the research questions. For example, they emphasised the need to operationalise assessment effectively as a tool to support teaching and learning. Advanced teachers might use various strategies to assess their assessment practices. This was anticipated in the refined assessment literacy framework by including the elaboration, 'know and understand the many ways to encourage students to be involved in self- and peer-assessment'. Understanding multiple ways to assess can improve the reliability of teachers' assessment practices by utilising a variety of sources of evidence and this, in turn, encourages teachers to demonstrate the full breadth requirements imbedded in the NPST. Driven by these considerations, further elaborations and their related details were included to support teachers in thinking deeply about the effective methods that can help to engage students in self- and peer-assessment. For example, in the refined assessment literacy framework, this is reflected by incorporating the elaboration for Advanced teachers: 'select, develop and apply strategies to reinforce student's participation in self- and peer-assessment'.

The refined assessment literacy framework also has the added advantage that it is consistent with the recommendations of the OECD (2020), which state that the KSA educational system should develop a national assessment framework. Consistent with this OECD challenge,

the refined assessment framework identifies whole-country and classroom assessment activities, strategies, methods, and purposes by incorporating assessment principles into it as a foundation.

The *Standards for Teacher Competence in Educational Assessment of Students* (AFT et al., 1990) created much debate when it was launched. Fulcher (2012) embarked on an empirical study to discover the types of skills and knowledge teachers need to conduct effective assessment. Nichols et al. (2009) also developed a framework for evaluating assessment literacy, in which they emphasised the educational context as an essential component in formative assessment or assessment for learning. K. Hill and McNamara (2012) conducted an empirical study to establish a framework in classroom-based assessment. Their framework incorporated three dimensions of assessment: evidence, interpretation and use. These dimensions included the various aspects of classroom assessment, which have similarities to this study's current framework; such as, planning, conducting and using assessment, feedback and reporting. This was the stimulus to conduct research into the measurement of teachers' assessment literacy and to develop their skills and knowledge in assessment literacy by developing an instrument and assessment literacy framework to characterise and analyse assessment literacy in the KSA education system.

Recall that the assessment literacy framework focuses on only what is considered the essential knowledge and skills needed to be assessment literate (see Chapter 3 for details regarding the process used to establish essential skills) and that previously published literature (see Chapters 1 and 2) indicates that KSA teachers have, in general, low assessment literacy. The assessment literacy framework was developed to focus only on the most essential skills as perceived by teachers in the KSA. Such a context might mean that more advanced aspects of

assessment were beyond the scope of this study and hence its assessment literacy framework and questionnaire.

To simplify the interactions between the knowledge and skills teachers need and the assessment standards as described in the KSA NPST, a new assessment literacy framework was developed. It was also necessary to build the MTALQ as one potential data collection instrument. One of the unique aspects of the current refined assessment literacy framework is that it describes two critical aspects of assessment literacy related to NPST emphasising the use of assessment to engage students in evaluation processes (i.e., self- and peer-assessment), and to report student achievement. Hence, the current refined assessment literacy framework could entice teachers in KSA to utilise strategies to evaluate the extent to which their performances has achieved the NPST standards.

The current refined assessment literacy framework also crucially provides KSA teachers with a way to better understand the many skills and knowledge in assessment literacy that are compatible with the KSA context. Moreover, these assessment literacy skills and knowledge are also recognised globally as essential skills because of the similarity of standards across many countries (such as the purpose, application and use of assessment data, and reporting of student achievement), which importantly enhances the transfer of such knowledge and skills outside of the KSA educational context. Furthermore, one of the best aspects of the refined assessment literacy framework is how the descriptions it incorporates can demonstrate the importance of the evaluation process when considering whether the assessment strategies used match teachers' and students' needs, as in 'Describe ways to evaluate whether the feedback has been understood by the students.' For example, this can be done through interacting with students to ensure there is a

shared understanding of terminology and language, and reflecting on how the feedback that has been given may assist students in improving their language.

The focus group discussions additionally raised some issues related to operationalising the current skills and knowledge in the refined assessment literacy framework to promote teachers who meet the NPST. The participants worked together to utilise the elaborations in the refined assessment literacy framework and additional details as interdependent skills, which means that the skills are connected in a cohesive manner. For example, a Practitioner teacher should first be able to ‘Describe the effective ways that assessments are used in schools, such as monitoring certifying student and teacher performance.’ Following this, the Practitioner teacher could then move to another set of skills, such as ‘Explain the main difference between a whole assessment strategies system (such as assessment for and of learning, school-based and standardised assessment), and relevant terms: standards, performance standards, outcomes and higher order thinking skills.’

Issues like the social aspects of assessment and the assessment context have continued to gain attention in assessment practices. Similarly, themes like standards-referenced assessment systems have emerged to be prominent topics in the assessment field. Therefore, the current refined assessment literacy framework should also cover the important skills and knowledge associated with data literacy as fundamental knowledge to use to conduct classroom assessment. This includes knowledge that allows teachers to know the meaning of data, information and evidence, as well as to understand that evidence is used to inform educational practice and reporting. Additionally, teachers should recognise that knowledge and understanding are a range of skills needed to transform data into information, evidence and strategies for teaching and learning.

During the process of revising the elaborations in the assessment literacy framework, the social and accountability assessment context of KSA were considered because of the importance of these factors. For example, the discussion of the survey results suggested that KSA teachers are struggling with a heavy workload, which impacts their ability to effectively provide opportunities for students to collaboratively develop standards and rubrics or for teachers to know the structure of national and international scales. The focus group participants, decided to revise some skills in the assessment literacy framework that were included initially as elaborations, but did not fit well into the KSA context even though they could be implied from the NPST. A contributing factor to the KSA teachers not having the time to take on the social and accountability type elaborations was that the education system of KSA is centrally controlled and it does not encourage or help the teachers to conduct more innovative assessment practices.

The current refined assessment literacy framework contains specific skills and knowledge that are aligned with the KSA context such as ‘Advanced teachers know and understand how to interpret the basic results of programs like SAAT and the NAP’. In doing this, the refined assessment literacy framework provides pathways for policymakers to include more specific details in the NPST and assessment standards that encourage and support teachers to improve their assessment literacy in accord with the implicit requirements of the NPST. In this way there may be an opportunity to better align the intent of the NPST with the KSA context.

Finally, during the focus groups, the participants found it necessary to establish systemic ways to integrate the NPST standards, MTALQ and the refined assessment literacy framework – ways that provide a pathway from basic level assessment skills to more advanced knowledge and assessment skills which provides the stimulus to expand and extend teachers’ capacity to conduct

the assessment process effectively, improve teaching and increase students' academic performance.

After outlining the strategies, processes and needs for developing the refined assessment literacy framework, to make it more valid for use in the KSA context as was advised during the discussion with teachers, the following discussion deals with the specific knowledge and skills KSA teachers need to demonstrate to show that they have met the assessment standard described in the NPST. The focus group participants refined and produced the final set of elaborations based on the NPST that was the basis for the refined assessment literacy framework.

The refined assessment literacy framework was used to align the KSA NPST with the MTALQ. By deliberately considering the context of the KSA education system, the refined assessment literacy framework aims to provide consistency and facilitate a comprehensive understanding of the assessment standards outlined in the NPST.

The absence of research and a clear assessment literacy framework specific to KSA teachers underscores the significance of conducting this study to determine and potentially enhance the current level of assessment literacy among KSA teachers (Alanazi, 2017; MOE, 2017). These efforts align with the educational reforms in KSA and recognise the vital role of teacher assessment practices in improving student learning outcomes (Barnes et al., 2015; Deneen & Brown, 2016; Fulmer et al., 2015).

To demonstrate their proficiency in the assessment standards described in the NPST, it is essential that teachers in the KSA possess knowledge and skills in key aspects of assessment literacy. In the context of the KSA government's desire to raise the quality of education—in which sound levels of teacher assessment literacy are key to long-term success in student outcomes—this would include having a sound knowledge of assessment principles such as

validity, reliability, fairness and assessment for learning. Teachers of the KSA should also be skilled in designing and developing assessments that align with curriculum objectives as well accurately measure student knowledge, skills and understanding. Teachers should also be proficient in designing formative assessment strategies, including monitoring student progress, providing timely feedback and using assessment data to inform instruction. Additionally, they need to demonstrate skills in data literacy and analysis, design inclusive assessments, provide effective feedback, report assessment results, adhere to ethical guidelines and uphold professional standards in assessment practices. By possessing these assessment literacy competencies, KSA teachers have a greater chance of effectively meeting the assessment standards outlined in the NPST and contribute to improved education outcomes.

A purposive sample of teachers was asked to judge the elaborations of the NPST standards according to their importance using a simple Likert scale where ‘Essential’ were scored as 2, ‘Desirable’ were scored as 1, and ‘Not Necessary’ were scored as 0 (Table 4.1). Clearly, the results from these judgements suggest that many of the elaborations were aligned with the NPST, with 62 of the 65 elaborations scoring as ‘Desirable’ or ‘Essential’ rating with only one elaboration scoring below ‘Desirable’. As such, it would be reasonable to contend that the current refined assessment literacy framework, on the balance of probability, would capture the most essential knowledge and skills of the NPST as perceived by KSA teachers.

The focus group discussions with participants considered the importance of each of elaborations and the rating that the teachers who responded to the survey ascribed to them. The feedback gathered from the participants informed the efficacy of the assessment literacy framework because elaborations that participants deemed as ‘problematic’ were revised while elaborations that were rated as ‘not essential’ were also discussed. The outcome of this process is

the refining of the assessment literacy framework which was better suited to the KSA educational context as it is one that has been moderated by the perspectives of KSA teachers themselves. Therefore, in effect, this process led to a revision of the original TALQ to produce the MTALQ.

Primary, intermediate, and secondary teachers recognised the significance of a wide range of assessment literacy knowledge and skills. Primary teachers generally regarded all the elaborations as important, secondary teachers attributed higher levels of importance to a larger number of elaborations across both career levels.

The mean ranges obtained from the survey responses provide a representation of the perceived importance of different elaborations among teachers at different career levels (Practitioner and Advanced) in both primary intermediate and secondary schools. Table 4.1 illustrates the mean ranges obtained from the judgements of primary teachers on the relative importance of the elaborations associated with the Practitioner and Advanced levels of Standard 10 in the NPST. Notably, none of the elaborations received mean scores below 0.49 for either career level, indicating that primary teachers generally regarded all the elaborations as important. One elaboration (Item 23) fell within the mean range of 0.5 to 0.99 at the Practitioner career level, suggesting a moderate level of importance. A substantial number of elaborations (17 in total) received mean scores between 1.0 and 1.49, indicating a higher level of importance as perceived by primary teachers. Moreover, 14 elaborations were rated within the mean range of 1.5 to 2.0, reflecting the highest level of importance attributed to these elaborations (Table 4.1).

In examining the primary teachers' focus group data, several key findings emerged regarding the ratings of specific elaborations within the NPST. Elaborations 14 (formative assessment understanding), 16 (self- and peer-assessment), and 43 (knowledge of major testing

programs) were rated as desirable but not essential by survey respondents. The discussions highlighted the need for clearer definitions and improved understanding among teachers regarding formative assessment concepts, the purpose of self- and peer-assessment, and the relevance of extensive knowledge about major testing programs. These findings emphasise the importance of refining these elaborations to enhance teachers' comprehension and ensure their effective application in the KSA context.

In contrast, primary teachers' responses to Elaborations 23 (knowledge of statistical terms), 29 (use of feedback in teaching), 51 (providing opportunities for students to construct criteria) and 57 (Knowing and understanding how to interpret the results of large-scale testing programs) varied significantly from what the focus group participants believed. Elaboration 23, despite being rated lower than expected, was recognised as important by focus group participants, leading to a revision that emphasised understanding basic statistical terms. Elaboration 29 caused some confusion because of its wording, prompting the recommendation to revise it to focus solely on the use of feedback in teaching. Participants acknowledged the challenges associated with elaboration 51 and suggested emphasising the importance of providing opportunities for students to construct criteria. Finally, elaboration 57 generated different opinions between focus group participants and survey respondents, with primary teachers advocating for its removal at certain career levels while acknowledging its overall desirability. These diverse findings underscore the need for tailored approaches and clearer language in the elaborations to effectively assess teachers' assessment knowledge and skills in line with the standards prescribed in the KSA.

Participants in the focus group expressed that several elaborations did not require any amendments based on their importance and alignment with the required skills and knowledge for

teachers. Elaboration 20, which focuses on feedback techniques to engage students, was deemed essential as teachers should understand the requirements and contextual use of feedback. Elaborations 21, 22, 24, and 25 regarding data literacy and the transformation of data into information and evidence were considered critical for teaching and learning. Elaborations 30 and 31 related to understanding purposes and strategies for reporting student assessment information were deemed important without requiring changes. Elaboration 32, emphasising the meaning of grades and communicating student achievement, was recognised as significant for all teachers. Elaboration 36 is focused on developing items and tasks based on outcome analyses and Elaboration 37 on developing high-quality assessments were both seen as important at both Advanced and Practitioner levels. Elaborations 44 and 47 related to recording student results and understanding self-regulation were considered essential skills for Advanced teachers. Lastly, Elaboration 56 highlighting the understanding of different types of data and its application to decision-making was viewed as crucial for designing assessments aligned with students' capabilities.

Table 4.2 in Chapter 4 presents the mean ranges derived from the judgment of secondary and intermediate teachers regarding the importance of the elaborations associated with the Practitioner and Advanced Career Levels of Standard 10 in the NPST. A comparison between Table 4.1 and Table 4.2 reveals that secondary and intermediate teachers generally rated more elaborations at higher levels of importance compared to primary teachers which might reflect that primary teachers in the KSA context lack important skills and knowledge related to assessment literacy and might need support to enhance their assessment literacy. At the Practitioner level, one elaboration fell within the mean range of 0.5 to 0.99, indicating a moderate level of importance as perceived by secondary and intermediate teachers. Fourteen

elaborations received mean scores between 1.0 and 1.49, signifying a higher level of importance. Additionally, 17 elaborations were rated within the mean range of 1.5 to 2.0, denoting the highest level of importance attributed to these elaborations by secondary teachers. For the Advanced level, six elaborations obtained mean scores between 1.0 and 1.49, indicating their significance, while a substantial number of elaborations (27 in total) fell within the mean range of 1.5 to 2.0, reflecting the highest level of importance as perceived by secondary teachers (Table 4.2).

A summative statement for these results would be that secondary teachers in KSA rated most elaborations in the survey as essential or desirable. Yet, the literature review indicated that teachers in KSA have in general, a low level of assessment literacy (OECD, 2020). What can be inferred from this is that there seems to be a gap between KSA teachers' knowledge and their actual assessment practices, especially when the questionnaire results (i.e., teachers' responses to the MTALQ) corroborates what the literature (e.g., OECD report (2020)) states. That is, this relatively small, purposefully sampled group of KSA teachers has low assessment literacy. Hence, what could be said is that it seems evident that teachers in the KSA have a positive attitude towards assessment (Albalawi, 2018) and the importance of certain aspects of assessment literacy in the KSA educational context. However, it seems to be the case that professional learning programs (as discussed in Chapter 2) might not have focused much on developing KSA teachers' assessment literacy and the consequence of this is the low levels of assessment literacy currently exist amongst those KSA teachers who participated in this MTALQ.

During the focus group discussion with secondary and intermediate teachers, there were suggestions for amendments to two specific elaborations, namely item 14 and item 16. In line

with the primary focus group participants, the secondary teachers recommended changing the term ‘constructive evaluation’ to ‘formative evaluation’. Their rationale behind this proposal stemmed from the fact that ‘formative evaluation’ is a concept that is more commonly used and understood within the education system of the KSA. This differential terminology may account for the KSA teachers performing relatively poorly on the formative assessment questions in the survey.

Furthermore, the secondary and intermediate teachers also expressed the view that self-assessment and peer-assessment should be combined into a single category, like the suggestions put forward by the primary focus group. This consolidation of self- and peer-assessment aligns with their belief that both approaches are integral to the assessment process and carry significant value in promoting student learning and development. By proposing these changes, the secondary and intermediate teachers aimed to enhance the clarity and relevance of the elaboration to the local educational context. The adoption of the term ‘formative evaluation’ and the amalgamation of self- and peer-assessment reflect their understanding of the prevailing practices and terminology in the KSA education system. These suggestions contribute to the ongoing efforts to refine the assessment literacy framework and ensure its effectiveness and applicability in the secondary education landscape, and to enhance teachers’ capacity to demonstrate that they have met the NPST Standard 10.

The mean ranges derived from the survey responses also provide insights into the assessment literacy knowledge and skills considered important by KSA teachers at different career levels to meet the assessment standards outlined in the NPST. These findings contribute to a deeper understanding of the assessment practices required to improve educational outcomes in

KSA schools and can inform the design of targeted interventions to enhance teachers' assessment literacy.

Furthermore, from the analysis of teachers' comments and feedback from focus group discussions on the skills and knowledge in the refined assessment literacy framework, it was found that the KSA education system limits the opportunities for teachers to utilise the appropriate assessment methods to assess student achievement effectively. For example, one hindrance is asking teachers to cover the whole of the textbook in a subject within a certain time. Thus, the discussions with teachers emphasised all elaborations in the refined assessment literacy framework that were deemed relevant and important. The teacher focus group discussions also supported the incorporation of a 'column' in the refined assessment literacy framework that describes (what teachers can do) to demonstrate the NPST.

Other issues such as workload, the top-down approach and socio-economic and curricular activities were raised during the discussions, which were aimed to discover the context in which the refined assessment literacy framework would be used to capture the Standard 10 in the NPST – this was discussed in Chapter 2 and is further elaborated below. In the case of developing assessment literacy frameworks in which the assessment methods may be affected by internal and external factors, using the 'Think Aloud Protocol' method might be useful to investigate the validity of the framework. Given that all these issues were raised by and discussed with the participants in the focus groups, it is believed that they have the potential to impact on the teachers' responses to the MTALQ. However, the results do not reflect the validity of the MTALQ (illustrated below) given that it was developed based on focus group participants' comments and feedback.

The findings also seem to illustrate a paradoxical tension between what is recognised as an essential skill to demonstrate the standard and teachers' ability to show that they possess the knowledge or skills. This is shown by their performance in the MTALQ where results indicate that the sample of KSA teachers taking the questionnaire have a low level of assessment literacy. Yet, they considered most of the elaborations as essential or desirable. Given this, there is a need to create a strong assessment culture that recognises the importance of assessment literacy and assessment practices at all levels in the KSA, such as at the macro- and micro-levels and amongst policy makers and all their key stakeholders.

In summary, based on the focus group discussions it seems evident that KSA teachers, at least in the sample who sat the MTALQ, are competent in being able to recognise the skills and knowledge of assessment literacy that are deemed essential to demonstrate in the NPST at the various teacher levels found therein. However, the results of the questionnaire (the MTALQ) indicate that overall, teacher participants possess a low level of assessment literacy. This finding corroborates with the findings found in the OECD report on the KSA educational system (OECD, 2020) which indicate that KSA teachers possess a low level of assessment literacy.

6.2.1 Utility of the Refined Assessment Literacy Framework

In addition to being sufficient to summarise the most important knowledge and skills of the refined assessment literacy framework, another idea that emerged from the analysis of the results is related to the utility of the refined assessment literacy framework - which is multifaceted. For example, the refined assessment literacy framework can be used by staff developers to create assessment literacy programs. Before teachers enter the teaching profession, teacher supervisors or principals can use the refined assessment literacy framework to obtain a

sense of teachers' knowledge in the fundamental assessment practices that demonstrate the respective teacher standards.

Furthermore, teacher professional developers can use the instrument offered by this research (i.e., the MTALQ) to assess teachers' assessment literacy prior to joining the program. Such employment of the questionnaire, which is based on the NPST, can help diagnose the level of assessment literacy a teacher possesses and the extent and kinds of professional development that are needed to develop a teacher's assessment literacy so that the professional learning can be targeted at the career stage in which the teacher aspires to achieve (i.e., zone of proximal development; Vygotsky, 1978).

Drawing on the results of the MTALQ in the study, the refined assessment literacy framework will also be able to function as a guide or a tool for teachers to evaluate their own assessment practices. Doing this will allow them to be able to gain insights into the strengths of their practices as well as aspects that they would need to improve. Therefore, by utilising the refined assessment literacy framework, teachers would be able to diagnose their individual assessment literacy in a manner that can help inform the kinds and extent of professional learning programs they might need to meet the NPST standards.

The refined assessment literacy framework has the potential to fulfil two other important functions. One, it can be a descriptive table of specifications that can be used to develop an instrument that achieves the desirable psychometric evidence and two, it can be used to provide assurance as to the construct and content validity of its standards and items. This is another process that adds to the evidential argument to support the efficacy of the MTALQ.

Furthermore, teachers could use the refined assessment literacy framework in moderation meetings to evaluate current assessment practices against curriculum standards to align the

assessment strategies used within current curriculum frameworks. Employing the refined assessment literacy framework in this manner has the significant potential to promote the essential comparability process across the classroom and the school level. Hence, teachers of various levels will be able to demonstrate that they have met the relevant NPST. In terms of teacher qualifications and readiness, teachers who intend to practise could also benefit from the current refined assessment literacy framework by enhancing their relevant knowledge and skills, which require development before sitting for the teacher licensing examination.

The next section addresses study research question 2.

6.3 Research Question 2: The Validity of the TALQ for Assessing NPST

RQ2. To what extent is the Teacher Assessment Literacy Questionnaire (the original TALQ) a valid measure of the assessment knowledge and skills required by KSA teachers to demonstrate that they have met the standards described in the in the NPST?

As discussed in Chapter 2, the TALQ was developed by Plake and Impara (1992) in the US and used in Oman. To determine if the TALQ was fit for purpose for use in the KSA and specifically for use in this study, validity information for all items as they applied across all NPST assessment standards was systematically investigated. Significant attention was paid to evidence which shows content alignment to the standards.

The primary vehicle for the collection of content validity related evidence comes primarily from evaluating the focus group discussions in which teachers in the KSA were asked to judge the alignment of TALQ test questions to the refined assessment literacy framework as well as provide their views as to the importance of these items as measures of assessment literacy.

The Standards for Educational and Psychological Testing (American Educational Research Association et al., 2014) define validity evidence as:

The body of evidence that supports the adaptation and the use of the test for its intended purposes (i.e., validity argument drawing from evidence based on test content, response processes, internal structure, relations to other variables, and consequences of testing; see American Educational Research Association et al. 2014). Note that not all types of evidence are necessary for every instrument.

The sample of KSA teachers participating in the focus groups were asked to judge the alignment and importance of each of the questions of the original TALQ relative to the NPST via the refined assessment literacy framework developed. The results of these judgements represent the primary validity evidence associated with the use of TALQ (and under the third research question presented later, the modification of TALQ) as a data collection device measuring performance of teachers against the NPST. Starting with TALQ, focus groups were asked to evaluate each of the TALQ items to determine, not only the extent of alignment with the refined assessment literacy framework but also how important they were in practice. Based on the analysis of the responses shared by teacher participants in the focus groups, it could be said that they provide strong evidence that the items selected by the focus groups were not only well aligned but that they measured important aspects of assessment literacy in the KSA.

In-service teachers were selected to be the participants in the focus groups in preference to pre-service teachers because the breadth and depth of experience of these teachers in classroom assessment provides more authentic data, which is more practical, meaningful, and likely to be more consistent when capturing judgements for the construction of both the refined assessment literacy framework and the instrument.

The participants worked together to provide evidence about the extent to which the original TALQ items might be consistent with the KSA context. The original TALQ was developed based on Teachers' Standards 1990, which implies that all items in the original TALQ may have some alignment and relevance to the content and for the construct of assessment literacy in at least the US context. Since it is very likely that there would be a large degree of overlap in the construct of assessment literacy for teachers across both the US and KSA educational contexts, adopting TALQ as the primary measurement instrument made logical sense (the latter sections of this chapter discuss modifications to TALQ to better align with the NPST - the MTALQ).

During the focus group discussions, participants made the decision to distribute items related to assessment ethics across all the Standards to increase the content and construct validity. This is premised on the argument that it is critical to recognise that ethical aspects of assessment, such as marking, scoring, administering, and using assessment data, are necessary skills and knowledge in any assessment practice and help to establish the validity of the instrument. For example, during the focus group discussions that were related to measuring Standard10.1 in NPST (preparing evaluation tools), the participants in the focus group emphasised that Items 3, 4 and 5 were the best ones to measure the ability of teachers to prepare the assessment 'evaluation' tools, because in the TALQ they were identified as measuring teachers' ability to choose assessment methods. The characteristics related to these 3 items are shown in Table 6.1.

Table 6.1

Connections Between the NPST, Items in the Original TALQ, Elaborations in the Refined Assessment Literacy Framework, Corrected Item-To-Total Score Correlation and Cronbach's Alpha if an Item is Deleted

NPST	Item in the original TALQ	Corrected item-to-total score correlation/ Cronbach's alpha if item deleted	Elaborations in the refined assessment literacy framework
10.1 Preparing evaluation tools	Item 3: Mrs Bruce wished to assess her students' understanding of the method of problem solving she had been teaching. Which assessment strategy below would be most valid?	−0.091 / 0.570	• Know and understand how to select assessment tasks which clearly relate to learning objectives • Know and understand how to use a range of assessment items that are fit-for-purpose • Teachers use a range of assessment items which effectively assess a range of skills in a unit of learning and mirror the relative weights of content materials or skills required by the unit of learning
	Item 4: What is the most effective use a teacher can make of an assessment strategy that requires students to show their work, e.g., the way they arrived at a solution to a problem or the logic used to arrive at a conclusion?	−0.055 / 0.578	• Know the meaning of feedback in assessment • Define and provide relevant examples of how feedback is used in schools • Know and understand the features of effective feedback • Know and understand of some feedback techniques to engage students in the instruction • Know and understand the requirements for effective feedback

NPST	Item in the original TALQ	Corrected item-to-total score correlation/ Cronbach's alpha if item deleted	Elaborations in the refined assessment literacy framework
	Item 5: Ms. Green, the principal, was evaluating the teaching performance of Mr Wajesa, the fourth-grade teacher. One of the things Ms. Green wanted to learn was if the students were being demonstrating higher order thinking skills. What documentation would be the most valid to help Ms. Green to make this decision?	-0.010 / 0.572	• Know and understand how feedback should be tailored to the needs of individual students • Know how to assess whether the feedback that is given in class is understood by the students • Teacher should explain the main differences between whole assessment strategies systems (such as assessment for and of learning, school-based and standardised assessments) and relevant terms: standards, performance standards, content, outcomes and higher order thinking skills • Know how to write, develop and use a range of item types and assessment tasks that are technically sound, editorially correct and that appropriately assess the outcomes and performance standard levels being assessed • Use assessment methods to identify current student performance, skills, achievement to be aligned with the lesson plan • Know and understand what it means to tailor assessments to the individual needs of all students

Item 3 asks:

Mrs Bruce wished to assess her students' understanding of the method of problem solving she had been teaching. Which assessment strategy below would be most valid?

- a. Select a textbook that has a 'teacher's guide' with a test developed by the authors.
- b. Develop an assessment consistent with an outline of what she has actually taught in the class.
- c. Select a standardised test that provides a score on problem solving skills.
- d. Select an instrument that measures students' attitudes about problem solving strategies.

Item 4 asks:

What is the most effective use a teacher can make of an assessment strategy that requires students to show their work, e.g., the way they arrived at a solution to a problem or the logic used to arrive at a conclusion?

- a. Assigning grades for a unit of instruction on problem solving.
- b. Providing instructional feedback to individual students.
- c. Motivating students to attempt innovative ways to solve problems.
- d. None of the above.

Item 5 asks:

Ms. Green, the principal, was evaluating the teaching performance of Mr Wajesa, the fourth-grade teacher. One of the things Ms. Green wanted to learn was if the students were demonstrating higher order thinking skills. What documentation would be the most valid to help Ms. Green to make this decision?

- a. Mr Wajesa's lesson plans.
- b. The state curriculum guides for fourth grade.
- c. Copies of Mr Wajesa's unit tests or assessment strategies used to assign grades.
- d. Worksheets completed by Mr Wajesa's students, but not used for grading.

The results of the focus group discussions and all the processes to develop the refined assessment literacy framework and related elaborations demonstrate the validity of evidence of these three items. This is because their content is closely related to Standard 10.1 in NPST. Participants used several strategies for evaluating the relationship between these 3 items and the first Standard in NPST. These include assessing the extent to which the item assessed whether the teacher realised the need to develop assessments consistent with the curriculum; making sure that they know the value of providing feedback to individual students that may lead to improved learning; and, they have an idea of the type of the assessment methods that may be used to measure skills that underpin higher-order thinking, rather than just those skills that have been directly taught in class. What can be said about the views of the focus group participants is that their judgement about the alignment of the items was consistent with previous reviews conducted when developing the TALQ.

Table 6.1 also provides some evidence obtained from an analysis of the MTALQ of the direct connection between the standard in the NPST, the items in the original TALQ, the skills identified in the refined assessment literacy framework, the Corrected Item-To-Total Score Correlation as well as Cronbach's Alpha. These data can then be used in conjunction with the focus group results to further understand the functioning of the items.

It is interesting to note that the teachers' performances across all three items obtained negative Corrected Item-To-Total Score Correlation and Cronbach's Alpha if the item is deleted will increase the reliability of the instrument. This might indicate that the items were not actually measuring what they were intended to measure or that the items were too difficult (even though they did assess the appropriate skill or knowledge) and consequently teachers tended to guess the correct answer.

Data obtained from both focus groups (i.e., the first focus group to validate the assessment literacy framework and the second to validate the original TALQ items against the refined assessment literacy framework) illustrate that Item 6 in the original TALQ was one of the important items in relation to validity and scoring procedures.

Item 6 asks:

A teacher wants to document the validity of the scores from a classroom assessment strategy she plans to use for assigning grades on a class unit. What kind of information would provide the best evidence for this purpose?

- a. Have other teachers judge whether the assessment strategy covers what was taught.
- b. Match an outline of the instructional content to the content of the assessment strategy.
- c. Let students in the class indicate if they thought the assessment was valid.
- d. Ask parents if the assessment reflects important learning outcomes.

Item 6 measures the ability of teachers to document the validity of the scores from classroom assessment strategies and it is seen as an essential skill to achieve the first Standard in the NPST (Standard 10.1). Additionally, validity is evidenced in relation to Item 6's content and

construct validity obtained from the second focus group discussion that aimed to validate the original TALQ. The participants decided that the item best measures the skills in the refined assessment literacy framework relative to the ability of the teachers to know and understand the scoring procedures and the implications of validity in classroom practices (for example, the validity of the scores from a classroom assessment method; see Table 6.2). Item analysis results indicate that Item 6 has a negative corrected item-to-total score correlation as well as a positive impact on Cronbach's alpha if an item is deleted.

Table 6.2

Item 6 in the Original TALQ, Corrected Item-To-Total Score Correlation/Cronbach's Alpha if Item Deleted, and Elaborations in the Refined Assessment Literacy Framework

NPST	Item in the original TALQ	Corrected item-to-total score correlation / Cronbach's alpha if item deleted	Elaborations in the Refined Assessment Literacy framework
10.1 Preparing evaluation tools	Item 6: A teacher wants to document the validity of the scores from a classroom assessment strategy she plans to use for assigning grades on a class unit. What kind of information would provide the best evidence for this purpose?	-0.046/0.561	• Know and understand how to develop items and tasks based on analyses of outcomes in terms of the factors that may affect these items • Know and understand the meaning of validity and reliability in relation of selecting assessment tasks • Know and understand the purposes of assessment results recording. • Know and understand the importance of collecting reliable and valid assessment data about the achievement of students and report it in a manner which is consistent with the reporting strategy(ies) used within a school • Know and understand how to work with students to analyse assessment results to identify their strengths and weaknesses • Explain the meaning of the various grades/marks that characterise most school reports: A to E grades, qualitative

NPST	Item in the original TALQ	Corrected item-to-total score correlation / Cronbach's alpha if item deleted	Elaborations in the Refined Assessment Literacy framework
			grades, benchmark-related grades, marks, ranks and percentages. • Describe and explain for teachers a. the process for developing grades and marks that have been collected to stakeholders; b. the process of reporting within scope of subject/syllabus being assessed; c. that the sequencing of learning is linked to reporting; d. appropriate next steps / aims for learning are identified; and, e. the technical specifications for formal assessments (e.g. mean, standard deviation, reliability, validity) to parents in a manner that helps them better interpret the report.

To substantiate the issue of validity, for example, the current findings of Items 8 and 9 (Table 6.3) indicate that teachers who participated in the focus group processes of the original TALQ argued that Items 8 and 9 would be appropriate to measure the capability of teachers to develop test items and generate ideas rather than recall the information. The item analysis results summarised in Table 6.3 show positive corrected item-to-total score correlation for both items and it also shows that the reliability of the test would decrease if they were deleted.

Item 8 asks:

Ms. Guardia wants to assess her students' skills in organizing ideas rather than just repeating facts. Which words should she use in formulating essay exercises to achieve this goal?

- a. compare, contrast, criticize
- b. identify, specify, list
- c. order, match, select
- d. define, recall, restate

Item 9 asks:

Mr. Woodruff wanted his students to appreciate the literary works of Edgar Allen Poe. Which of his test items shown below will best measure his instructional goal?

- a. 'Spoke the raven, nevermore.' comes from which of Poe's works?
- b. True or False: Poe was an orphan and never knew his biological parents.
- c. Edgar Allen Poe wrote:
 - 1. Novels
 - 2. Short stories
 - 3. Poems

- 4. All of the above.
- d. Discuss briefly your view of Poe's contribution to American literature.

Table 6.3

Items 8 And 9 in the Original TALQ, Corrected Item-To-Total Score Correlation/Cronbach's Alpha if Item Deleted, and Elaborations in the Refined Assessment Literacy Framework

NPST	Item in the original TALQ	Corrected item-to-total score correlation / Cronbach's alpha if item deleted	Elaborations in the refined Assessment Literacy framework
10.1 Preparing evaluation tools	Item 8: Ms. Guardia wants to assess her students' skills in generating ideas rather than just repeating facts. Which words should she use in formulating essay exercises to achieve this goal?	0.047 / 0.545	• Know and understand how to design summative assessment tools that will enable the results to be used to improve teaching and learning • Know the relationship between item difficulty and the level of cognition based and can use classification systems and taxonomies in building assessments • Explain how classification and taxonomies like Bloom's Taxonomy can be used to show growth of cognition in learning. • Know and understand some of the more common purposes of assessment, including how the assessment is likely to be used and what are likely to be the intended and unintended consequences of using these assessments • Know and understand some of the more common assessment dichotomies used in education and the meaning of relevant terms associated with them.
	Item 9: Mr Woodruff wanted his students to appreciate the literary works of Edgar Allen Poe. Which of his test items shown below will best measure his instructional goal?	0.062 / 0.565	

In relation to research question 2, there is evidence aggregated from the source of the focus groups and the item analysis of the MTALQ that the items that were retained after the review of the TALQ by the focus group have functioned in a way that gives confidence that they are effectively assessing the standards described in the KSA Teacher Professional Standards.

The next section reflects on some important aspects of validity to set the scene for the discussion of research question 3, which addresses the validity of the modified questionnaire referred to as the MTALQ.

6.4 Validity of the MTALQ

The process of validating any instrument related to assessment is a necessary and important process in research. It determines the extent to which the test can accurately measure what it is designed and expected to measure. However, it consists of, as Messick argues, infinite operations (1995). The current study, as presented in the first and second chapters, provides evidence that teacher assessment literacy in KSA is low. To this end, this study proposes and predicts interpretation of the test results, working as a foundation that could be built upon when analysing the test results to validate the MTALQ. This proposed interpretative framework is a method designed to enhance and make sure that the warrants (logical arguments as can be seen in the discussion chapter) contribute to the discussion regarding the extent to which the MTALQ is a valid measure to gauge the assessment literacy of teachers in the KSA. It draws on the analysis of documents, previous work, expert judgments, teacher focus group discussions, assessment theories, and the performance of the sample of teachers themselves in the test. Similarly, Chapters 1 and 2 contain several inferences and assumptions that can allow this proposed interpretative framework to be validated based on the test results, which provides further evidence of the validity of the test. It is also important to mention here that this chapter does not

seek to expand the discussion of all the issues related to the validity of the test because the proposed interpretative framework seems to initially align with the results. For example, it does not discuss in detail the reasons and factors that might threaten validity as these have already been discussed in Chapter 2. These include factors such as the lack of assessment literacy professional development, teachers' jobs demands and the structure of the education system.

The MTALQ is designed to measure teachers' assessment literacy. It is not a high-stakes test. The MTALQ assesses the seven standards that relate Standard 10 of the NPST. These standards all fall within the field of educational measurement, assessment, and evaluation, meaning that they form a single unitary structure. This unitary structure was made explicit in the original TALQ and the MTALQ and captures the target domain of 'assessment literacy'.

As explained in the first, second and third chapters, the test results are not generalisable in other contexts as the test participants are only a purposive sample of KSA teachers. Therefore, the validity of the test results and the instrument is only applicable to that group of teachers who agreed to respond to the questionnaire in the KSA educational context.

It could also be said that the case for validity is reflected in the various steps undertaken in the study. To recap, first, a set of elaborations was developed that led to an assessment literacy framework that articulated the requirements of assessment literacy in relation to the NPTS. Second, focus groups were assembled with the aim of aligning the TALQ with the elaborations and the refined assessment literacy framework developed. Teachers in the focus group opted to retain most of the items, but there were others that were rejected and consequently further items were developed and reviewed for inclusion in the MTALQ. Third, based on all this information a new version of the TALQ (called the MTALQ) was developed, which included some of the

original TALQ items, some of the original TALQ items as modified by the focus groups and some new items were written.

The inductive aspect of participant responses is another focus when generating evidence to support the validity of the MTALQ. These formative response process (i.e., the teachers cognitive process) seeks to elicit information as to why teachers chose the incorrect option in the MTALQ over the correct one. For example, in building the assessment literacy framework the participants identified formative assessment as an important aspect of assessment literacy although they did not seem to fully understand what it is. In terms of validity, it must be said here that, the sources of validity evidence stem from the focus group discussions, the survey and the MTALQ results albeit not in the same capacity. This is because while the MTALQ results are one source of evidence about the validity of the instrument, there are other multiple sources evidence (focus group discussions and the survey results) that inform the validity of the MTALQ before it is administered.

As this study aims to verify the construction processes and the validity of the instrument more than the results themselves, the proposed interpretative framework and associated interpretation are used as a method of predicting reliability weaknesses (Messick, 1995). It is understood that there may be other factors that may have influenced teachers' choices in their responses to the questionnaire that are not related to general construction of the MTALQ and its validity. These factors, raised by teachers during the discussion, include the general philosophy of assessment, job requirements, general educational policies, and others.

The next section addresses study question 3.

6.5 Research Question 3: Validity of the MTALQ for Assessing the NPST.

RQ3: What evidence is available to support the validity of a new questionnaire (MTALQ) based upon the original TALQ, but augmented by items that better align it with the requirements of the KSA assessment literacy framework?

As explained earlier in this chapter the focus group engagements led to several modifications of the original TALQ questionnaire to make it more content aligned and fit-for-purpose in the KSA context. First, items that were not aligned (as judged by the focus groups participants) were eliminated from the TALQ. Second, some wording modifications to remaining items (stems and distractors) were made to enhance the likelihood that the items would resonate with KSA educators.

Finally, new items were developed to represent the fundamental aspect of assessment that the participants in the study felt so strongly about. Such a representation was crucial to develop an instrument which was tailored to the KSA context. Hence, the MTALQ, which was revised in response to KSA teachers' views, represents the final measurement instrument used to ascertain assessment literacy as described in the NPST. All in all, and as reported in the previous chapter, nine TALQ items were deleted for various reasons to do with alignment. Also 10 new items were created and reviewed and agreed upon by the participants. These new items were constructed to permeate a range of assessment literacy elaborations within the refined assessment literacy framework and the NPST. Because of all these factors, the MTALQ resulted in 36 test items.

The writing process described in chapter 3 and 4 included participants' suggestions to incorporate new items that measured specific skills in the refined assessment literacy framework and the content for each new item. Furthermore, the supervisors also reviewed the new items and refined them so that they would be better measures of the NPST. Following this, the new items

were reviewed and discussed again with the focus group participants. The current study utilised teachers' feedback and suggestions as a guide to write new items in which participants raised many issues related to formative assessment and its difference from summative assessment. Importantly, this process led to the inclusion of items that measure teacher assessment literacy (see Table 6.4).

In addition, in terms of the Standards that were included in the MTALQ, the idea of the new Standard, that is, the seventh Standard in the MTALQ to measure Standard 10.3 in NPST, was founded on building a new Standard that contributed to supporting the validity of the construction of the test through its connection with all other standards, either explicitly or implicitly. At the same time, the items built to assess the seventh standard in the MTALQ were designed to work in an overlapping and interdependent way to enhance the validity of the construction of the MTALQ. From the analysis of the focus group discussions with KSA teachers to validate the MTALQ, it became evident that the teachers stressed that most test items might appear to be difficult for teachers to answer correctly. This, in their opinion would be due to the low assessment literacy of KSA teachers. Their opinion was supported by what the OECD identified in their 2020 report (OECD, 2020).

However, the findings of this study showed that most of the new items in the MTALQ functioned more consistently with expectation for the sample of KSA teachers who responded to the questionnaire than the items in the original TALQ. This was the case particularly in relation to corrected item-to-total score correlation as well as Cronbach's alpha if an item was deleted, and hence the contribution of the item to overall test reliability was positive. For example, all items in Standard 7 in MTALQ (Standard 7 being the new standard in MTALQ) were endorsed by teacher participants to be included in MTALQ and were subsequently

reviewed by teachers and experts. This demonstrates that content and construct validity measures matched skills in the refined assessment literacy framework. That Standard 7 had the highest contribution to reliability and the best corrected item-to-total score correlation indicates that the processes and methods to establish validity were useful and effective strategies. Other researchers could consider them as valuable methods to conduct the validation process in their own studies.

The results from the survey suggest that teachers considered self- and peer-assessment as just ‘Desirable’ and the mean is close to ‘Not necessary’. Although this may be the case, the study findings show that teachers in KSA are not familiar with these concepts or the associated standard in the NPST about ‘Engaging students in the evaluation processes’. Focus group discussions with KSA participants showed that they felt that formative assessment and related strategies, such as self- and peer-assessment, might need further investigation in the KSA context because it appears that most teachers primarily employ summative assessment strategies in their regular classroom practices. Hence, some items were added to MTALQ related to standard 3 (Standard 10.3) in NPST (i.e., Engaging students in evaluation processes) so that this important skill is incorporated into the assessment literacy framework.

Because of the focus group discussions, a new item was (i.e., item 6 see Table 6.4) added to assess teachers’ knowledge and understanding of formative assessment, relating it directly to self- and peer-assessment. The analysis of the questionnaire for this item shows that most teachers chose option A. This selection can be explained by the fact that there is a difference between formative assessment and self- and peer-assessment which KSA teachers may not understand. This lack of knowledge is consistent with the OECD report (OECD, 2020) that highlight the low level of assessment literacy amongst teachers in KSA when it comes to

formative assessment. The Corrected Item-To-Total Score Correlation is positive which means that the item discriminates between those teachers who scored higher on the assessment literacy construct than those who scored lower. The reliability of the MTALQ if the item is deleted, decreases very slightly.

Table 6.4

Items 6 in the MTALQ, Corrected Item-To-Total Score Correlation/Cronbach's Alpha if Item Deleted, and Elaborations in the Refined Assessment Literacy Framework

NPST	Item in the MALQ	Corrected item-to-total score correlation / Cronbach's alpha if item deleted	Elaborations in the refined Assessment Literacy framework
10.1 Preparing evaluation tools	Item 6 A principal wants to assess a ninth-grade teacher's ability to use 'formative' assessment. The teacher uses peer and self-assessment as part of the set of 'formative assessments. The principal should make it clear to the teacher that: ...	0.028 / 0.550	• Know and understand the requirements, structure, purpose and action of conducting formative assessment appropriately.

To investigate the MTALQ validity, the reliability of a test must also be discussed (Kane, 2016). As the previous chapter outlined, the study data indicate that the reliability of the MTALQ is in fact slightly higher than that of the original TALQ. The KR20 reliability estimate

of the original TALQ was 0.54, whereas the reliability of the MTALQ, using Coefficient Alpha,¹ was 0.56.

Given that the MTALQ was modified from its original form for use in this study and that it was applied to a selection of KSA teachers who could be considered as roughly representative of KSA teachers in general, it was pleasing that the internal consistency increased, although it is considered still to be much too low. If these reliability estimates were ‘test-retest’ correlations, they would explain less than one-third of the variance in results across the two tests. What this essentially means is that, if given again without changes in context, the likelihood of producing similar results would be quite low, which would be a troubling finding. Despite this, the apparent revision did no harm to the construct in the instrument it intended to measure, as originally developed.

For the MTALQ, 13 of the 36 test questions had item-to-total correlations of less than 0.01. With nearly a third of the items having such low internal agreement, it would be likely that they were greatly suppressing the overall reliability estimate. For Standard 1 in the MTALQ this could be partially explained by the large variances in item difficulty, with one item having a large p-value of 92.8 (a very easy item) with a low item-to-total correlation of only 0.057. But another item for Standard 1 was very difficult, with a p-value of 14.8. Extreme p-values are known to suppress item-to-total correlations owing to a restricted item variance. Furthermore, Standard 6 had three item-to-total correlations which were negative, with the only positive item showing a correlation of 0.063. This was most likely attributed to their extreme difficulty.

¹ For multiple choice assessments like the TALQ and revised TALQ and without missing data, the expectation is that Coefficient Alpha and KR20 should be almost identical estimates of internal consistency reliability.

It could be said that the relatively low internal consistency measure of reliability was a concern given the overlap in content found within the original TALQ. For example, Item 2 and Item 7 both measured the ability of teachers to know and understand the concept of ‘reliability’ as it relates to the choice of assessment methods and development of assessment methods, respectively.

Item 2 asked:

When scores from a standardized test are said to be reliable, what does it imply?

- a. Student scores from the test can be used for a large number of educational decisions.
- b. If a student retook the same test, he or she would get a similar score on each retake.
- c. The test score is a more valid measure than teacher judgments.
- d. The test score accurately reflects the content of instruction in classes where the test is administered

Item 7 asked:

Which of the following actions would most likely increase the reliability of Mrs.

Lockwood's multiple-choice end-of-unit examination in physical science?

- a. Use a blueprint to develop the test questions.
- b. Change the test format to true-false questions.
- c. Add more items like those already in the test.
- d. Add an essay component.

The teachers who participated in the assessment literacy framework’s validation processes evaluated the elaborations pertaining to reliability as fundamental abilities, and they

stressed that they were ‘extremely familiar with the notions of ‘validity and reliability’’ in the follow-up discussion session. Yet only 52.9% of the participants scored Item 2 correctly and an even lower number (8.2%) scored Item 7 correctly. Participants who were involved in the process of eliminating some items from the original TALQ advocated for the retention of both items because they were of the view that the items measured the reliability skills in the refined assessment literacy framework and related standards in the NPST (see Table 6.5). This finding is encouraging as it implies that the participants judged the importance of the items and their alignment to the NPST independently of their perception of performance.

A discussion led here through a psychometric lens might require further investigation related to response processes (i.e., most teachers, 68.86%, chose an incorrect Option, A for item 2), which might call into question the validity of the item format and content. As seen in Table 6.5, this technique was developed by talking to teacher participants in the study and this process undoubtedly provides strong evidence for the validity of the instrument.

Furthermore, informed by the discussions with teachers during the focus group sessions, the MTALQ included a new standard of modern assessment that is appropriate for the KSA educational context and aligns well with the NPST. This inclusion was driven by the insights of a group of KSA teachers themselves. In the focus group discussions, they mentioned that some skills in the refined assessment literacy framework were considered essential skills and they therefore suggested that the new standard could be incorporated to measure these skills in the refined assessment literacy framework. Besides, the participants in the focus group worked to identify these skills to establish and construct validity and make appropriate connections between their feedback and the quality of the content validity. The new items in the new standard achieved psychometric metrics results higher than items accepted by the teachers in the original

TALQ, which also proved to be a more reliable and effective process to increase the validity of the instrument.

Table 6.5

Items 2 and 8 in the MTALQ, Corrected Item-to-total Score Correlation/Cronbach's Alpha if Item Deleted, and Elaborations in the Refined Assessment Literacy Framework

NPST	Item in the MTALQ	Corrected item-to-total score correlation / Cronbach's alpha if item deleted	Elaborations in the Refined Assessment Literacy framework
10.1 Preparing evaluation tools	Item 2: When scores from a standardised test are said to be reliable, what does it imply?	0.033 / 0.540	• Know and understand how to develop assessment items and tasks in order to ensure high quality in their validity and reliability • Describe the importance of validity and reliability in classroom assessment
	Item 8: Which of the following actions would most likely increase the reliability of Mrs Lockwood's multiple-choice end-of-unit examination in physical science?	0.001 / 0.571	• Know the features of the range of assessment forms in relation of its strengths and weaknesses. • Know and understand how to improve items and tasks that have been written for assessing achievement and learning outcomes • Know and understand the strengths and weaknesses of a range of forms of currently used assessments, such as multiple-choice questions, observation and portfolio

Since this study began with an existing measure of assessment literacy (TALQ) as a starting point to develop a revised instrument with the aim of validating its use (or validating its modification) in the KSA, comparisons of its properties with other similar measures are important. To achieve this, a literature review of the instruments based on the original TALQ was undertaken. The original TALQ worked as a source for ALI (Campbell et al., 2002), CALI (the Classroom Assessment Literacy Inventory; Mertler, 2003) and the second version of ALI (Mertler & Campbell, 2005) to provide a picture of comparison, Table 6.6 compares the different instruments to the original TALQ.

The internal structure is supported by a strong internal consistency reliability as well as large specific item-to-total correlations and a large contribution of each item-to-total test reliability. While it is recognised that this is not the same or as strong as linear modelling, factor analyses or structural modelling (given that the TALQ is an existing instrument), more detailed structural analyses were beyond the scope of this study.

Overall, there is relatively strong evidence of the validity of the MTALQ. However, this is the start of the process of building an evidential case for supporting the validity of the MTALQ.

Table 6.6*Comparison of TALQ and Similar Instruments' Reliability*

Instrument	Guiding framework	Reliability	Teacher category	Number of teacher participants
Original TALQ	1990 Standards for Teachers	0.54	in-service	555
ALI (Assessment Literacy Inventory)	1990 Standards for Teachers	0.74	pre-service	220
CALI (Classroom Assessment Literacy Inventory)	1990 Standards for Teachers	In-service = 0.57 Pre-service = 0.74	both in-service and pre-service	In-service = 197 Pre-service = 220
ALI (second version)	1990 Standards for Teachers	0.74	in-service	250
The MTALQ	NPST	0.56	in-service	291

6.6 Chapter Summary

This chapter provides a summary of the responses and data gathered from the study, with the primary objective of addressing the study's research questions. The study is centred around the development of a refined assessment literacy framework and instrument. The chapter outlines several key points. In terms of significance of the refined assessment literacy framework and the MTALQ. It emphasises the importance of the newly developed assessment literacy framework and instrument. These contributions are considered valuable not only in the context of education in the Kingdom of Saudi Arabia (KSA) but also in other educational settings.

The chapter also explains how the refined assessment literacy framework can be practically employed to make positive differences in teacher learning and education. It underscores the potential impact of this framework in enhancing teachers' understanding of assessment practices. In terms of validity and reliability of the instrument, the study presents evidence to demonstrate the validity and reliability of the assessment instrument that was created. The chapter emphasises that the processes undertaken during the study collectively establish a strong level of validity and reliability for the Modified Teachers' Assessment Literacy Questionnaire (MTALQ).

In addition, the chapter suggests that, due to its validity and reliability, the MTALQ can be a valuable tool in efforts and initiatives aimed at improving teacher assessment literacy in the KSA. This implies that the instrument can contribute to enhancing the quality of assessment practices and, by extension, the quality of education in the KSA.

In essence, this chapter sets the stage for the final chapter of the thesis, highlighting the importance and potential impact of the refined assessment literacy framework and the MTALQ. It underscores their value in improving teacher assessment literacy and, consequently, educational quality in the KSA and beyond.

The next and final chapter will provide a summary of the study's contributions to knowledge and discuss its limitations. It serves as a conclusion of this thesis where the overall findings and insights of the study will be discussed.

Chapter 7: Conclusion

7.1 Introduction

The original overarching aims of this thesis were to develop a framework of assessment literacy for KSA teachers that operationalises the NPST and validate a teacher assessment literacy questionnaire that can reliably measure the assessment literacy of teachers in schools in the KSA education system. This final chapter will summarise the results of the study, discuss the study recommendations that are informed by the study findings, highlight the contributions the study makes, suggest areas for future research, and acknowledge the limitations of the study.

7.2 Summary of Results

After the analysis of the data and assessment of the results, as discussed in Chapters 4 and 5, this study found that there was a critical need for a valid assessment literacy framework that can provide the basis for delineating the NPST into elaborations and at the same time help teachers meet the teaching standards in the KSA. These elaborations and the refined assessment literacy framework will allow for more directed feedback and engagement with teachers for improved teaching and learning. The refined assessment literacy framework provides the teachers with a road map regarding how to apply their knowledge and skills in assessment in a practical way. Because the refined assessment framework is inextricably linked to the NPST, it is also a guide for teachers towards targeted professional learning that will enable them to move from where they are now in terms of assessment literacy to where they would like to be.

The refined assessment literacy framework and the associated assessment literacy instrument (the MTALQ) were not designed to be global tools, but rather were developed for specific application in the KSA, because educational assessment and evaluation processes are dependent on various cultural, economic, and other factors. In making this point, however, it

is also important to reinforce that the procedure that was developed in this study is generalisable to all countries that have professional teacher standards that include assessment and data literacy and would like to produce a measure to answer the question, ‘How assessment literate are our teachers at this time?’ This question must be answered to produce baseline measures that can be used to ascertain the extent to which professional learning interventions will have a positive impact on the assessment literacy of their teachers. The required knowledge and skills in the refined assessment literacy framework are complementary to one another and they shouldn't be considered as separate skills. This is something about which professional learning program developers should be aware because educational assessment is an integrative process. As a result, professional learning programs using the refined assessment literacy framework will hopefully lead to the improvement of all assessment skills required by KSA teachers.

It is important to mention that the refined assessment literacy framework is intended to incorporate most of the skills related to educational assessment, but not all. In different contexts, the skills can be reduced to merely include basic skills, and in others the assessment literacy framework can be developed to include skills needed by expert teachers. Finally, this study does not imply that the refined assessment literacy framework is a ‘magical solution’ for raising the level of teachers’ understanding of educational measurement and assessment literacy. Rather, the refined assessment literacy framework is intended as an additional ‘auxiliary tool’ that helps - via its elaborations - to make the NPST standards related to assessment literacy more consumable and more understandable for the teachers in the KSA. Use of the refined assessment literacy framework in a judicious manner will likely lead to improved performance by the KSA teachers as measured by the MTALQ over time.

7.3 Implications and Recommendations of the Study

Findings from this study have led to the development of a refined assessment literacy framework that makes explicit what is required in terms of assessment and data literacy by the KSA's National Professional Standards for Teachers; and the development of an instrument that can help measure the level of assessment literacy KSA teachers currently possess. As a result of this work, the study has also been able to provide some preliminary insights into the current state of teacher assessment literacy in KSA, based on a small non-generalisable sample of KSA teachers. To this end, the next two sections discuss some of the major implications of these insights as well as the recommendations for the improvement of assessment literacy within the education system of KSA in general.

7.4 Assessment Literacy: A Vital Cog in the Efficacy of the KSA

Educational System

It is evident that recent educational reforms in the KSA have demonstrated the Government's commitment to improving the educational system in the country. In the KSA's 'Vision 2030' (2016), education is seen as a key instrument in transforming the country in the 21st century. This is because it is perceived as an integral and vital part of government initiatives that aim to pivot the Saudi Arabian economy's dependence on oil towards the development of a knowledge-based economy—an economy that has been argued to be a more sustainable pathway to economic growth and progress in the country (Bashehab & Buddhapriya, 2013).

However, given that recent years have seen numerous educational reforms and initiatives being launched, the education system in the KSA is at a critical juncture. For instance, a result of the KSA government's policies has been an increase in the universal access to education across a highly dispersed and vast school-age population. However, the attempts to provide high-quality education have come at an expense, as it appears that KSA

teachers and education administrators are being pushed to their limits. Such a situation implies that for teachers, there is little time to attend to the professional learning programs designed to target key areas for improvement, including, as argued in this study, assessment literacy. Thus, even if there are programs that are designed to support and develop KSA teachers' competency in assessment literacy, there is a likelihood that teachers may not have the 'bandwidth' nor the time to attend.

The Education, Training Evaluation Commission (ETEC) is an important body in the KSA's educational ecosystem as it has been established to specialise in assessment, measurement and evaluation. It was ETEC's partnership with King Abdullah bin Abdulaziz's Public Education Development Project ('Tatweer') that led to the articulation of the NPST in the KSA, in the first place. To reiterate, these standards encompass three main areas: (1) 'Professional Knowledge', (2) 'Professional Practices' and (3) 'Professional Values and Responsibilities'. ETEC is a primary driver of the creation and establishment of the NPST. Assessment comes under area two of the NPST, 'Professional Practices'. Accordingly, it is recommended that they partner with professional development agencies to develop professional learning programs for assessment literacy. In addition, the refined assessment literacy framework and the results from its associated MTALQ could be used to increase the probability that professional learning programs can be better tailored to teachers needs at the different locations on the assessment literacy progression defining improved assessment literacy skill.

The KSA's 'Vision 2030' (2016), which is a landmark national initiative, also promised substantial reforms to the entire education system. In 2019, the KSA's MOE reformed the structure of the standards of public education and the curricula used in public educational institutions. As discussed, such changes were aimed at raising KSA's educational system to be in line with other progressive educational systems around the world. The

changes included reforming assessment practices to create a more assessment-based culture in the educational system (Assalahi, 2021). KSA's reforms in its educational assessment regimes underscore the importance that the Government places on assessment as part of raising the quality of education in the KSA. However, while this seems to be the case, it is unclear what the MOE in the KSA or a body under its jurisdiction, such as ETEC, has committed to in terms of supporting KSA teachers' professional learning in assessment literacy. They could adopt the refined assessment literacy framework and the MTALQ developed in this study and utilise them to improve teacher training and proficiency in assessment literacy in much the same way as advocated in the study.

This potential lack of coordination between policy and its enablers has been suggested by an OECD (2020) report on KSA education as a reason for the system not reaching its full potential. For instance, it indicates that while the various sub-units under ETEC are tasked with the development of resources to aid teachers, close collaboration among them is lacking. The report also suggests that there are clear avenues for the various units under ETEC to move more in concert in terms of coordination and collaboration. For the KSA Government to act on the OECD's recommendations for quality of teaching in schools, it will be essential that teachers are supported in assessment literacy. This is because assessment is key in raising the quality of an educational system such as the KSA.

However, the challenge of realising KSA's overriding reform mandate is the fact that there seems to be a lack of articulation of its vision of quality schooling (OECD, 2020). The absence of a vision means that there is no collective reference for the various agents and stakeholders in the KSA education system that are working towards the aim of enhancing the quality of schooling. The findings of the present study implies (from a small non-generalisable purposive) that KSA teachers are lacking in assessment literacy and have limited skills and knowledge in the area. Therefore, notions such as teacher competencies in

areas that include assessment should form part of a clear vision of schooling. Moreover, the current study not only supports the KSA government acting on OECD's recommendation (OECD, 2020), focused on improved student outcomes and how an effective school develops their students' cognitive, emotional, and social dimensions, it provides a tool which would be essential to determining the current effectiveness of the teacher and data literacy of KSA teachers. There is little doubt that to achieve the broader OECD aim (OECD, 2020) teachers must be able to broaden their knowledge and skills to utilise more formative assessment into their everyday teaching, assess higher order thinking skills, etc. because these are all related to improving student assessment in the classroom (Black & Wiliam, 1998; Hattie, 2008).

OECD (2020) also highlighted that the National Transformation Program (NTP) has not assisted schools in comprehending and attaining student achievement goals, even though it is the entity that stipulates student attainment goals on a national level. What this implies is that the NTP has also not been able to help schools in understanding and mobilising the aim of schooling under Vision 2030. For this reason, Directorates and Education Offices do not have reference points by which to prioritise their actions when working with schools and, consequently, each Directorate lists its own priorities to define the key elements of effective schooling. The refined assessment literacy framework and MTALQ developed through this study gives them 'common ground' to start standardising an approach to improved teacher assessment literacy.

Before the study was completed, it was found that several KSA education stakeholders were of the view that, in general, KSA teachers had weak assessment literacy (OECD, 2020). The study findings support this, even if it is based on a small, non-generalisable sample. Given that assessment effectively supports learning better than any other intervention (Black & Wiliam, 1998; Hattie, 2008) it is very important that KSA teacher literacy programs should also be consistent with the key principles of effective

assessment requirements that are based on the NPST and the refined assessment literacy framework. Reforms to teacher assessment literacy and professional learning programs on assessment and data literacy in the KSA should be designed around the key principles of effective assessment. For instance, the Assessment Reform Group (1999, p. 7) suggests the elements that inform effective teacher assessment literacy reforms and professional learning. These are:

1. AFL is embedded in curriculum and assessment institutionally and pedagogically;
2. assessment reform goals are explicitly shared with all stakeholders, and stakeholders together work to identify how to know and to recognise the standards they are aiming for;
3. constructive qualitative feedback helps stakeholders to recognise the next steps needed for reform and how to take them;
4. all stakeholders regularly review and reflect on assessment data;
5. it is assumed every school, teacher and student can improve.

In applying the above precepts of assessment to the context of KSA, this would mean the following. First, effective teacher assessment literacy reforms and professional learning should consider where teachers are in terms of their assessment literacy competence, just as effective student learning is formative. The refined assessment literacy framework and MTALQ developed in this study can provide support for this. Effective assessment literacy reforms and professional learning that aim to develop KSA teachers' assessment literacy competence must be informed by understandings of their existing competence before designing programs that meet their various needs. To this end, the findings of this study can significantly contribute to such understandings as this study has provided some evidence that the MTALQ is a relatively valid and reliable questionnaire to assess teachers' assessment

literacy in the KSA and provide a baseline for measuring the effectiveness of any interventions used to improve teacher assessment literacy.

Second, the goals of assessment reforms must be clearly communicated with all stakeholders in the educational system so that there is a common understanding about assessment standards and goals that exists among all stakeholders. In this regard, it is argued that the Ministry of Education (MOE) or ETEC ought to lead by putting into place system-wide initiatives to establish common understanding of how to go about knowing and recognising the benchmarks these organisations aspire towards, and how their teachers' assessment literacy plays a role in such achievements. Again, the refined assessment literacy framework and the MTALQ, developed in this study, could help teachers, leaders and policy makers in understanding these benchmarks.

Third, it is critical that qualitative feedback is instrumental to stakeholders' understanding of identifying what is required to effect change and how to go about implementing what is needed to improve assessment literacy. Fourth, consistent review and analysis of assessment data. Fifth, as a tenet of formative assessment, there must be an explicit assumption that the key stakeholders such as the school, teacher and student can make progress in understanding and using it in the classroom.

Ultimately, the efficacy of a professional program in teacher assessment literacy starts with a clear understanding of the basic principles of effective assessment by stakeholders. In the case of the KSA's educational context, this implies an important role for ETEC in coordinating such understandings and principles and implementing them so that the requirements imbedded in the refined assessment framework can be executed in harmony with all levels of education. The current refined assessment literacy framework and instrument developed in this study could be used to help all five of these reforms.

It is unsurprising that the OECD (2020) review on education in KSA found that there is currently no assessment framework that defines assessment activity at a national level. The lack of a clear framework could explain why the OECD found that at the level of the classroom, teacher assessment is predominantly summative. Moreover, assessment materials strongly feature memorisation and closed-ended item types, which limit critical thinking. The OECD review recommended that to strengthen classroom assessments, teacher appraisal and school evaluation should be used to promote more diverse - especially formative - assessment practices. Such findings demonstrate the need for professional learning programs in teacher assessment literacy in the KSA. The study outcomes such as the refined assessment literacy framework and MTALQ, which have been developed and validated in this study could play a key role in the KSA moving forward in the direction of the ‘SAUDI VISION 2030’.

7.5 Contributions of the Study

The present study has led to several potentially significant contributions. The refined assessment literacy framework and the MTALQ are argued as invaluable tools very beneficial to the KSA and, the process can be generalised to other contexts where systems are searching for answers to the question like ‘How assessment and data literate are our teachers?’. The current refined assessment literacy framework could be utilised as a specification table to create customised training programs in educational assessment. It is preferable to have professional learning course developers integrate all skills in the refined assessment literacy framework into the training programs in a considered way such that foundational learning (i.e., learning how to learn) can take place. It may be possible to concentrate more on the abilities identified as urgently needed by the study in the early stages with more sustainable training coming online later.

This study also demonstrates the usefulness of employing Classical Test Theory (CTT) in educational studies by comparing the results from this study with other, previously

published outcomes. Because no such framework existed previously in the KSA context, one of the extremely beneficial applications of this study is using the refined assessment literacy framework during moderation and comparability community practices in schools. Of course, the refined framework may be less useful if these strategies are not employed correctly.

Furthermore, the current refined assessment literacy framework may contribute to unifying views on assessment literacy at all levels- micro as well as macro. Issues that are conflated with this aspect of assessment literacy are described by McMillan (2003) and include the tension between teachers' beliefs, values and external factors, such as macro policies, parents and accountabilities related testing. In the KSA context, such issues are argued to be more influential because teacher supervisors and school principals assess and use students' results as part of the appraisal of a teacher's work performance (OECD, 2020). Such a practice could be argued to hinder the confidence of teachers which, in turn, stops them from being innovative in applying various assessment methods and to align their assessments more effectively with the requirements of the curriculum and make better informed decisions about student learning.

Even though the refined assessment literacy framework was developed for the needs of the mainstream classroom, there is no reason it cannot be applied to teaching and assessing in special needs areas like teaching gifted and talented students or those learners with diverse and special needs.

The MTALQ that was developed is also the first in the KSA context to measure assessment literacy as defined by the NPST, and it could evolve while the educational system evolves, making it the *ex-post facto* measure for assessment literacy going forward. Alternatively, it may inspire stakeholders to develop alternative or even similar instruments, whether they be based on a philosophical, theoretical or other underpinnings. This is a vital area to improve because of its expected potential implications for the development and

improvement of teachers' assessment literacy and the role this plays in raising the quality of education in the KSA.

7.6 Areas for Future Research

It is possible for future researchers to use the refined assessment literacy framework and the MTALQ to empirically measure the impact of assessment literacy training programs. For example, the refined assessment literacy framework is presented to teachers to test their assessment literacy before and after training programs, measuring, for example, the relationship between the refined assessment literacy framework and the standards. It will also be possible to use other psychometric methods to discover the characteristics of the MTALQ and its development. Future researchers could add an option in the Likert survey that provides an opportunity for teachers who wish to voice their opinion about the best instrument that is consistent with the current refined assessment literacy framework, NPST and the Saudi educational context by including these instruments in the Likert survey.

In addition, future researchers might like to consider calibration of the instrument using more modern, and applicable measurement theories like Rasch Theory.

The current study should inspire future researchers to build standards-based frameworks and instruments to assist teachers in achieving standards in areas such as data literacy, teaching methods and pedagogy. This study could be the catalyst for future research focused on measuring the impact of the refined assessment literacy framework in influencing the attitudes and beliefs of the teachers who use it, bearing in mind that great knowledge in a specific field may contribute positively to a person's beliefs, attitudes, and tendencies towards the desired measurement implementation.

7.7 Limitations of the Study

This study used a purposive non-generalisable sample to make indicative conclusions about the assessment literacy of KSA teachers. Drawing conclusions from such a small

sample selected in this way is not defensible. That is why this is stressed throughout the study. This is indeed a limitation of the study. However, it is a necessary one, given the scope of this study was a PhD thesis. Selecting a random sample and getting Government approval to use such a sample (which would have required significant intrusion into school programs) would have been beyond the scope of the study. Drawing conclusions from the results of the study about the level of assessment literacy of KSA teachers was therefore done in an indicative manner to demonstrate how the results could be interpreted and used to make policy decisions. The value of this thesis is that it demonstrates a process for taking standards that describe increasing levels of teaching effectiveness, the KSA National Professional Teaching Standards (NPST) and delineating them into a coherent set of elaborations that can be used to build an assessment literacy framework; evaluating the utility of an assessment literacy questionnaire that has been designed for a different context; and, then adapting it for this new context by designing, constructing and validating the adapted questionnaire so that it can be used confidently in a new context.

A second limitation is that the process after the initial delineation of the standards into elaborations involved using teacher participants in the focus groups, inviting them to make decisions about the relative importance of each of the elaborations for demonstrating the standards, and having teachers complete the MTALQ to generate validity evidence to support the MTALQ and draw indicative conclusions about the quality of teacher assessment literacy in the KSA. The results demonstrate that the teachers who completed the MTALQ were not very competent in assessment literacy. Some of these same teachers were the ones who had participated in the focus groups and survey. It could be inferred that preliminary decisions upon which the whole study was built were made by teachers who had limited assessment literacy knowledge and skill. This is a limitation, but it also points out the need to have measures of any construct, because if the question cannot be answered regarding how

competent teachers (in this study) are in relation to the construct of assessment literacy, how can appropriate samples of participants be used in studies such as the one carried out in this PhD thesis?

A third limitation is that Classical Test Theory (CTT) was used to produce the evidence to support the validity of the MTALQ. While such information is appropriate for this purpose, the fact that the MTALQ is potentially going to be used as a measurement tool to monitor assessment literacy over time and could involve increasing the number of items in the MTALQ or generating different versions of the MTALQ, means that a measurement model (the Rasch Model) should have been used to govern the construction of the MTALQ. In this case, the results would be very similar because there were a fixed set of items and the study focused on the alignment between the items and the elaborations and standards, the reliability of the single questionnaire. The conclusions would be the same, given the nature of this study.

7.8 Concluding Remarks

The KSA government's major reforms, not only in education but also in other domains such as the economy, indicate that education is the key in pivoting KSA's economic trajectory to one that is more sustainable, and weaning it off from its almost singular reliance on oil. The KSA government's initiatives towards the liberalisation and diversification of its economy can already be seen in the 'easing up' of its traditionally strict societal norms (e.g., opening of cinemas, allowing music concerts and females to drive) and the hosting of international sports events.

However, while such pivots have their benefits in realising aspects of KSA's Vision 2030, it is argued that at the core of it all remains the people of the KSA. Developing the KSA's talents and human capital will be crucial in ensuring that the nation can thrive long into the future without significant reliance on outside expertise. And as the economic success

of other countries has shown, the quality of education that is provided to a society is the key to its economic and social transformation. For instance, Singapore's success story in developing itself from a third world country to a first world one in a single generation has been described as an 'economic miracle'. However, a closer inspection reveals that this transformation is no miracle. Rather, it has been the aggregate result of sound policies, good governance and, ultimately, the quality of its education, which have transformed the country. Thus, for the KSA, which is a country that has far greater resources than a country like Singapore, the potential is infinitely more valuable.

Having said that, the KSA's situation in terms its educational outcomes cannot be on a trajectory of success given the issues that continue to beset it. The government still struggles to attract quality candidates and its current teachers appear to be lacking in professional support and development. All these issues can be said to be manifested in KSA students lagging other students in other countries in international testing results. For all the investment that the government puts into education, its returns can be described as rather poor.

It is argued that at the heart of the issues that beset KSA's education system is the quality of its teachers and the support they access, be it financially or professionally. Much research has shown that the quality of teachers is the single most impactful variable in determining student learning outcomes (Hattie, 2015, p. 2, 2017). Research also shows that assessment practices are key in effective teaching and learning (Tognolini, 2020; Tognolini & Stanley, 2007). A teacher's assessment literacy is a vital factor in the complex matrix of skills and knowledge that determines the quality of teachers.

This study does not make the claim about the quality of teacher assessment literacy, it has provided a roadmap to raise the quality of teachers and KSA education. But some of its significant outcomes have the clear potential to play a transformative role when it comes to

raising the level of KSA teachers' assessment literacy. This study was driven by research gaps in the KSA, such as the absence of measurements of teacher assessment literacy despite the recognised importance of teacher assessment literacy in education systems around the world. It was also driven by a commitment to find solutions to such critical gaps in a manner that could benefit the KSA.

Now at its completion, this study has provided a refined assessment literacy framework for teacher assessment literacy, and an instrument to measure the competence of teacher assessment literacy in the context of KSA - that is, the MTALQ. Considering findings from this study, as well as those in the OECD (2020) report that indicate that KSA teachers have low levels of assessment literacy, the current refined assessment literacy framework has the potential to play a transformative role in effecting successful educational reforms. Moreover, as this study's findings indicate, the instrument that is the MTALQ can measure the level of teachers' assessment literacy in the KSA. The development of a refined and more comprehensive assessment literacy framework and an instrument to measure KSA teachers' assessment literacy is akin to making incursions on uncharted grounds in KSA's educational landscape. This is because prior to this study, no such assessment literacy framework or instrument that was fully tailored to the KSA educational context existed.

What both the refined assessment literacy framework and MTALQ now mean is that KSA teachers are now able to have access to a more comprehensive framework of assessment literacy that has especially been tailored to the KSA educational context. The refined assessment literacy framework can be a common and unifying reference point for KSA teachers to reflect on as well as have professional dialogues with colleagues on assessment skills and knowledge. Complementing this is the instrument, which can play an essential role in diagnosing gaps in teachers' knowledge and skills so that professional learning programs can be targeted and structured. This importantly means that the results of the MTALQ along

with the refined assessment literacy framework can be the means to personalise professional learning programs for KSA teachers, which are more effective than professional development of the general variety.

Furthermore, by using the refined assessment literacy framework, teachers in the KSA can move away from merely summative assessment practices and measuring students' memorisation to more advanced assessment practices that could transform the way teachers teach while also developing students' 21st-century skills and knowledge, such as high order thinking skills.

From a policy perspective, the refined assessment literacy framework and the MTALQ can in tandem provide a more systematic means for policy makers and other stakeholders to connect teachers' standards, teachers' skills and knowledge, professional learning and student learning. In addition, it is expected that the outcomes of this study can pave the way for a move to a standard referenced assessment system. This would indeed be a crucial positive move in the KSA context.

Finally, as a KSA national and an educator, I take the view that recent educational reforms in the KSA can be seen as a form of negotiation: a professional compact as well as a dialogue between policy makers on the one hand, and those which they intend to effect on the other - that is, teachers. The outcomes of this study are not only about attempting to contribute to efforts to raise the quality of education in the KSA. Crucially, too, they are about enriching, deepening and extending the dialogue that has already begun in earnest in the shape of the landmark reforms encapsulated in Saudi Arabia's 'Vision 2030'.

References

- Adie, L. (2013). The development of teacher assessment identity through participation in online moderation. *Assessment in Education: Principles, Policy & Practice*, 20(1), 91–106.
- Alafaleq, M., & Fan, L. (2014). *A new assessment approach in mathematics classrooms in Saudi Arabia* [Paper presentation]. 8th British Congress of Mathematics Education, University of Nottingham, United Kingdom.
- Alamri, M. (2011). Higher education in Saudi Arabia. *Journal of Higher Education Theory and Practice*, 11(4), 88–91.
- Alanazi, F. H. (2017). Effectiveness of the proposed training formative assessment programme and its impact on teaching style improvements of Saudi science teachers in Saudi Arabia. *Journal of Turkish Science Education*, 14(1), 35–56.
- Albalawi, A. M. (2018). The beliefs of mathematics teachers in Riyadh about formative assessment. *International Journal of Educational and Psychological Studies*, 4(1), 42–56.
- Alhawsawi, S. (2013). *Investigating student experiences of learning English as a foreign language in a preparatory programme in a Saudi university* [Unpublished doctoral dissertation]. University of Sussex.
- Al-Johani, H. M. (2009). *Finding a way forward: The impact of teachers' strategies, beliefs and knowledge on teaching English as a foreign language in Saudi Arabia* [Unpublished doctoral dissertation]. University of Strathclyde.
- Alkharashi, M. A. (2012). The oil economy of Saudi Arabia. *Allied Academies International Conference. Academy for Economics and Economic Education. Proceedings*, 15(2), 27–30.

- Alkharusi, H. (2011). Psychometric properties of the teacher assessment literacy questionnaire for preservice teachers in Oman. *Procedia –Social and Behavioral Sciences*, 29, 1614–1624.
- Almuntasheri, S. (2016). Saudi teachers' practices of formative assessment: A qualitative study. *Problems of Education in the 21st Century*, 74(6), 6–15.
http://www.scientiasocialis.lt/pec/files/pdf/vol74/6-15.Almuntasheri_Vol.74_PEC.pdf
- Almutairi, N. H. (2008). *The influence of educational and sociocultural factors on the learning styles and strategies of female students in Saudi Arabia* [Unpublished doctoral dissertation]. University of Leicester.
- Alnahdi, G. H. (2014). Educational change in Saudi Arabia. *Journal of International Education Research*, 10(1), 1–6.
- Alotabi, K. A. (2014). Student assessment strategies in Saudi Arabia: A case study of pre and post classroom practices. *Literacy Information and Computer Education Journal (LICEJ)*, 3(1), 1758–1763.
- Al-Otaibi, G. (2004). *Language learning strategy use among Saudi EFL students and its relationship to language proficiency level, gender and motivation* [Unpublished doctoral dissertation]. Indiana University of Pennsylvania.
- Alquraini, T. (2010). Special education in Saudi Arabia: Challenges, perspectives, future possibilities. *International Journal of Special Education*, 25(3), 139–147.
- Al-Sadan, I. A. (2000). Educational assessment in Saudi Arabian schools. *Assessment in Education: Principles, Policy & Practice*, 7(1), 143–155.
- Alsharif, K. M. (2011). *Towards quality teacher education: Productive pedagogies as a framework for Saudi pre-service teachers' training in mathematics education* [Unpublished doctoral dissertation]. Curtin University.

- Al-Shikh, K. (2017). Comparative study between students enrolled in science education diploma program at Jeddah in Saudi Arabia, and Iowa in the United States: universities in the implementation of skills lessons. *ASEP*, 81, 51–78
- Althobeti, O. (2018). The evaluation methods used by a sample of faculty members at Shakra University and their relation to the quality of the outputs of university education. *Journal of Education*, 51(51), 321–351.
https://edusohag.journals.ekb.eg/article_15237.html?lang=en
- American Educational Research Association, American Psychological Association, National Council on Measurement in Education, & Joint Committee on Standards for Educational and Psychological Testing (U.S.). (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- American Federation of Teachers, National Council on Measurement in Education, & National Education Association. (1990). *Standards for teacher competence in educational assessment of students*. ERIC.
<https://files.eric.ed.gov/fulltext/ED323186.pdf>
- Arksey, H., & Knight, P. (1999). *Interviewing for social scientists*. SAGE Publications.
- Assalahi, H. (2021). The professional standards for teachers in Saudi Arabia as a framework for EFL teachers' professional development needs assessment. *TESOL International Journal*, 16(6), 161–186.
- Assessment Reform Group. (1999). *Assessment for learning: Beyond the black box*. University of Cambridge School of Education
- Australian Institute of Teachers and School Leadership. (2018). *Australian professional standards for teachers*. <https://www.aitsl.edu.au/standards>

- Barnes, N., Fives, H., & Dacey, C. M. (2015). Teachers' beliefs about assessment. In H. Fives & M. G. Gill (Eds.), *International handbook of research on teacher beliefs* (pp. 284–300). Routledge.
- Bashehab, O. S., & Buddhapriya, S. (2013). Status of knowledge based economy in the Kingdom of Saudi Arabia: An analysis. *Journal of Social and Development Sciences*, 4(6), 268–277.
- Best, S., & Kahn, R. (2004). Trial by fire: The SHAC7, globalization, and the future of democracy. *Animal Liberation Philosophy and Policy Journal*, 2(2), 1–36.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policies and Practice*, 5(1), 7–74.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31.
<https://doi.org/10.1007/s11092-008-9068-5>
- Black, P., & Wiliam, D. (2010). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 92(1), 81–90.
- Boekaerts, M., & Corno, L. (2005). Self-regulation in the classroom: A perspective on assessment and intervention. *Applied Psychology*, 54(2), 199–231.
- Boren, T., & Ramey, J. (2000). Thinking aloud: Reconciling theory and practice. *IEEE Transactions on Professional Communication*, 43(3), 261–278.
- Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice*, 30(1), 3–12.
- Brookhart, S. M., & Durkin, D. T. (2003). Classroom assessment, student motivation, and achievement in high school social studies classes. *Applied Measurement in Education*, 16(1), 27–54. https://doi.org/10.1207/S15324818AME1601_2

- Brookhart, S. M., Walsh, J. M., & Zientarski, W. A. (2006). The dynamics of motivation and effort for classroom assessments in middle school science and social studies. *Applied Measurement in Education*, 19(2), 151–184.
https://doi.org/10.1207/s15324818ame1902_5
- Brown, G. T. L. (2008a). Assessment literacy training and teachers' conceptions of assessment. In C. Rubie-Davies & C. Rawlinson (Eds.), *Challenging thinking about teaching and learning* (pp. 285–302). Nova Science.
- Brown, G. T. L. (2008b). *Conceptions of assessment: Understanding what assessment means to teachers and students*. Nova Science.
- Burns, B., Filmer, D., & Patrinos, H. (2011). *Making schools work: New evidence on accountability reforms*. World Bank.
<http://documents.worldbank.org/curated/en/483221468147581570/Makingschools-work-new-evidence-on-accountability-reforms>
- Campbell, C., Murphy, J. A., & Holt, J. K. (2002). *Psychometric analysis of an assessment literacy instrument: Applicability to preservice teachers* [Paper presentation]. Annual Meeting of the Mid-Western Educational Research Association, Columbus, OH.
- Connelly, L. M. (2014). Ethical considerations in research studies. *MEDSURG Nursing*, 23(1), 54–56.
- Coombs, A. J., 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>
- Cowie, B., Cooper, B., & Ussher, B. (2014). Developing an identity as a teacher-assessor: Three student teacher case studies. *Assessment Matters*, 7, 64–89.

<https://www.nzcer.org.nz/nzcerpress/assessment-matters/articles/developing-identity-teacher-assessor-three-student-teacher>

- Davies, A. (2008). *Assessing academic English: Testing English proficiency 1950–1989—The IELTS solution*. UCLES/Cambridge University Press.
- Davison, C. (2013). Innovation in assessment: Common misconceptions and problems. In K. Hyland & L. Wong (Eds.), *Innovation and change in English language education* (pp. 263–275). Routledge.
- 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), 107–126.
- 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.
- DeLuca, C., Klinger, D., Pyper, J., & Woods, J. (2015). Instructional rounds as a professional learning model for systemic implementation of assessment for learning. *Assessment in Education: Principles, Policy & Practice*, 22(1), 122–139.
- 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.
- Deneen, C. C., & Brown, G. T. L. (2016). The impact of conceptions of assessment on assessment literacy in a teacher education program. *Cogent Education*, 3, 1–14.

- Education and Training Evaluation Commission. (2016). *Professional standards for teachers in Saudi Arabia*. <https://hdl.handle.net/1959.11/52216>
- Fryer, R. G., & Levitt, S. D. (2010). An empirical analysis of the gender gap in mathematics. *American Economic Journal: Applied Economics*, 2(2), 210–240.
<https://doi.org/10.1257/app.2.2.210>
- Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113–132. <https://doi.org/10.1080/15434303.2011.642041>
- Fullan, M. (2002). Principals as leaders in a culture of change. *Educational Leadership*, 59(8), 16–20.
- Fullan, M., & Watson, N. (2000). School-based management: Reconceptualizing to improve learning outcomes. *School Effectiveness and School Improvement*, 11, 453–473.
- 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, and Practice*, 22, 475–494.
<https://doi.org/10.1080/0969594X.2015.1017445>
- Gipps, C. (1999). Socio-cultural aspects of assessment. *Review of Research in Education*, 24, 355–392.
- Glaser, R. (1981). The future of testing: A research agenda for cognitive psychology and psychometrics. *American Psychologist*, 36(9), 923–936.
- Gotch, C. M., & French, B. F. (2014). A systematic review of assessment literacy measures. *Educational Measurement: Issues and Practice*, 33(2), 14–18.
- Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.
- Hattie, J. (2015). *What works best in education: The politics of collaborative expertise*. British Columbia Teachers' Federation.

- Hattie, J. (2017). *250+ influences on student achievement*. <https://visible-learning.org/wp-content/uploads/2018/03/VLPLUS-252-Influences-Hattie-ranking-DEC-2017.pdf>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112.
- Heitink, M. C., Van der Kleij, F. M., Veldkamp, B. P., Schildkamp, K., & Kippers, W. B. (2016). A systematic review of prerequisites for implementing assessment for learning in classroom practice. *Educational Research Review*, 17, 50–62.
- Hill, K., & McNamara, T. (2012). Developing a comprehensive, empirically based research framework for classroom-based assessment. *Language Testing*, 29(3), 395–420. <https://doi.org/10.1177/0265532211428317>
- Hill, M. F., Cowie, B., Gilmore, A., & Smith, L. F. (2010). Preparing assessment-capable teachers: What should pre-service teachers know and be able to do? *Assessment Matters*, 7, 44–64.
- Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, 25(3), 385–402.
- Kane, M. (2016). Validation strategies delineating and validating proposed interpretations and uses of test scores. In S. Lane, M. R. Raymond, & T. M. Haladyna, (Eds.), *Handbook of test development* (pp. 64–80). Taylor and Francis.
- Kingdom of Saudi Arabia. (2016). *Vision 2030*. <http://www.vision2030.gov.sa/en>
- Kroll, T., & Neri, M. (2009). Designs for mixed methods research. In S. Andrew & E. J. Halcomb (Eds.), *Mixed methods research for nursing and the health sciences* (pp. 31–49). Blackwell.
- Lam, R. (2015). Language assessment training in Hong Kong: Implications for language assessment literacy. *Language Testing*, 32(2), 169–197.

- 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.
- McMillan, J. H. (2001). *Essential assessment concepts for teachers and administrators*. Corwin Press.
- McMillan, J. H. (2003). Understanding and improving teachers' classroom assessment decision making: Implications for theory and practice. *Educational Measurement: Issues and practice*, 22(4), 34–43.
- McNamara, T., & Roever, C. (2006). *Language testing: The social dimension*. Blackwell.
- Mertler, C. A. (2003). *Preservice versus inservice teachers' assessment literacy: Does classroom experience make a difference?* [Paper presentation]. The Annual Meeting of the Mid-Western Educational Research Association, Columbus, OH. ERIC <https://files.eric.ed.gov/fulltext/ED482277.pdf>
- Mertler, C. A., & Campbell, C. (2004). Secondary teachers' assessment literacy: Does classroom experience make a difference? *American Secondary Education*, 33, 49–64.
- Mertler, C. A., & Campbell, C. (2005). *Measuring teachers' knowledge & application of classroom assessment concepts: Development of the assessment literacy inventory* [Paper presentation]. Annual Meeting of the American Educational Research Association, Montreal, Canada.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *The American Psychologist*, 50(9), 741–749. <https://doi.org/10.1002/j.2333-8504.1994.tb01618.x>
- Ministry of Culture and Information. (2013). *Kingdom of Saudi Arabia: Facts*. <http://www.info.gov.sa/portals/Kingdom/AKingdom.html>

Ministry of Education. (2013). *Directorates of education in the Kingdom*.

<https://www.moe.gov.sa/Arabic/Pages/default.aspx>

Ministry of Education. (2014). *Statistical reports*.

<https://www.moe.gov.sa/Arabic/Ministry/Pages/Statistical-reports.aspx>

Ministry of Education (2022). *Student Evaluation regulation*.

General Administration of Evaluation and Admission.

Ministry of Finance. (2013). *Government budget data*.

<http://www.mof.gov.sa/english/DownloadsCenter/Pages/Budget.aspx>

Ministry of Finance. (2014). *Saudi Arabian budget*.

[http://www.mof.gov.sa/English/DownloadsCenter/Budget/Statement%20Details%20\(PDF\).pdf](http://www.mof.gov.sa/English/DownloadsCenter/Budget/Statement%20Details%20(PDF).pdf)

Newfields, T. (2006). Teacher development and assessment literacy. In *Authentic communication: Proceedings of the 5th Annual JALT Pan-SIG Conference* (pp. 48–73).

Nichols, P. D., Meyers, J. L., & Burling, K. S. (2009). A framework for evaluating and planning assessments intended to improve student achievement. *Educational Measurement: Issues and Practice*, 28(3), 14–23.

Onsman, A. (2010). Dismantling the perceived barriers to the implementation of national higher education accreditation guidelines in the Kingdom of Saudi Arabia. *Journal of Higher Education Policy and Management*, 32(5), 511–519.
<https://doi.org/10.1080/1360080X.2010.511123>

Organisation for Economic Co-operation and Development. (2013). *Synergies for better learning: An international perspective on evaluation and assessment*. OECD Publishing. <https://doi.org/10.1787/9789264190658-en>

- Organisation for Economic Co-operation and Development. (2016a). *PF3.1: Public spending on childcare and early education*. <http://www.oecd.org/els/family/database.htm>.
- Organisation for Economic Co-operation and Development. (2016b). *Starting Strong IV: Monitoring quality in early childhood education and care country note: Australia*. <http://www.oecd.org/education/school/Monitoring-Quality-in-ECEC-Australia.pdf>
- Organisation for Economic Co-operation and Development. (2019). *PISA 2018 Database*. <https://www.oecd.org/pisa/data/2018database/>
- Organisation for Economic Co-operation and Development. (2020). *Education in Saudi Arabia*. OECD Publishing. <https://doi.org/10.1787/76df15a2-en>
- O'Sullivan, R. S., & Johnson, R. L. (1993). *Using performance assessments to measure teachers' competence in classroom assessment* [Paper presentation]. Annual Meeting of the American Educational Research Association, Atlanta, GA.
- Oyaid, A. (2009). *Education policy in Saudi Arabia and its relation to secondary school teachers' ITC use, perceptions, and views of the future of ICT in education* [Unpublished doctoral dissertation]. University of Exeter.
- Pastore, S., & Andrade, H. L. (2019). Teacher assessment literacy: A three-dimensional model. *Teaching and Teacher Education*, 84, 128–138.
- Paterno, J. (2001). Measuring success: A glossary of assessment terms. In *Building cathedrals: Compassion for the 21st century* (pp. 1–26).
- Plake, B. S., Impara, J., & Fager, J. (1993). Assessment competencies of teachers: A national survey. *Educational Measurement: Issues and Practice*, 12(4), 10–39. <https://doi.org/10.1111/j.1745-3992.1993.tb00548.x>
- Plake, B. S., & Impara, J. C. (1992). *Teacher competencies questionnaire description*. University of Nebraska.

- Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory and Practice*, 48(1), 4–11.
- 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>
- Royal Embassy of Saudi Arabia in Washington. (2014). *About Saudi Arabia*.
http://www.saudiembassy.net/about/country-information/economy_global_trade/trade.aspx (retrieved 27 October 2014).
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119–144. <https://doi.org/10.1007/BF00117714>
- Scarino, A. (2013). Language assessment literacy as self-awareness: Understanding the role of interpretation in assessment and in teacher learning. *Language Testing*, 30(3), 309–327. <https://doi.org/10.1177/0265532213480128>
- Schafer, W. D., & Lissitz, R. W. (1987). Measurement training for school personnel: Recommendations and reality. *Journal of Teacher Education*, 18(3), 57–63.
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4–14.
- Stiggins, R. (2009). Essential formative assessment competencies for teachers and school leaders. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 233–250). Routledge.
- Stiggins, R. J. (1991a). Assessment literacy. *Phi Delta Kappan*, 72, 534–539.
- Stiggins, R. J. (1991b). Relevant classroom assessment training for teachers. *Educational Measurement: Issues and Practice*, 10(1), 7–12.
- Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan*, 77(3), 238–245.

- Stiggins, R. J. (1997). Dealing with the practical matter of quality performance assessment. *Measurement in Physical Education and Exercise Science*, 1(1), 5–17.
- Stiggins, R. J. (1999). Evaluating classroom assessment training in teacher education programs. *Educational Measurement: Issues and Practice*, 18(1), 23–27.
- Stiggins, R. J. (2005). From formative assessment to assessment for learning: A path to success in standards-based schools. *Phi Delta Kappan*, 87(4), 324–328.
- Stoynoff, S., & Chapelle, C. A. (2005). *ESOL tests and testing*. Teachers of English to Speakers of Other Languages, Inc.
- Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, 48, 1273–1296.
- Tatweer. (2008). *A comparative analysis of national education development and e-learning in various countries*. King Abdullah Bin Abdulaziz Project for Public Educational Development.
- Tierney, R. D. (2006). Changing practices: Influences on classroom assessment. *Assessment in Education: Principles, Policy and Practice*, 13(3), 239–264.
<https://doi.org/10.1080/09695940601035387>
- Tognolini, J. (2020). *The beginning of a journey to assessment and data literacy for teachers*.
<https://cpl.nswtf.org.au/journal/semester-2-2020/the-beginning-of-a-journey-to-assessment-and-data-literacy-for-teachers/>
- Tognolini, J., & Davidson, M. (2012). Assessment, standards-referencing and standard setting. In M. Mok (Ed.), *Self-directed learning oriented assessments in the Asia-Pacific* (pp. 23–41). Springer. https://doi.org/10.1007/978-94-007-4507-0_2
- Tognolini, J., & Stanley, G. (2007). Standards-based assessment: A tool and means to the development of human capital and capacity building in education. *Australian Journal of Education*, 51(2), 129–145.

United Nations Educational, Scientific and Cultural Organization Institute for Statistics.

(2019). *Education database*. <http://data.uis.unesco.org/>

United Nations Educational, Scientific and Cultural Organization Institute for Statistics.

(2020). *Education database*. <http://data.uis.unesco.org/>

Volante, L., & Fazio, X. (2007). Exploring teacher candidates' assessment literacy:

Implications for teacher education reform and professional development. *Canadian Journal of Education*, 30(3), 749–770.

Vygotsky, L. (1978). *Mind and society: The development of higher mental processes*.

Harvard University Press.

Webb, N. (2002). *Assessment literacy in a standard-based urban education setting* [Paper presentation]. American Educational Research Association Annual Meeting, New Orleans.

Willis, J., Adie, L., & Klenowski, V. (2013). Conceptualising teachers' assessment literacies in an era of curriculum and assessment reform. *Australian Educational Researcher*, 40(2), 241–256. <https://doi.org/10.1007/s13384-013-0089-9>

Wiseman, A. W. (2010). The uses of evidence for educational policymaking: Global contexts and international trends. *Review of Research in Education*, 34(1), 1–24. <https://doi.org/10.3102/0091732X09350472>

World Bank & Education and Training Evaluation Commission. (2016). Measuring the state of education [Unpublished].

Xu, Y., & Brown, G. T. (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>

Yousuf, Y. (2018). Obstacles of implementing authentic assessment in Islamic schools in the Tabuk Region. *International Journal of Educational Psychological Studies*, 3(2), 292–316.

Appendices

Appendix A: Ethics Approval

Research Integrity & Ethics Administration



HUMAN RESEARCH ETHICS COMMITTEE

Tuesday, 22 December 2020

Mr James Tognolini

Social Work; Faculty of Arts and Social Sciences Email: james.tognolini@sydney.edu.au

Dear James,

The University of Sydney Human Research Ethics Committee (HREC) has considered your application.

I am pleased to inform you that after consideration of your response, your project has been approved.

Details of the approval are as follows:

Project No.: 2020/771
Project Title: The Development and validation of an assessment literacy framework and instrument to measure the assessment literacy of teachers in Saudi Arabia
Authorised Personnel: Tognolini James; Alzhrany Abdullah; Twing Jon S.;
Approval Period: 22 December 2020 to 22 December 2024
First Annual Report Due: 22 December 2021

Documents Approved:

Date Uploaded	Version Number	Document Name
30/11/2020	Version 1	CONFIDENTIALITY AGREEMENT . TRANSLATOR 2
30/11/2020	Version 1	LETTER TO MINISTRY OF EDUCATION
30/11/2020	Version 1	PLS
30/11/2020	Version 2	PIS
30/11/2020	Version 2	INVITATION TO SCHOOL PRINCIPAL
30/11/2020	Version 2	SF
25/10/2020	Version 1	Focus group interview questions
25/10/2020	Version 1	Participant Consent Form
25/10/2020	Version 1	Likert Survey Sample
25/10/2020	Version 1	Letter to Ministry of Education

Special Condition/s of Approval

- It will be a condition of approval that data not be stored on eNotebooks as it cannot be deleted.

Please use a similar data storage platform, e.g. the Research Data Store (RDS), for storage of participant information so that it can be deleted after the five year data retention period.

Condition/s of Approval

- Research must be conducted according to the approved proposal.
- An annual progress report must be submitted to the Ethics Office on or before the anniversary of approval and on completion of the project.
- You must report as soon as practicable anything that might warrant review of ethical approval of the project including:
 - Serious or unexpected adverse events (which should be reported within 72 hours).
 - Unforeseen events that might affect continued ethical acceptability of the project.
- Any changes to the proposal must be approved prior to their implementation (except where an amendment is undertaken to eliminate *immediate* risk to participants).

Research Integrity & Ethics Administration
Research Portfolio

Level 3, F23 Administration Building
The University of Sydney

NSW 2006 Australia

T + 61 2 9036 9161

E human.ethics@sydney.edu.au

W sydney.edu.au/ethics

ABN 15 211 513 464
CRICOS 00026A

- Personnel working on this project must be sufficiently qualified by education, training and experience for their role, or adequately supervised. Changes to personnel must be reported and approved.
- Personnel must disclose any actual or potential conflicts of interest, including any financial or other interest or affiliation, as relevant to this project.
- Data and primary materials must be retained and stored in accordance with the relevant legislation and University guidelines.
- Ethics approval is dependent upon ongoing compliance of the research with the *National Statement on Ethical Conduct in Human Research*, the *Australian Code for the Responsible Conduct of Research*, applicable legal requirements, and with University policies, procedures and governance requirements.
- The Ethics Office may conduct audits on approved projects.
- The Chief Investigator has ultimate responsibility for the conduct of the research and is responsible for ensuring all others involved will conduct the research in accordance with the above.

This letter constitutes ethical approval only.

Please contact the Ethics Office should you require further information or

clarification. Sincerely,

[REDACTION]

Dr Haryana Dillon, Chair

Human Research Ethics Committee (HREC 3)

The University of Sydney of Sydney HRECs are constituted and operate in accordance with the National Health and Medical Research Council's (NHMRC) [National Statement on Ethical Conduct in Human Research \(2018\)](#) and the NHMRC's [Australian Code for the Responsible Conduct of Research \(2018\)](#)

Appendix B: The Likert Survey

Page 2 of 2

PRACTITIONER STANDARD 10.1

Prepare a diverse and comprehensive set of direct and in-direct evaluation tools

1. To what extent do practitioner teachers have to **‘Know and understand the principles of assessment in the literature including related definitions such as: evaluation, measurement and testing’** to demonstrate **PRACTITIONER STANDARD 10.1** ?

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

2. To what extent do practitioner teachers have to **‘Know and understand some of the more common purposes of assessment, including how the assessment is likely to be used and what are likely to be the intended and unintended consequences of using these assessments’** to demonstrate **PRACTITIONER STANDARD 10.1** ?

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

3. To what extent do practitioner teachers have to **‘know the features of the range of assessment forms in relation of its strengths and weaknesses’** to demonstrate **PRACTITIONER STANDARD 10.1** ?

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

4. To what extent do practitioner teachers have to **‘know and understand how to select assessment tasks which clearly related to learning objectives’** to demonstrate **PRACTITIONER STANDARD 10.1** ?

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

PRACTITIONER STANDARD 10.1

Prepare a diverse and comprehensive set of direct and in-direct evaluation tools

5. To what extent do practitioner teachers have to **‘Know and understand some of the more common assessment dichotomies used in education and the meaning of relevant terms associated with them’** to demonstrate **PRACTITIONER STANDARD 10.1** ?

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

6. To what extent do practitioner teachers have to **‘Know and understand the meaning of reliability and validity in relation of selecting assessment tasks’** to demonstrate **PRACTITIONER STANDARD 10.1** ?

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

7. To what extent do practitioner teachers have to **‘Know and understand illegal, inappropriate, and unethical assessment practices’** to demonstrate **PRACTITIONER STANDARD 10.1** ?

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

PRACTITIONER STANDARD 10.2

Evaluate students' performance and learning using diagnostic, formative and summative evaluation methods. Identify and improve students' strengths and challenges

8. To what extent do practitioner teachers have to **'Know and understand how to use a range of assessment items that are fit-for-purpose.'** to demonstrate **PRACTITIONER STANDARD 10.2** ?

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

9. To what extent do practitioner teachers have to **'Know how to administer a range of assessments and understand the need for consistent instructions to ensure a consistent testing environment'** to demonstrate **PRACTITIONER STANDARD 10.2** ?

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

10. To what extent do practitioner teachers have to **'Know when and how to assess and report student progress against standards'** to demonstrate **PRACTITIONER STANDARD 10.2** ?

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

PRACTITIONER STANDARD 10.2

Evaluate students' performance and learning using diagnostic, formative and summative evaluation methods. Identify and improve students' strengths and challenges

11. To what extent do practitioner teachers have to **'Know and understand how to Use diagnostic assessment to assess prior students' knowledge'** to demonstrate **PRACTITIONER STANDARD 10.2?**

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

12. To what extent do practitioner teachers have to **'Know and understand how to design summative assessments tools that will enable the results to be used to improve teaching and learning'** to demonstrate **PRACTITIONER STANDARD 10.2 ?**

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

13. To what extent do practitioner teachers have to **'Know and understand the purposes of assessment results recording'** to demonstrate **PRACTITIONER STANDARD 10.2 ?**

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

14. To what extent do practitioner teachers have to **'know and understand the requirements, structure, purpose and action of conducting formative assessment appropriately'** to demonstrate **PRACTITIONER STANDARD 10.2 ?**

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

PRACTITIONER STANDARD 10.3

Apply self-assessment strategies effectively and support students to participate and making use of its results

15. To what extent do practitioner teachers have to '**Know and understand the definition of self-assessment and peer assessment**' to demonstrate **PRACTITIONER STANDARD 10.3** ?

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

16. To what extent do practitioner teachers have to '**know and understand the main purpose of self and peer assessment**' to demonstrate **PRACTITIONER STANDARD 10.3** ?

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

17. To what extent do practitioner teachers have to '**Know and understand how to work with students to analyse assessment results to identify their strengths and weaknesses**' to demonstrate **PRACTITIONER STANDARD 10.3** ?

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

PRACTITIONER STANDARD 10.3

Apply self-assessment strategies effectively and support students to participate and making use of its results

18. To what extent do practitioner teachers have to **use assessment methods to identify current student performance, skills, achievement to be aligned with the lesson plan** to demonstrate **PRACTITIONER STANDARD 10.3?**

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

19. To what extent do practitioner teachers have to **'Know and understand how to develop student's assessment skills'** to demonstrate **PRACTITIONER STANDARD 10.3 ?**

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

20. To what extent do practitioner teachers have to **'Know and understand of some feedback techniques to engage students in the instruction'** to demonstrate **PRACTITIONER STANDARD 10.3?**

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

PRACTITIONER STANDARD 10.4

use the evaluation results to organise and modify study units, provide constructive feedback to students and parents in a timely manner, and pursue its applications

21. To what extent do practitioner teachers have to **'Know the definition of data literacy'** to demonstrate **PRACTITIONER STANDARD 10.4?**

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

22. To what extent do practitioner teachers have to **'Know and understand that there are a range of skills needed to transform data into information, evidence and strategies for teaching and learning'** to demonstrate **PRACTITIONER STANDARD 10.4?**

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

23. To what extent do practitioner teachers have to **'Know and understand the meaning of some of the more commonly used statistical terms used for analysing and summarising student performance data'** to demonstrate **PRACTITIONER STANDARD 10.4?**

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

PRACTITIONER STANDARD 10.4

use the evaluation results to organise and modify study units, provide constructive feedback to students and parents in a timely manner, and pursue its applications

24. To what extent do practitioner teachers have to 'Know the meaning of data, information and evidence, as well understanding that evidence is used to inform educational practice and reporting ' to demonstrate **PRACTITIONER STANDARD 10.4** ?

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

25. To what extent do practitioner teachers have to 'Know and understand how to collect and organise data for classroom use' to demonstrate **PRACTITIONER STANDARD 10.4**?

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

26. To what extent do practitioner teachers have to ' Know the meaning of feedback in assessment' to demonstrate **PRACTITIONER STANDARD 10.4** ?

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

PRACTITIONER STANDARD 10.4

use the evaluation results to organise and modify study units, provide constructive feedback to students and parents in a timely manner, and pursue its applications

27. To what extent do practitioner teachers have to '**Know and understand the features of effective feedback**' to demonstrate **PRACTITIONER STANDARD 10.4**?

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

28. To what extent do practitioner teachers have to '**Know and understand some of the main requirements for delivering effective feedback**' to demonstrate **PRACTITIONER STANDARD 10.4**?

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

29. To what extent do practitioner teachers have to '**Know the main outcomes from research on feedback**' to demonstrate **PRACTITIONER STANDARD 10.4** ?

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

PRACTITIONER STANDARD 10.5

provide clear and accurate learning reports to students and parents using accurate and reliable technology and records about students' learning outcomes

30. To what extent do practitioner teachers have to '**Know and understand the purposes of reporting**' to demonstrate **PRACTITIONER STANDARD 10.5**?

Essential

Desirable

Not Necessary

☐☐☐

Comments

31. To what extent do practitioner teachers have to '**Know and understand some of the more common reporting strategies used in schools for disseminating assessment information to parents.**' to demonstrate **PRACTITIONER STANDARD 10.5**?

Essential

Desirable

Not Necessary

☐☐☐

Comments

32. To what extent do practitioner teachers have to '**Know and understands the meaning of the grades that are used to summarise performance**' to demonstrate **PRACTITIONER STANDARD 10.5** ?

Essential

Desirable

Not Necessary

☐☐☐

Comments



THE UNIVERSITY OF
SYDNEY

Elaborations of Assessment

Advanced Standards

ADVANCED STANDARD 10.1

Prepare a variety of direct and in-direct evaluation tools and develop effective evaluation tools to support his colleagues.

33. To what extent do advanced teachers have to **'Know the relationship between item difficulty and the level of cognition and can use classification systems and taxonomies in b assessments'** to demonstrate **ADVANCED STANDARD 10.1?**

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

34. To what extent do advanced teachers have to **'Know and understand the assumptions that underpin educational assessment'** to demonstrate **ADVANCED STANDARD 10.1?**

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

35. To what extent do advanced teachers have to **'Know and understand how to improve items and tasks that have been written for assessing achievement and learning outcomes'** to demonstrate **ADVANCED STANDARD 10.1?**

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

ADVANCED STANDARD 10.1

Prepare a variety of direct and in-direct evaluation tools and develop effective evaluation tools to support his colleagues.

36. To what extent do advanced teachers have to **'Know and understand how to develop items and tasks based on analyses of outcomes in terms of the factors that may affect these items'** to demonstrate **ADVANCED STANDARD 10.1?**

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

37. To what extent do advanced teachers have to **'Know and understand how to develop assessment items and tasks in order to ensure high quality in their validity and reliability'** to demonstrate **ADVANCED STANDARD 10.1?**

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

38. To what extent do advanced teachers have to **'Know how to develop assessments that minimise sources of bias that can distort assessment results'** to demonstrate **ADVANCED STANDARD 10.1?**

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

39. To what extent do advanced teachers have to **'Know and understand and carry out their legal and ethical responsibilities in assessment at the school- and classroom levels'** to demonstrate **ADVANCED STANDARD 10.1?**

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

ADVANCED STANDARD 10.2

Evaluate students' performance and learning using the diagnostic formative and summative evaluation, identifies and improve students' strengths and challenges, and develop effective evaluation methods to support his colleagues.

40. To what extent do advanced teachers have to '**Know how to construct, develop and use a range of item types and assessment tasks that are technically sound, editorially correct and that appropriately assess the outcomes and performance standard levels being assessed**' to demonstrate **ADVANCED STANDARD 10.2**?

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

41. To what extent do advanced teachers have to '**Know the process on how to accurately mark assessments and collect , record student results**' to demonstrate **ADVANCED STANDARD 10.2**?

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

42. To what extent do advanced teachers have to '**Know why moderation and comparability is needed at the school- and classroom levels**' to demonstrate **ADVANCED STANDARD 10.2**?

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

ADVANCED STANDARED 10.2

Evaluate students' performance and learning using the diagnostic formative and summative evaluation, identifies and improve students' strengths and challenges, and develop effective evaluation methods to support his colleagues.

43. To what extent do advanced teachers have ' **Know the content, purpose and structure of major state, national and international testing programs** ' to demonstrate **ADVANCED STANDARD 10.2?**

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

44. To what extent do advanced teachers have to ' **Know various ways to record results to capture an image of student performance that can be compared against performance standards** ' to demonstrate **ADVANCED STANDARD 10.2?**

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

45. To what extent do advanced teachers have to **Know some of the more common ways of achieving comparability at the school-level in a standards-referenced system and its strengths, weaknesses** ' to demonstrate **ADVANCED STANDARD 10.2?**

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

46. To what extent do advanced teachers have to **Know and understand some of the ways that assessment methods can be misused and misinterpreted** ' to demonstrate **ADVANCED STANDARD 10.2?**

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

ADVANCED STANDARD 10.3

apply self-assessment strategies on students, develop effective strategies that enhance students' participation in the assessment process, and support his colleagues with them.

47. To what extent do advanced teachers have to 'know and understand the definition of self-regulation' to demonstrate **ADVANCED STANDARD 10.3**?

Essential

Desirable

Not Necessary

☐☐☐

Comments

48. To what extent do advanced teachers have to 'use and develop assessment approaches to engaging prior students' knowledge and skills in lesson plans ' to demonstrate **ADVANCED STANDARD 10.3**?

Essential

Desirable

Not Necessary

☐☐☐

Comments

49. To what extent do advanced teachers have to 'Know and understand many ways to encourage students to be involved in self and peer assessment' to demonstrate **ADVANCED STANDARD 10.3**?

Essential

Desirable

Not Necessary

☐☐☐

Comments

50. To what extent do advanced teachers have to 'Know and understand what it means to tailor assessments to the individual needs of all students' to demonstrate **ADVANCED STANDARD 10.3**?

Essential

Desirable

Not Necessary

☐☐☐

Comments

ADVANCED STANDARD 10.3

apply self-assessment strategies on students, develop effective strategies that enhance students' participation in the assessment process, and support his colleagues with them.

51. To what extent do advanced teachers have to **'Know and understand how to provide opportunity for students to construct criteria** to demonstrate **ADVANCED STANDARD 10.3?**

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

52. To what extent do advanced teachers have to **'Sharing learning outcomes with student to ensure their standing before heading to the next steps.'** to demonstrate **ADVANCED STANDARD 10.3?**

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

ADVANCED STANDARD 10.4

use the evaluation results to organise and modify study units, provide constructive feedback to students and parents in a timely manner, and pursue its application. Develop effective methods aimed at improving learning and supports his colleagues with them.

53. To what extent do advanced teachers have to **'Know and understand some of the terms associated with classroom data'** to demonstrate **ADVANCED STANDARD 10.4?**

Essential

Desirable

Not Necessary

☐☐☐

Comments

54. To what extent do advanced teachers have to **'Know the strengths and weaknesses of some of the more common statistics used to summarise student performance'** to demonstrate **ADVANCED STANDARD 10.4?**

Essential

Desirable

Not Necessary

☐☐☐

Comments

55. To what extent do advanced teachers have to **'Know and understand how to generate teaching and learning interventions from an analysis of assessment data'** to demonstrate **ADVANCED STANDARD 10.4?**

Essential

Desirable

Not Necessary

☐☐☐

Comments

56. To what extent do advanced teachers have to **'Know and understand that different kinds of data exist and have a basic understanding of which kind of data is useful for particular decisions'** to demonstrate **ADVANCED STANDARD 10.4?**

Essential

Desirable

Not Necessary

☐☐☐

Comments

ADVANCED STANDARD 10.4

use the evaluation results to organise and modify study units, provide constructive feedback to students and parents in a timely manner, and pursue its application. Develop effective methods aimed at improving learning and supports his colleagues with them.

57. To what extent do advanced teachers have to **'Know and understand how to interpret the results of large-scale testing programs'** to demonstrate **ADVANCED STANDARD 10.4?**

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

58. To what extent do advanced teachers have to **'Define and provide relevant examples of how feedback is used in schools'** to demonstrate **ADVANCED STANDARD 10.4?**

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

59. To what extent do advanced teachers have to **'Know and understand the requirements for effective feedback'** to demonstrate **ADVANCED STANDARD 10.4?**

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

60. To what extent do advanced teachers have to **'Know and understand how feedback should be tailored to the needs of individual students'** to demonstrate **ADVANCED STANDARD 10.4?**

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

61. To what extent do advanced teachers have to **'Know how to assess whether the feedback that is given in class is understood by the students'** to demonstrate **ADVANCED STANDARD 10.4?**

Essential

Desirable

Not Necessary



Comments

ADVANCED STANDARD 10.5

provide clear and accurate learning reports to students and parents using accurate and reliable technology and records about students' learning outcomes. Develop effective methods of preparing evaluation reports and support colleagues with them.

62. To what extent do advanced teachers have to **'Know and understand how to prepare written reports that report accurately, clearly and respectfully to parents/careers'** ADVANCED STANDARD 10.5?

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

63. To what extent do advanced teachers have to **' Know and understand the importance of collecting reliable and valid assessment data about the achievement of students and report it in a manner which is consistent with the reporting strategy(ies) used within a school'** to demonstrate ADVANCED STANDARD 10.5?

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

64. To what extent do advanced teachers have to **'Know and understand the meaning of the grades/marks on the reports'** to demonstrate ADVANCED STANDARD 10.5?

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

65. To what extent do advanced teachers have to **'Know and understand the strengths and weaknesses of different ways of recording data for reporting purposes'** to demonstrate ADVANCED STANDARD 10.5?

Essential	Desirable	Not Necessary
☐	☐	☐

Comments

Appendix C: The Assessment Literacy Framework

Standard	Practitioner teachers			Advanced teachers		
	Sub-Standard	Knowledge and understanding	Is able to ...	Sub-Standard	Knowledge and understanding	Is able to ...
10.1 Preparing evaluation tools	10–1-1 Prepare a diverse and comprehensive set of direct and indirect evaluation tools	1-Know and understand the principles of assessment in the literature including related definitions such as evaluation, measurement and testing	Explain these principles and provide examples about their application inside the school	10–1-2 Prepare a variety of direct and indirect evaluation tools and develop effective evaluation tools to support colleagues	1-Know the relationship between item difficulty and the level of cognition based and can use classification systems and taxonomies in building assessments	Explain how classification and taxonomies like Bloom’s Taxonomy can be used to show growth of cognition in learning
		2-Know and understand some of the more common purposes of assessment, including how the assessment is likely to be used and what are likely to be the intended and unintended consequences of using these assessments	Describe the effective ways that assessments are used in schools, such as: monitoring and certifying student and teacher performance		2- Know and understand the assumptions that underpin educational assessment	Describe the major assumptions that reinforce educational assessment such as: assessment must be fair and professional judgement

Standard	Practitioner teachers			Advanced teachers		
	Sub-Standard	Knowledge and understanding	Is able to ...	Sub-Standard	Knowledge and understanding	Is able to ...
		3-Know the features of the range of assessment forms and their strengths and weaknesses	Discuss the strengths and weaknesses of a range of forms of currently used assessments, such as multiple choice questions, observation and portfolio		3- Know and understand how to improve items and tasks that have been written for assessing achievement and learning outcomes	Support colleagues to optimise the writing of assessment items and tasks
		4-Know and understand how to select assessment tasks which clearly related to learning objectives	Select assessment tasks and items that are compatible with learning strategies		4- Know and understand how to develop items and tasks based on analyses of outcomes in terms of the factors that may affect these items	Know and identify some of the main threats to validity such as the item structures in terms of its quality, sufficient time, clarifying, cheating and inappropriate items

Standard	Practitioner teachers			Advanced teachers		
	Sub-Standard	Knowledge and understanding	Is able to ...	Sub-Standard	Knowledge and understanding	Is able to ...
			Explain the main differences between whole assessment strategies systems such as assessment for and of learning, school-based and standardised assessments, and relevant terms: standards; performance standards; content standards; outcomes; higher order thinking skills			
		5- Know and understand some of the more common assessment dichotomies used in education and the meaning of associated terms			5- Know and understand how to develop assessment items and tasks in order to ensure high quality in their validity and reliability	Explain to colleagues the role that these assessment terms play in school-based and other assessments
		6- Know and understand the meaning of reliability and validity in relation to selecting assessment tasks	Describe the importance of validity and reliability in classroom assessment		6-Know how to develop assessments that minimise sources of bias, which can distort assessment results	Develop assessments that have taken account of potential sources of bias such as that caused by differences in gender, or race

Standard	Practitioner teachers			Advanced teachers		
	Sub-Standard	Knowledge and understanding	Is able to ...	Sub-Standard	Knowledge and understanding	Is able to ...
10–2 Applying evaluation		7-Know and understand illegal, inappropriate, and unethical assessment practices	Explain the importance of assessment ethics and their impact on students’ performance		7-Know and understand and carry out legal and ethical responsibilities in assessment at the school and classroom levels	Support colleagues to be more knowledgeable in assessment ethics
	10–2-1 Evaluate students’ performance and learning using diagnostic, formative and summative evaluation methods. •Identifies and improves students’ strengths and challenges	1- Know and understand how to use a range of assessment items that are fit-for-purpose	Use a range of assessment items that effectively assess a range of skills in a unit of learning and mirror the relative weights of content materials or skills required by the unit of learning	10–2-2 Evaluate students’ performance and learning using the diagnostic formative and summative evaluation, identify and improve students’ strengths and challenges, and develop effective evaluation methods to support colleagues	1-Know how to write, develop and use a range of item types and assessment tasks that are technically sound, editorially correct and that appropriately assess the outcomes and performance standard levels being assessed	Explain how to support colleagues in relation to know assessment tasks (items) that meet the requirements for good quality assessment tasks

Standard	Practitioner teachers			Advanced teachers		
	Sub-Standard	Knowledge and understanding	Is able to ...	Sub-Standard	Knowledge and understanding	Is able to ...
		2-Know how to administer a range of assessments and understand the need for consistent instructions to ensure a consistent testing environment	Demonstrate through relevant examples knowledge of the requirements for administering classroom assessments and more formal assessments		2-Know the process for accurately marking assessments and collecting and recording student results	Give exemplars on how to collect and record students results in a systematic way and identify the potential problems with marking in a range of assessment types
		3- Know when and how to assess and report student progress against standards	Complete and interpret reports to stakeholders in a standards-referenced system		3- Know why comparability is needed at the school and classroom levels	Describe situations in schools where comparability is likely to be required. For example, it may be needed to compare results across similar grades in a school or compare results from students within a class who have done different assessments

Standard	Practitioner teachers			Advanced teachers		
	Sub-Standard	Knowledge and understanding	Is able to ...	Sub-Standard	Knowledge and understanding	Is able to ...
		4-Know and understand how to use diagnostic assessment to assess prior student knowledge	Demonstrate some of the ways to figure out prior student knowledge and performance		4- Know the content, purpose and structure of major state, national and international testing programs	Describe the purposes, main characteristics and basic structures of these programs
		5- Know and understand how to design summative assessments tools that will enable the results to be used to improve teaching and learning	Describe some of the ways that marks can be used in the learning process to improve learning, including for the purposes of helping teachers to adapt and adjust teaching to accommodate students' learning needs		5- Know various ways to record results to capture an image of student performance that can be compared against performance standards	Demonstrate ways that results of student assessments are recorded and used to arrive at an overall grade representing a performance standard

Standard	Practitioner teachers			Advanced teachers		
	Sub-Standard	Knowledge and understanding	Is able to ...	Sub-Standard	Knowledge and understanding	Is able to ...
		6- Know and understand the purposes of assessment results recording	Describe the main purposes of recording, including A. reporting achievement; B. monitoring strengths and weaknesses and recording progress		6-Know some of the more common ways of achieving comparability at the school level in a standards-referenced system and its strengths and weaknesses	Describe the way that consensus moderation and moderation using common tasks can be applied within schools to obtain comparability across classes and the strengths and weaknesses of consensus and common-task moderation
		7-Know and understand the requirements, structure, purpose and action of conducting formative assessment appropriately	Give some examples of formative assessment instruments and how to use them to inform teaching and learning		7- Know and understand some of the ways that assessment methods can be misused and misinterpreted	Help teachers by explaining and giving examples of how assessment methods can be misused and misinterpreted in the context of their original purpose
10-3 Engaging students in evaluation processes	10–3-1 Applies self-assessment strategies effectively and supports students to participate and	1- Know and understand the definition of self-assessment and peer assessment	Describe how to apply the self and peer assessment during the instruction in a positive manner	10–3-2 Apply self-assessment strategies on students, develops effective strategies that enhance students’	1-Know and understand the definition of self-regulation	Support teachers to use assessment tasks to increase student ability in terms of organising their learning process, such as by sharing

Standard	Practitioner teachers			Advanced teachers		
	Sub-Standard	Knowledge and understanding	Is able to ...	Sub-Standard	Knowledge and understanding	Is able to ...
	make use of its results			participation in the assessment process and supports colleagues with them		standards, rubrics and criteria
		2-Know and understand the main purpose of self and peer assessment	Discuss the processes and ultimate goals of applying self and peer assessments based in the research's outcomes	2-Use and develop assessment approaches to engage prior student knowledge and skills in lesson plans	2-Use and develop assessment approaches to engage prior student knowledge and skills in lesson plans	Explain some of the assessment methods that will benefit from prior student knowledge to ensure more participation in current practices
		3- Know and understand how to work with students to analyse assessment results to identify their strengths and weaknesses	Describe some ways to support students to benefit from their assessment results and monitor their learning	3-Know and understand many ways to encourage students to be involved in self- and peer assessment	3-Know and understand many ways to encourage students to be involved in self- and peer assessment	Select, develop and apply some strategies to reinforce students' participation in self- and peer assessment
		4- Use assessment methods to identify current student performance, skills, achievement to be aligned with the lesson plan	Define and provide relevant examples of how assessment methods can empower students in new learning processes	4- Know and understand what it means to tailor assessments to the individual needs of all students	4- Know and understand what it means to tailor assessments to the individual needs of all students	Develop assessments that have been adapted to accommodate students with different needs
		5- Know and understand how to develop student assessment skills	Describe some methods of improving students' assessment knowledge	5-Know and understand how to provide opportunity for students to construct criteria	5-Know and understand how to provide opportunity for students to construct criteria	Support colleagues to explain how to use instruction environment to incorporate students in their learning by co-

Standard	Practitioner teachers			Advanced teachers		
	Sub-Standard	Knowledge and understanding	Is able to ...	Sub-Standard	Knowledge and understanding	Is able to ...
						constructing rubrics that can be used in assessment tasks
		6-Know and understand some feedback techniques to engage students in instruction	Describe how to improve the learning by using feedback to support students to adapt their learning strategies and to become independent learners		6-Sharing learning outcomes with students to ensure their standing before heading to the next steps	Help teachers to share with students the results of their works, such as tests, projects, worksheets and presentations
10–4 Employing evaluation results	10–4-1 Use the evaluation results to organise and modify study units, provide constructive feedback to students and parents in a timely manner, and pursue its applications	1- Know the definition of data literacy	Define data literacy as the ability to be able to understand and use data effectively	10–4-2 Use the evaluation results to organise and modify study units, provide constructive feedback to students and parents in a timely manner, and pursue its application. Develop effective methods aimed at improving	1- Know and understand some of the terms associated with classroom data	Define and describe terms associated with student assessment data, including: A. dichotomous (true / false) data; B. quantitative versus qualitative data; C. polytomously scored questions; and, D. Percentages and percentiles

Standard	Practitioner teachers			Advanced teachers		
	Sub-Standard	Knowledge and understanding	Is able to ...	Sub-Standard	Knowledge and understanding	Is able to ...
		2- Know and understand that there are a range of skills needed to transform data into information, evidence and strategies for teaching and learning	List some of the skills that are needed to convert data into actions for teaching and learning; identify, collect, analyse, summarise and prioritise data	learning and support colleagues with them	2-Know the strengths and weaknesses of some of the more common statistics used to summarise student performance	Describes for colleagues the strengths and weaknesses of, for example, A. mean (average), mode and median; B. standard deviation, variance and range; and, correlation
		3- Know and understand the meaning of some of the more common statistical terms used for analysing and summarising student performance data	Interpret summary data commonly used to represent student performance. For example, statistical terms such as: mean, median and mode; variance and standard deviation; correlation; range; percentiles and norms		3-Know and understand how to generate teaching and learning interventions from an analysis of assessment data	Modify teaching practice according to the data collected
		4- Know the meaning of data, information and evidence, as well understanding that evidence is used to inform educational	Distinguish between data (responses), information (responses as marks), evidence (information considered through		4- Know and understand that different kinds of data exist and have a basic understanding of which kind of data	Develop effective methods that align the needs for decision-making with the appropriate kind of data. For example, monitoring performance at a

Standard	Practitioner teachers			Advanced teachers		
	Sub-Standard	Knowledge and understanding	Is able to ...	Sub-Standard	Knowledge and understanding	Is able to ...
		practice and reporting	the lens of outcomes or other educational aims). Identify where and how evidence is used in schools to inform teaching		is useful for particular decisions	summative level usually requires statistical data; monitoring the performance on a day-to-day basis generally requires more qualitative and anecdotal data
		5- Know and understand how to collect and organise data for classroom use	Effectively take raw response data and collect it in different forms (using technology, where appropriate) so that it can be used to evaluate student learning		5- Know how to interpret the results of large-scale testing programs	Interpret the basic results of programs like SAAT and the NAP
		6- Know the meaning of feedback in assessment	Describe the importance of feedback in classroom assessment		6-Define and provide relevant examples of how feedback is used in schools	Demonstrate how feedback from a range of assessments is meaningful in terms of student learning with respect to the assessed outcomes

Standard	Practitioner teachers			Advanced teachers		
	Sub-Standard	Knowledge and understanding	Is able to ...	Sub-Standard	Knowledge and understanding	Is able to ...
		7- Know and understand the features of effective feedback	Describe the features of feedback that make it an effective tool to assist student learning. For examples, it should be A. clear; B. purposeful; C. meaningful; D. timely; and E. compatible		7-Know and understand the requirements for effective feedback	Help teachers by providing examples of feedback to students that consider its timing and its appropriateness for the type of feedback given

Standard	Practitioner teachers			Advanced teachers		
	Sub-Standard	Knowledge and understanding	Is able to ...	Sub-Standard	Knowledge and understanding	Is able to ...
			Give some examples of feedback that illustrate some of the most effective delivery modes. For example, feedback should be			
		8- Know and understand some of the main requirements for delivering effective feedback	A. delivered in an appropriate setting; B. focused on performance and not the individual; C. delivered using neutral, non-judgemental language; D. focused on highlighting areas for improvement and suggesting strategies to improve; and, aimed to empower and encourage learning		8-Know and understand how feedback should be tailored to the needs of individual students	Design and administer a range of feedback strategies that can be used in class to accommodate individual student needs and at the same time take into account the nature of the task

Standard	Practitioner teachers			Advanced teachers		
	Sub-Standard	Knowledge and understanding	Is able to ...	Sub-Standard	Knowledge and understanding	Is able to ...
		9- Know the main outcomes from research on feedback	Discuss some of the outcomes from research related to the impact that feedback has on learning in classrooms		9- Know how to assess whether the feedback that is given in class is understood by the students	Describe ways to evaluate whether the feedback has been understood by the students. Some of the ways that this can be done include: interacting with students to ensure there is a shared understanding of terminology and language; and, reflecting on how the feedback that has been given may assist students in improving their learning

Standard	Practitioner teachers			Advanced teachers		
	Sub-Standard	Knowledge and understanding	Is able to ...	Sub-Standard	Knowledge and understanding	Is able to ...
10–5 Preparing evaluation reports	10–5-1 provide clear and accurate learning reports to students and parents using accurate and reliable technology and record students’ learning outcomes	1- Know and understand the purposes of reporting	Describe the main purposes of reporting, including A. reporting achievement; B. monitoring strengths and weaknesses; and, recording progress	10–5-2 Provide clear and accurate learning reports to students and parents using accurate and reliable technology and record students’ learning outcomes. Develop effective methods of preparing evaluation reports and support colleagues with them	1- Knows and understands how to prepare written reports that report accurately, clearly and respectfully to parents/carers	Prepare reports that A. use accurate, appropriate language to report skills and progress; B. use language that is respectful and appropriate to audiences; C. bases qualitative feedback on a range of assessments; and, uses data which assesses academic, skill, social and emotional progress as appropriate

Standard	Practitioner teachers			Advanced teachers		
	Sub-Standard	Knowledge and understanding	Is able to ...	Sub-Standard	Knowledge and understanding	Is able to ...
		2- Know and understand some of the more common reporting strategies used in schools for disseminating assessment information to parents	Describe the characteristics (including strengths and weaknesses) of strategies for reporting to parents, including A. written reports; B. parent-teacher meetings and interviews; and, portfolios and e-Portfolios Explain the meaning of the various grades/marks that characterise most school reports		2-Knows and understands the importance of collecting reliable and valid assessment data about the achievement of students and reports it in a manner which is consistent with the reporting strategy(ies) used within a school	Explain to teachers why it is necessary to collect, organise and report data in an accurate, consistent and meaningful way, and that selecting reporting formats as appropriate to the task and assessment policy is important Describe and explain for teachers
		3- Knows and understands the meaning of the grades that are used to summarise performance	A. A to E grades; B. Qualitative grades; C. Benchmark-related grades; D. Marks; E. Ranks; F. Percentages and, Percentiles		3- Knows and understands the meaning of the grades/marks on the reports	A. the process for developing grades and marks that have been collected to stakeholders; B. the process of reporting within scope of subject/syllabus being assessed; C. that the sequencing of learning is linked to reporting;

Standard	Practitioner teachers			Advanced teachers		
	Sub-Standard	Knowledge and understanding	Is able to ...	Sub-Standard	Knowledge and understanding	Is able to ...
						D. appropriate next steps / aims for learning are identified; and, the technical specifications for formal assessments (e.g. mean, standard deviation, reliability, validity) is explained to parents in a manner that helps them better interpret the report
				4- Know and understand the strengths and weaknesses of different ways of recording data for reporting purposes		Support teachers to know the strengths and weaknesses of different levels of reporting for various purposes and contexts. For example, A. Student-item level; B. Summary-task level; C. Qualitative-descriptor level; and, Video or photograph

Appendix D: The Refined Assessment Literacy Framework

Standard	Practitioner teachers			Advanced teachers		
	Sub-Standard	Knowledge and understanding	Is able to ...	Sub-Standard	Knowledge and understanding	Is able to ...
10.1 Preparing evaluation tools	10–1-1 Prepare a diverse and comprehensive set of direct and indirect evaluation tools	1-Know and understand the principles of assessment in the literature including related definitions such as evaluation, measurement and testing	Explain these principles and provide examples about their application inside the school	10–1-2 Prepare a variety of direct and indirect evaluation tools and develop effective evaluation tools to support colleagues	1-Know the relationship between item difficulty and the level of cognition based and can use classification systems and taxonomies in building assessments	Explain how classification and taxonomies like Bloom’s Taxonomy can be used to show growth of cognition in learning
		2-Know and understand some of the more common purposes of assessment, including how the assessment is likely to be used and what are likely to be the intended and unintended consequences of using these assessments	Describe the effective ways that assessments are used in schools, such as: monitoring and certifying student and teacher performance		2- Know and understand the assumptions that underpin educational assessment	Describe the major assumptions that reinforce educational assessment such as: assessment must be fair and professional judgement

Standard	Practitioner teachers			Advanced teachers		
	Sub-Standard	Knowledge and understanding	Is able to ...	Sub-Standard	Knowledge and understanding	Is able to ...
		3-Know the features of the range of assessment forms and their strengths and weaknesses	Discuss the strengths and weaknesses of a range of forms of currently used assessments, such as multiple choice questions, observation and portfolio		3- Know and understand how to improve items and tasks that have been written for assessing achievement and learning outcomes	Support colleagues to optimise the writing of assessment items and tasks
		4-Know and understand how to select assessment tasks which clearly related to learning objectives	Select assessment tasks and items that are compatible with learning strategies		4- Know and understand how to develop items and tasks based on analyses of outcomes in terms of the factors that may affect these items	Know and identify some of the main threats to validity such as the item structures in terms of its quality, sufficient time, clarifying, cheating and inappropriate items

Standard	Practitioner teachers			Advanced teachers		
	Sub-Standard	Knowledge and understanding	Is able to ...	Sub-Standard	Knowledge and understanding	Is able to ...
			Explain the main differences between whole assessment strategies systems such as assessment for and of learning, school-based and standardised assessments, and relevant terms: standards; performance standards; content standards; outcomes; higher order thinking skills			
		5- Know and understand some of the more common assessment dichotomies used in education and the meaning of associated terms			5- Know and understand how to develop assessment items and tasks in order to ensure high quality in their validity and reliability	Explain to colleagues the role that these assessment terms play in school-based and other assessments
		6- Know and understand the meaning of reliability and validity in relation to selecting assessment tasks	Describe the importance of validity and reliability in classroom assessment		6-Know how to develop assessments that minimise sources of bias, which can distort assessment results	Develop assessments that have taken account of potential sources of bias such as that caused by differences in gender, or race

Standard	Practitioner teachers			Advanced teachers		
	Sub-Standard	Knowledge and understanding	Is able to ...	Sub-Standard	Knowledge and understanding	Is able to ...
10–2 Applying evaluation		7-Know and understand illegal, inappropriate, and unethical assessment practices	Explain the importance of assessment ethics and their impact on students’ performance		7-Know and understand and carry out legal and ethical responsibilities in assessment at the school and classroom levels	Support colleagues to be more knowledgeable in assessment ethics
	10–2-1 Evaluate students’ performance and learning using diagnostic, formative and summative evaluation methods. •Identifies and improves students’ strengths and challenges	1- Know and understand how to use a range of assessment items that are fit-for-purpose	Use a range of assessment items that effectively assess a range of skills in a unit of learning and mirror the relative weights of content materials or skills required by the unit of learning	10–2-2 Evaluate students’ performance and learning using the diagnostic formative and summative evaluation, identify and improve students’ strengths and challenges, and develop effective evaluation methods to support colleagues	1-Know how to write, develop and use a range of item types and assessment tasks that are technically sound, editorially correct and that appropriately assess the outcomes and performance standard levels being assessed	Explain how to support colleagues in relation to know assessment tasks (items) that meet the requirements for good quality assessment tasks

Standard	Practitioner teachers			Advanced teachers		
	Sub-Standard	Knowledge and understanding	Is able to ...	Sub-Standard	Knowledge and understanding	Is able to ...
		2-Know how to administer a range of assessments and understand the need for consistent instructions to ensure a consistent testing environment	Demonstrate through relevant examples knowledge of the requirements for administering classroom assessments and more formal assessments		2-Know the process for accurately marking assessments and collecting and recording student results	Give exemplars on how to collect and record students results in a systematic way and identify the potential problems with marking in a range of assessment types
		3- Know when and how to assess and report student progress against standards	Complete and interpret reports to stakeholders in a standards-referenced system		3- Know why comparability is needed at the school and classroom levels	Describe situations in schools where comparability is likely to be required. For example, it may be needed to compare results across similar grades in a school or compare results from students within a class who have done different assessments

Standard	Practitioner teachers			Advanced teachers		
	Sub-Standard	Knowledge and understanding	Is able to ...	Sub-Standard	Knowledge and understanding	Is able to ...
		4-Know and understand how to use diagnostic assessment to assess prior student knowledge	Demonstrate some of the ways to figure out prior student knowledge and performance		4- Know the content and purpose of major state, national and international testing programs	Describe the purposes, main characteristics and content of these programs
		5- Know and understand how to design summative assessments tools that will enable the results to be used to improve teaching and learning	Describe some of the ways that marks can be used in the learning process to improve learning, including for the purposes of helping teachers to adapt and adjust teaching to accommodate students' learning needs		5- Know various ways to record results to capture an image of student performance that can be compared against performance standards	Demonstrate ways that results of student assessments are recorded and used to arrive at an overall grade representing a performance standard

Standard	Practitioner teachers			Advanced teachers		
	Sub-Standard	Knowledge and understanding	Is able to ...	Sub-Standard	Knowledge and understanding	Is able to ...
		6- Know and understand the purposes of assessment results recording	Describe the main purposes of recording, including A. reporting achievement; B. monitoring strengths and weaknesses and recording progress		6-Know some of the more common ways of achieving comparability at the school level in a standards-referenced system and its strengths and weaknesses	Describe the way that consensus moderation and moderation using common tasks can be applied within schools to obtain comparability across classes and the strengths and weaknesses of consensus and common-task moderation
		7-Know and understand the requirements, structure, purpose and action of conducting formative assessment appropriately	Give some examples of formative assessment instruments and how to use them to inform teaching and learning		7- Know and understand some of the ways that assessment methods can be misused and misinterpreted	Help teachers by explaining and giving examples of how assessment methods can be misused and misinterpreted in the context of their original purpose
10-3 Engaging students in evaluation processes	10–3-1 Applies self-assessment strategies effectively and supports students to participate and	1- Know and understand the definition and the main purpose of self-assessment and peer assessment	Describe how to apply the self and peer assessment during the instruction in a positive manner	10–3-2 Apply self-assessment strategies on students, develops effective strategies that enhance students’	1-Know and understand the definition of self-regulation	Support teachers to use assessment tasks to increase student ability in terms of organising their learning process, such as by sharing

Standard	Practitioner teachers			Advanced teachers		
	Sub-Standard	Knowledge and understanding	Is able to ...	Sub-Standard	Knowledge and understanding	Is able to ...
	make use of its results		Discuss the processes and ultimate goals of applying self and peer assessments based in the research's outcomes	participation in the assessment process and supports colleagues with them	2-Use and develop assessment approaches to engage prior student knowledge and skills in lesson plans	standards, rubrics and criteria Explain some of the assessment methods that will benefit from prior student knowledge to ensure more participation in current practices
		2- Know and understand how to work with students to analyse assessment results to identify their strengths and weaknesses	Describe some ways to support students to benefit from their assessment results and monitor their learning		3-Know and understand many ways to encourage students to be involved in self- and peer assessment	Select, develop and apply some strategies to reinforce students' participation in self- and peer assessment
		3- Use assessment methods to identify current student performance, skills, achievement to be aligned with the lesson plan	Define and provide relevant examples of how assessment methods can empower students in new learning processes		4- Know and understand what it means to tailor assessments to the individual needs of all students	Develop assessments that have been adapted to accommodate students with different needs
		4- Know and understand how to develop student assessment skills	Describe some methods of improving students' assessment knowledge		5-Know and understand the importance of providing opportunity for	Support colleagues to explain how to use instruction environment to incorporate students in their learning by co-

Standard	Practitioner teachers			Advanced teachers		
	Sub-Standard	Knowledge and understanding	Is able to ...	Sub-Standard	Knowledge and understanding	Is able to ...
					students to construct criteria	constructing rubrics that can be used in assessment tasks
		5-Know and understand some feedback techniques to engage students in instruction	Describe how to improve the learning by using feedback to support students to adapt their learning strategies and to become independent learners		6-Sharing learning outcomes with students to ensure their standing before heading to the next steps	Help teachers to share with students the results of their works, such as tests, projects, worksheets and presentations
10–4 Employing evaluation results	10–4-1 Use the evaluation results to organise and modify study units, provide constructive feedback to students and parents in a timely manner, and pursue its applications	1- Know the definition of data literacy	Define data literacy as the ability to be able to understand and use data effectively	10–4-2 Use the evaluation results to organise and modify study units, provide constructive feedback to students and parents in a timely manner, and pursue its application. Develop effective methods aimed at improving	1- Know and understand some of the terms associated with classroom data	Define and describe terms associated with student assessment data, including: A. dichotomous (true / false) data; B. quantitative versus qualitative data; C. polytomously scored questions; and, D. Percentages and percentiles

Standard	Practitioner teachers			Advanced teachers		
	Sub-Standard	Knowledge and understanding	Is able to ...	Sub-Standard	Knowledge and understanding	Is able to ...
		2- Know and understand that there are a range of skills needed to transform data into information, evidence and strategies for teaching and learning	List some of the skills that are needed to convert data into actions for teaching and learning; identify, collect, analyse, summarise and prioritise data	learning and support colleagues with them	2-Know the strengths and weaknesses of some of the more common statistics used to summarise student performance	Describes for colleagues the strengths and weaknesses of, for example, A. mean (average), mode and median; B. standard deviation, variance and range; and, correlation
		3- Know and understand the meaning of basic statistical terms used for analysing and summarising student performance data	Interpret summary data commonly used to represent student performance. For example, statistical terms such as: mean, median and mode; variance and standard deviation; correlation; range; percentiles and norms		3-Know and understand how to generate teaching and learning interventions from an analysis of assessment data	Modify teaching practice according to the data collected
		4- Know the meaning of data, information and evidence, as well understanding that evidence is used to inform educational	Distinguish between data (responses), information (responses as marks), evidence (information considered through		4- Know and understand that different kinds of data exist and have a basic understanding of which kind of data	Develop effective methods that align the needs for decision-making with the appropriate kind of data. For example, monitoring performance at a

Standard	Practitioner teachers			Advanced teachers		
	Sub-Standard	Knowledge and understanding	Is able to ...	Sub-Standard	Knowledge and understanding	Is able to ...
		practice and reporting	the lens of outcomes or other educational aims). Identify where and how evidence is used in schools to inform teaching		is useful for particular decisions	summative level usually requires statistical data; monitoring the performance on a day-to-day basis generally requires more qualitative and anecdotal data
		5- Know and understand how to collect and organise data for classroom use	Effectively take raw response data and collect it in different forms (using technology, where appropriate) so that it can be used to evaluate student learning			
		6- Know the meaning of feedback in assessment	Describe the importance of feedback in classroom assessment		5-Define and provide relevant examples of how feedback is used in schools	Demonstrate how feedback from a range of assessments is meaningful in terms of student learning with respect to the assessed outcomes

Standard	Practitioner teachers			Advanced teachers		
	Sub-Standard	Knowledge and understanding	Is able to ...	Sub-Standard	Knowledge and understanding	Is able to ...
		7- Know and understand the features of effective feedback	Describe the features of feedback that make it an effective tool to assist student learning. For examples, it should be A. clear; B. purposeful; C. meaningful; D. timely; and E. compatible		6-Know and understand the requirements for effective feedback	Help teachers by providing examples of feedback to students that consider its timing and its appropriateness for the type of feedback given

Standard	Practitioner teachers			Advanced teachers		
	Sub-Standard	Knowledge and understanding	Is able to ...	Sub-Standard	Knowledge and understanding	Is able to ...
			Give some examples of feedback that illustrate some of the most effective delivery modes. For example, feedback should be			
		8- Know and understand some of the main requirements for delivering effective feedback	A. delivered in an appropriate setting; B. focused on performance and not the individual; C. delivered using neutral, non-judgemental language; D. focused on highlighting areas for improvement and suggesting strategies to improve; and, aimed to empower and encourage learning		7-Know and understand how feedback should be tailored to the needs of individual students	Design and administer a range of feedback strategies that can be used in class to accommodate individual student needs and at the same time take into account the nature of the task

Standard	Practitioner teachers			Advanced teachers		
	Sub-Standard	Knowledge and understanding	Is able to ...	Sub-Standard	Knowledge and understanding	Is able to ...
		9- Know the impact that feedback has on learning in classrooms	Discuss some of the outcomes from research related to the impact that feedback has on learning in classrooms		8- Know how to assess whether the feedback that is given in class is understood by the students	Describe ways to evaluate whether the feedback has been understood by the students. Some of the ways that this can be done include: interacting with students to ensure there is a shared understanding of terminology and language; and, reflecting on how the feedback that has been given may assist students in improving their learning

Standard	Practitioner teachers			Advanced teachers		
	Sub-Standard	Knowledge and understanding	Is able to ...	Sub-Standard	Knowledge and understanding	Is able to ...
10–5 Preparing evaluation reports	10–5-1 provide clear and accurate learning reports to students and parents using accurate and reliable technology and record students’ learning outcomes	1- Know and understand the purposes of reporting	Describe the main purposes of reporting, including A. reporting achievement; B. monitoring strengths and weaknesses; and, recording progress	10–5-2 Provide clear and accurate learning reports to students and parents using accurate and reliable technology and record students’ learning outcomes. Develop effective methods of preparing evaluation reports and support colleagues with them	1- Knows and understands how to prepare written reports that report accurately, clearly and respectfully to parents/carers	Prepare reports that A. use accurate, appropriate language to report skills and progress; B. use language that is respectful and appropriate to audiences; C. bases qualitative feedback on a range of assessments; and, uses data which assesses academic, skill, social and emotional progress as appropriate

Standard	Practitioner teachers			Advanced teachers		
	Sub-Standard	Knowledge and understanding	Is able to ...	Sub-Standard	Knowledge and understanding	Is able to ...
		2- Know and understand some of the more common reporting strategies used in schools for disseminating assessment information to parents	Describe the characteristics (including strengths and weaknesses) of strategies for reporting to parents, including A. written reports; B. parent-teacher meetings and interviews; and, portfolios and e-Portfolios Explain the meaning of the various grades/marks that characterise most school reports		2-Knows and understands the importance of collecting reliable and valid assessment data about the achievement of students and reports it in a manner which is consistent with the reporting strategy(ies) used within a school	Explain to teachers why it is necessary to collect, organise and report data in an accurate, consistent and meaningful way, and that selecting reporting formats as appropriate to the task and assessment policy is important Describe and explain for teachers
		3- Knows and understands the meaning of the grades that are used to summarise performance	A. A to E grades; B. Qualitative grades; C. Benchmark-related grades; D. Marks; E. Ranks; F. Percentages and, Percentiles		3- Knows and understands the meaning of the grades/marks on the reports	A. the process for developing grades and marks that have been collected to stakeholders; B. the process of reporting within scope of subject/syllabus being assessed; C. that the sequencing of learning is linked to reporting;

Standard	Practitioner teachers			Advanced teachers		
	Sub-Standard	Knowledge and understanding	Is able to ...	Sub-Standard	Knowledge and understanding	Is able to ...
						D. appropriate next steps / aims for learning are identified; and, the technical specifications for formal assessments (e.g. mean, standard deviation, reliability, validity) is explained to parents in a manner that helps them better interpret the report
				4- Know and understand the strengths and weaknesses of different ways of recording data for reporting purposes	Support teachers to know the strengths and weaknesses of different levels of reporting for various purposes and contexts. For example, A. Student-item level; B. Summary-task level; C. Qualitative-descriptor level; and, Video or photograph	

Appendix E: The NPST

Standard	practitioner teachers- Sub-Standard	Advanced teacher- Sub-Standard
10.1 preparing evaluation tools	prepare a diverse and comprehensive set of direct and in direct evaluation tools	prepare a variety of direct and in direct evaluation tools and develops effective evaluation tools to support his colleagues
10-2 Applying evaluation	evaluate student' performance and learning using the diagnostic formative and summative evaluation, identifies and improve students' strengths and challenges	evaluate student' performance and learning using the diagnostic formative and summative evaluation, identifies and improve students' strengths and challenges, and develop effective evaluation methods to support his colleagues
10-3 Engaging students in evaluation processes	apply self-evaluation strategies effectively and supports students to participate and making use of its results	apply self- evaluation strategies on students, develops effective strategies that enhance students' participation in the evaluation process, and support his colleagues with them

10-4 employing evaluation results	use the evaluation results to organize and modify study units, provide constructive feedback to students and parents in a timely manner, and pursue its applications	use evaluation results to organize and modify study units, provide constructive feedback to students and parents in a timely manner , and pursue its application. He develops effective methods aimed at improving learning and supports his colleagues with them
10-5 preparing Evaluation Reports	provide clear and accurate learning reports to students and parents using accurate and reliable technology and records about students' learning outcomes	provide clear and accurate learning reports to students and parents using accurate and reliable technology and records about students' learning outcomes. He develops effective methods of preparing evaluation reports and supports his colleagues with them

Appendix F: The Original TALQ

Teacher Assessment Literacy Questionnaire

Developed by

Barbara S. Plake

James C. Impara

University of Nebraska-Lincoln

In cooperation with

The National Council on Measurement in Education

W.K. Kellogg Foundation

Brief description of instrument:

This questionnaire is divided into two parts. Part I consists of 35 items related to the seven Standards for Teacher Competence in the Educational Assessment of Students. Some of the items are intended to measure general concepts related to testing and assessment, including the use of assessment activities for assigning student grades and communicating the results of assessment to students and parents; other items are related to knowledge of standardized testing and the remaining items are related to classroom assessment. Part II consists of items intended to find out about how you think assessments should be used in the classroom, your level of teaching experience, and so on.

Directions for Administration:

Please read each item carefully and select the response you think is the best one. If you think you know which is best, even if you are not positive, mark that response.

KEY FOR TEACHER ASSESSMENT LITERACY TEST

1. C
2. B
3. B
4. B
5. C
6. B
7. C
8. A
9. D
10. D
11. A
12. D
13. D
14. B
15. A
16. C
17. B
18. B
19. C
20. A
21. B
22. D
23. C
24. D
25. B
26. C
27. A
28. D
29. A
30. A
31. D
32. A
33. C
34. C
35. B

Standard 1: Choosing assessment methods.

1. What is the most important consideration in choosing a method for assessing student achievement?
 - a. Ease of scoring the assessment.
 - b. Ease of preparing the method of assessment.
 - c. Accuracy of assessing attainment of instructional objectives.

- d. Acceptance by the school administration.
2. When scores from a standardized test are said to be reliable, what does it imply?
- a. Student scores from the test can be used for a large number of educational decisions.
 - b. If a student retook the same test, he or she would get a similar score on each retake.
 - c. The test score is a more valid measure than teacher judgments.
 - d. The test score accurately reflects the content of instruction in classes where the test is administered.
3. Mrs. Bruce wished to assess her students' understanding of the method of problem solving she had been teaching. Which assessment strategy below would be most valid?
- a. Select a textbook that has a "teacher's guide" with a test developed by the authors.
 - b. Develop an assessment consistent with an outline of what she has actually taught in the class.
 - c. Select a standardized test that provides a score on problem solving skills.
 - d. Select an instrument that measures students' attitudes about problem solving strategies.
4. What is the most effective use a teacher can make of an assessment strategy that requires students to show their work , e.g., the way they arrived at a solution to a problem or the logic used to arrive at a conclusion?
- a. Assigning grades for a unit of instruction on problem solving.
 - b. Providing instructional feedback to individual students.
 - c. Motivating students to attempt innovative ways to solve problems.
 - d. None of the above.

5. Ms. Green, the principal, was evaluating the teaching performance of Mr. Wajesa, the fourth grade teacher. One of the things Ms. Green wanted to learn was if the students were being encouraged to use higher order thinking skills in the class. What documentation would be the most valid to help Ms. Green to make this decision?
- a. Mr. Wajesa's lesson plans.
 - b. The state curriculum guides for fourth grade.
 - c. Copies of Mr. Wajesa's unit tests or assessment strategies used to assign grades.
 - d. Worksheets completed by Mr. Wajesa's students, but not used for grading.

Standard 2: Developing assessment methods

6. A teacher wants to document the validity of the scores from a classroom assessment strategy she plans to use for assigning grades on a class unit. What kind of information would provide the best evidence for this purpose?
- a. Have other teachers judge whether the assessment strategy covers what was taught.
 - b. Match an outline of the instructional content to the content of the assessment strategy.
 - c. Let students in the class indicate if they thought the assessment was valid.
 - d. Ask parents if the assessment reflects important learning outcomes.
7. Which of the following actions would most likely increase the reliability of Mrs. Lockwood's multiple choice end-of-unit examination in physical science?
- a. Use a blueprint to develop the test questions.
 - b. Change the test format to true-false questions.
 - c. Add more items like those already in the test.
 - d. Add an essay component.
8. Ms. Guardia wants to assess her students' skills in organizing ideas rather than just repeating facts. Which words should she use in formulating essay exercises to achieve this goal?

- a. compare, contrast, criticize
 - b. identify, specify, list
 - c. order, match, select
 - d. define, recall, restate
9. Mr. Woodruff wanted his students to appreciate the literary works of Edgar Allen Poe. Which of his test items shown below will best measure his instructional goal?
- a. "Spoke the raven, nevermore." comes from which of Poe's works?
 - b. True or False: Poe was an orphan and never knew his biological parents.
 - c. Edgar Allen Poe wrote:
 - 1. Novels
 - 2. Short stories
 - 3. Poems
 - 4. All of the above.
 - d. Discuss briefly your view of Poe's contribution to American literature.
10. Several students in Ms. Atwell's class got low scores on her end-of-unit test in doing multi-step story problems in mathematics. She wanted to know which students were having similar problems so she could group them for instruction. Which assessment strategy would be best for her to use for grouping students?
- a. Use the test provided in the "teacher's guide."
 - b. Have the students take a test that has separate items for each step of the process.
 - c. Look at the student's records and standardized test scores to see which topics the students had not performed well on previously.
 - d. Give students story problems to complete and have them show their work.

Standard 3: Administering, scoring and interpreting assessment results.

11. Many teachers score classroom tests using a 100-point percent correct scale. In general, what does a student's score of 90 on such a scale mean?

- a. The student answered 90% of the items on this test correctly.
- b. The student knows 90% of the instructional content of the unit covered by this test.
- c. The student scored higher than 90% of all the students who took the test.
- d. The student scored 90% higher than the average student in the class.

12. Students in Mr. Jakman's science class are required to develop a model of the solar system as part of their end of unit grade. Which scoring procedure below will maximize the objectivity of assessing these student projects?
- a. When the models are turned in, Mr Jakman identifies the most attractive models and gives them the highest grades, the next most attractive get a lower grade and so on.
 - b. Mr. Jakman asks other teachers in the building to rate each project on a 5 point scale based on their quality.
 - c. Before the projects are turned in, Mr. Jakman constructs a scoring key based on the critical features of the projects as identified by the highest performing students in the class.
 - d. Before the projects are turned in, Mr. Jakman prepares a model or blueprint of the critical features of the product and assigns scoring weights to these features. The models with the highest scores receive the highest grade.
13. At the close of the first month of school, Mrs. Friend gives her fifth grade students a test she developed in social studies. Her test is modeled after a standardized social studies test. It presents passages and then asks questions related to understanding and problem definition. When the test was scored, she noticed that two of her students who had been performing well in their class assignments, scored a lot lower than other students. Of the following types of additional information which would be most helpful in interpreting the results of this test?
- a. The gender of the students.
 - b. The age of the students.
 - c. Reliability data for the standardized social studies test she used as the model.
 - d. Reading comprehension scores for the students.
14. Frank, a beginning fifth grader, received a G. E. (grade equivalent score) of 8.0 on the Reading Comprehension subtest of a standardized test. This score should be interpreted to mean that Frank:
- a. can read and understand 8th grade reading level material.
 - b. scored as well as a typical beginning 8th grader scored on this test.
 - c. is performing in Reading Comprehension at the 8th grade level.
 - d. will probably reach maximum performance in Reading Comprehension at the beginning of the 8th grade.

15. When the directions indicate each section of a standardized test is timed separately, which of the following is acceptable test-taking behavior?
- a. John finishes the vocabulary section early; he then rechecks many of his answers in that section.
 - b. Mary finishes the vocabulary section early; she checks her answers on the previous test section.
 - c. Jane finishes the vocabulary section early; she looks ahead at the next test section but does not mark her answer sheet for any of those items.
 - d. Bob did not finish the vocabulary section; he continues to work on that section when the testing time is up.

Standard 4: Using assessment results for decision making.

16. Ms. Camp is starting a new semester with a factoring unit in her Algebra I class. Before beginning the unit, she gives her students a test on the commutative, associative, and distributive properties of addition and multiplication. Which of the following is the most likely reason she gives this test to her students?
- a. The principal needs to report the results of this assessment to the state testing director.
 - b. Ms. Camp wants to give the students practice in taking tests early in the semester.
 - c. Ms. Camp wants to check for prerequisite knowledge in her students before she begins the unit on factoring.
 - d. Ms. Camp wants to measure growth in student achievement of these concepts, and scores on this test will serve as the students' knowledge baseline.
17. To evaluate the effectiveness of the mathematics program for her gifted first graders, Ms. Allen gave them a standardized mathematics test normed on third graders. To decide how well her students performed, Ms. Allen compared her students' scores to those of the third-grade norm group. Why is this an incorrect application of standardized test norms?
- a. The norms are not reliable for first graders.
 - b. The norms are not valid for first graders.

- c. Third grade mathematics items are too difficult for first graders.
 - d. The time limits are too short for first graders.
18. When planning classroom instruction for a unit on arithmetic operations with fractions, which of these types of information have more potential to be helpful?

norm-referenced information: describes each student's performance relative to a other students in a group (e.g. percentile ranks, stanines), or

criterion-referenced information: describes each student's performance in terms of status on specific learning outcomes (e.g., number of items correctly answered for each specific objective)

- a. Norm-referenced information.
 - b. Criterion-referenced information.
 - c. Both types of information are equally useful in helping to plan for instruction.
 - d. Neither, test information is not useful in helping to plan instruction.
19. Students' scores on standardized tests are sometimes inconsistent with their performances on classroom assessments, e.g. teacher tests or other in-class activities. Which of the following is **NOT** a reasonable explanation for such discrepancies?
- a. Some students freeze up on standardized tests, but they do fine on classroom assessments.
 - b. Students often take standardized test less seriously than they take classroom assessment.
 - c. Standardized tests measure only recall of information while classroom assessments measure more complex thinking.
 - d. Standardized tests may have less curriculum validity than classroom assessment
20. Elementary school teachers in the Baker School system collectively designed and developed a new curricula in Reading, Mathematics, and Science that is based on locally developed objectives and objectives in state curriculum guides. The new curricula were not matched directly to the content of the fourth grade standardized test. A newspaper reports the fourth grade students in Baker Public Schools are among the lowest scoring districts in the State Assessment Program. Which of the following would **invalidate** the comparison between Baker Public Schools and other schools in the state?

- a. The curriculum objectives of the other districts may more closely match those of the State Assessment.
- b. Other school systems did not design their curriculum to be consistent with the State Assessment test.
- c. Instruction in Baker schools is poor.
- d. Other school systems have different promotion policies than Baker.

Standard 5: Using assessment in grading

21. Of the following, which choice typically provides the most reliable student-performance information a teacher might consider when assigning a unit grade?
- a. Scores from a teacher-made test containing two or three essay questions related directly to instructional objectives of the unit.
 - b. Scores from a teacher-made 20 item multiple-choice test designed to measure the specific instructional objectives of the unit.
 - c. Oral responses to questions asked in class of each student over the course of the unit.
 - d. Daily grades designed to indicate the quality of in-class participation during regular instruction.
22. A teacher gave three tests during a grading period and she wants to weight them all equally when assigning grades. The goal of the grading program is to rank order students on achievement. In order to achieve this goal, which of the following should be closest to equal?
- a. Number of items.
 - b. Number of students taking each test.
 - c. Average scores.
 - d. Variation (range) of scores.
23. When a parent asks a teacher to explain the basis for his or her child's grade, the teacher should:

- a. explain that the grades are assigned fairly, based on the student's performance and other related factors.
 - b. ask the parents what they think should be the basis for the child's grade.
 - c. explain exactly how the grade was determined and show the parent samples of the student's work.
 - d. indicate that the grading scale is imposed by the school board and the teachers have no control over grades.
24. Which of the following grading practices results in a grade that **least** reflects students' achievement?
- a. Mr. Jones requires students to turn in homework; however, he only grades the odd numbered items.
 - b. Mrs. Brown uses weekly quizzes and three major examinations to assign final grades in her class.
 - c. Ms. Smith permits students to redo their assignments several times if they need more opportunities to meet her standards for grades.
 - d. Miss Engle deducts 5 points from a student's test grade for disruptive behavior.
25. During the most recent grading period Ms. Johnson graded no homework and gave only one end-of-unit test. Grades were assigned only on the basis of the test. Which of the following is the major criticism of how she assigned the grades?
- a. The grades probably reflect a bias against minority students that exists in most tests.
 - b. Decisions like grade assignment should be based on more than one piece of information.
 - c. The test was too narrow in curriculum focus.
 - d. There is no significant criticism of this method providing the test covered the unit's content.

Standard 6: Communicating assessment results.

26. In a routine conference with Mary's parents, Mrs. Estes observed that Mary's scores on the state assessment program's quantitative reasoning tests indicate Mary is performing better in mathematics concepts than in mathematics computation. This probably means that:

- a. Mary's score on the computation test was below average.
- b. Mary is an excellent student in mathematics concepts.
- c. The percentile bands for the mathematics concepts and computation tests do not overlap.
- d. The mathematics concepts test is a more valid measure of Mary's quantitative reasoning ability.

27. Many states are revising their school accountability programs to help explain differences in test scores across school systems. Which of the following is **NOT** something that needs to be considered in such a program?

- a. The number of students in each school system.
- b. The average socio-economic status of the school systems.
- c. The race/ethnic distribution of students in each school system.
- d. The drop-out rate in each school systems.

28. The following standardized test data are reported for John.

<u>Subject</u>	<u>Stanine Score</u>
Vocabulary	7
Mathematics Computation	7
Social Studies	7

Which of the following is a valid interpretation of this score report?

- a. John answered correctly the same number of items on each of the three tests.
- b. John's test scores are equivalent to a typical seventh grader's test performance.
- c. John had the same percentile rank on the three tests.
- d. John scored above average on each of the three tests.

29. Mr. Klein bases his students' grades mostly on graded homework and tests. Mr. Kaplan bases his students' grades mostly on his observation of the students during

class. A major difference in these two assessment strategies for assigning grades can best be summarized as a difference in:

- a. formal and informal assessment.
 - b. performance and applied assessment.
 - c. customized and tailored assessment.
 - d. formative and summative assessment.
30. John scored at the 60th percentile on a mathematics concepts test and scored at the 57th percentile on a test of reading comprehension. If the percentile bands for each test are five percentile ranks wide, what should John's teacher do in light of these test results?
- a. Ignore this difference.
 - b. Provide John with individual help in reading.
 - c. Motivate John to read more extensively outside of school.
 - d. Provide enrichment experiences for John in mathematics, his better performance area.

Standard 7: Recognizing unethical practices.

31. In some states testing companies are required to release items from prior versions of a test to anyone who requests them. Such requirements are known as:
- a. open-testing mandates.
 - b. gag rules.
 - c. freedom-of-information acts.
 - d. truth-in-testing laws.
32. Mrs. Brown wants to let her students know how they did on their test as quickly as possible. She tells her students that their scored tests will be on a chair outside of her room immediately after school. The students may come by and pick out their graded test from among the other tests for their class. What is wrong with Mrs. Brown's action?

- a. The students can see the other students' graded tests, making it a violation of the students' right of privacy.
 - b. The students have to wait until after school, so the action is unfair to students who have to leave immediately after school.
 - c. Mrs. Brown will have to rush to get the tests graded by the end of the school day, hence, the action prevents her from using the test to identify students who need special help.
 - d. The students who were absent will have an unfair advantage, because her action allows the possibility for these students to cheat.
33. A state uses its statewide testing program as a basis for distributing resources to school systems. To establish an equitable distribution plan, the criterion set by the State Board of Education provides additional resources to every school system with student achievement test scores above the state average. Which cliché best describes the likely outcome of this regulation?
- a. Every cloud has its silver lining.
 - b. Into each life some rain must fall.
 - c. The rich get richer and the poor get poorer.
 - d. A bird in the hand is worth two in the bush.
34. In a school where teacher evaluations are based in part on their students' scores on a standardized test, several teachers noted that one of their students did not reach some vocabulary items on a standardized test. Which teacher's actions is considered ethical?
- a. Mr. Jackson darkened circles on the answer sheet at random. He assumed Fred, who was not a good student, would just guess at the answers, so this would be a fair way to obtain Fred's score on the test.
 - b. Mr. Hoover filled in the answer sheet the way he thought Joan, who was not feeling well, would have answered based on Joan's typical in-class performance.
 - c. Mr. Stover turned in the answer sheet as it was, even though he thought George, an average student, might have gotten a higher score had he finished the test.
 - d. Mr. Lund read each question and darkened in the bubbles on the answer sheet that represented what he believed Felicia, a slightly below average student, would select as the correct answers.
35. Mrs. Overton was concerned that her students would not do well on the State Assessment Program to be administered in the Spring. She got a copy of the

standardized test form that was going to be used. She did each of the following activities to help increase scores. Which activity was unethical?

- a. Instructed students in strategies on taking multiple choice tests, including how to use answer sheets.
- b. Gave students the items from an alternate form of the test.
- c. Planned instruction to focus on the concepts covered in the test.
- d. None of these actions are unethical.

Appendix G: The Modified Teacher Assessment Literacy Questionnaire

The MTALQ

PART I

- Teacher's Demographic Information

1- Specify your Gender?

- Male

- Female

2-school levels:

-Elementary

-Middle

-Secondary

3-What is your highest academic qualification?

-Diploma

-Bachelor

-Master

-Doctoral

4-Career Level:

-practitioner teacher

-advanced teacher

5-How many years experience as a classroom teacher do you have?

-1-5 years ago

-6-10 years ago

-More than ten years ago

-I don't remember

6-Have you ever taken a training course in pre-service or in-service in the field of educational assessment, measurement and evaluation?

-Yes

-No

7-How many classes do you give each week?

PART II

Standard 1: Choosing assessment methods.

1. What is the most important consideration in choosing a method for assessing student achievement?
 - a. Ease of scoring the assessment.
 - b. Ease of preparing the method of assessment.
 - c. Accuracy of assessing attainment of instructional objectives.
 - d. Acceptance by the school administration.
2. When scores from a standardized test are said to be reliable, what does it imply?
 - a. Student scores from the test can be used for a large number of educational decisions.

- b. If a student retook the same test, he or she would get a similar score on each retake.
 - c. The test score is a more valid measure than teacher judgments.
 - d. The test score accurately reflects the content of instruction in classes where the test is administered.

- 3. Mrs. Bruce wished to assess her students' understanding of the method of problem solving she had been teaching. Which assessment strategy below would be most valid?
 - a. Select a textbook that has a "teacher's guide" with a test developed by the authors.
 - b. Develop an assessment consistent with an outline of what she has actually taught in the class.
 - c. Select a standardized test that provides a score on problem solving skills.
 - d. Select an instrument that measures students' attitudes about problem solving strategies.

- 4. What is the most effective use a teacher can make of an assessment strategy that requires students to show their work , e.g., the way they arrived at a solution to a problem or the logic used to arrive at a conclusion?
 - a. Assigning grades for a unit of instruction on problem solving.
 - b. Providing instructional feedback to individual students.
 - c. Motivating students to attempt innovative ways to solve problems.
 - d. None of the above.

- 5. Ms. Green, the principal, was evaluating the teaching performance of Mr. Wajesa, the fourth grade teacher. One of the things Ms. Green wanted to learn was if the students were being demonstrating higher order thinking skills. What documentation would be the most valid to help Ms. Green to make this decision?
 - a. Mr. Wajesa's lesson plans.
 - b. The state curriculum guides for fourth grade.
 - c. Copies of Mr. Wajesa's unit tests or assessment strategies used to assign grades.
 - d. Worksheets completed by Mr. Wajesa's students, but not used for grading.

6 - A principal wants to assess a ninth-grade teacher's ability to use "formative" assessment. The teacher uses peer and self-assessment as part of the set of "formative assessments". The principal should make it clear to the teacher that:

- a. there is a difference between formative assessment, peer-assessment, and self-assessment.
- b. formative assessment does not need to be planned and occurs spontaneously during the lesson.
- c. peer-assessment and self-assessment are inappropriate for the age group of students in this class.
- d. these actions are consistent with the use of formative assessment.

Standard 2: Developing assessment methods

7. A teacher wants to document the validity of the scores from a classroom assessment strategy she plans to use for assigning grades on a class unit. What kind of information would provide the best evidence for this purpose?

- a. Have other teachers judge whether the assessment strategy covers what was taught.
- b. Match an outline of the instructional content to the content of the assessment strategy.
- c. Let students in the class indicate if they thought the assessment was valid.
- d. Ask parents if the assessment reflects important learning outcomes.

8. Which of the following actions would most likely increase the reliability of Mrs. Lockwood's multiple choice end-of-unit examination in physical science?

- a. Use a blueprint to develop the test questions.
- b. Change the test format to true-false questions.
- c. Add more items like those already in the test.
- d. Add an essay component.

9. Ms. Guardia wants to assess her students' skills in generating ideas rather than just repeating facts. Which words should she use in formulating essay exercises to achieve this goal?

- a. compare, contrast, criticize
- b. identify, specify, list
- c. order, match, select
- d. define, recall, restate

10. Mr. Woodruff wanted his students to appreciate the literary works of Edgar Allen Poe. Which of his test items shown below will best measure his instructional goal?

- a. "Spoke the raven, nevermore." comes from which of Poe's works?
- b. True or False: Poe was an orphan and never knew his biological parents.
- c. Edgar Allen Poe wrote:
 - 1. Novels
 - 2. Short stories
 - 3. Poems
 - 4. All of the above.
- d. Discuss briefly your view of Poe's contribution to American literature.

11. Several students in Ms. Atwell's class got low scores on her end-of-unit test in doing multi-step story problems in mathematics. She wanted to know which students were having similar problems so she could group them for instruction. Which assessment strategy would be best for her to use for grouping students?

- a. Use the test provided in the "teacher's guide."
- b. Have the students take a test that has separate items for each step of the process.
- c. Look at the standardized test scores to see which topics the students had not performed well on previously.
- d. Give students story problems to complete and have them show their work.

12. When a teacher administers an assessment at the end of an educational unit or semester for reporting home to parents, this test is considered to be:

- a. Assessment for learning.
- b. Assessment of learning.
- c. Formative assessment.

- d. None of the above.

Standard 3: Administering, scoring and interpreting assessment results.

13. Many teachers score classroom tests using a 100-point percent correct scale. In general, what does a student's score of 90 on such a scale mean?
- a. The student answered 90% of the items on this test correctly.
 - b. The student knows 90% of the instructional content of the unit covered by this test.
 - c. The student scored higher than 90% of all the students who took the test.
 - d. The student scored 90% higher than the average student in the class.
14. Students in Mr. Jakman's science class are required to develop a model of the solar system as part of their end of unit grade. Which scoring procedure below will maximize the objectivity of assessing these student projects?
- a. When the models are turned in, Mr Jakman identifies the most attractive models and gives them the highest grades, the next most attractive get a lower grade and so on.
 - b. Mr. Jakman asks other teachers in the building to rate each project on a 5 point scale based on their quality.
 - c. Before the projects are turned in, Mr. Jakman constructs a scoring key based on the critical features of the projects as identified by the highest performing students in the class.
 - d. Before the projects are turned in, Mr. Jakman prepares a model or blueprint of the critical features of the product and assigns scoring weights to these features. The models with the highest scores receive the highest grade.
15. At the close of the first month of school, Mrs. Friend gives her fifth grade students a test she developed in social studies. Her test is modeled after a standardized social studies test. It presents passages and then asks questions related to understanding and problem definition. When the test was scored, she noticed that two of her students who had been performing well in their class assignments, scored a lot lower than other students. Of the following types of additional information which would be most helpful in interpreting the results of this test?
- a. The gender of the students.

- b. The age of the students.
 - c. Reliability data for the standardized social studies test she used as the model.
 - d. Reading comprehension scores for the students.
16. In a school where teacher evaluations are based in part on their students' scores on a standardized test, several teachers noted that one of their students did not reach some vocabulary items on a standardized test. Which teacher's actions is considered ethical?
- a. Mr. Jackson darkened circles on the answer sheet at random. He assumed Fred, who was not a good student, would just guess at the answers, so this would be a fair way to obtain Fred's score on the test.
 - b. Mr. Hoover filled in the answer sheet the way he thought Joan, who was not feeling well, would have answered based on Joan's typical in-class performance.
 - c. Mr. Stover turned in the answer sheet as it was, even though he thought George, an average student, might have gotten a higher score had he finished the test.
 - d. Mr. Lund read each question and darkened in the bubbles on the answer sheet that represented what he believed Felicia, a slightly below average student, would select as the correct answers.
17. When the directions indicate each section of a standardized test is timed separately, which of the following is acceptable test-taking behavior?
- a. John finishes the vocabulary section early; he then rechecks many of his answers in that section.
 - b. Mary finishes the vocabulary section early; she checks her answers on the previous test section.
 - c. Jane finishes the vocabulary section early; she looks ahead at the next test section but does not mark her answer sheet for any of those items.
 - d. Bob did not finish the vocabulary section; he continues to work on that section when the testing time is up.

Standard 4: Using assessment results for decision making.

18. Before planning a lesson, a teacher wants to know what knowledge and skills the students already have in the unit of work. Which of the following would be the best way to find out?

- a. Do nothing as the teacher has already given instruction on the pre-requisite skills.
 - b. Review the past lessons homework to understand students depth of learning.
 - c. Conduct structured interviews with each student.
 - d. Administer a short quiz.
19. To evaluate the effectiveness of the mathematics program for her gifted first graders, Ms. Allen gave them a standardized mathematics test normed on third graders. To decide how well her students performed, Ms. Allen compared her students' scores to those of the third-grade norm group. Why is this an incorrect application of standardized test norms?
- a. The norms are not reliable for first graders.
 - b. The norms are not valid for first graders.
 - c. Third grade mathematics items are too difficult for first graders.
 - d. The time limits are too short for first graders.
20. When planning classroom instruction for a unit on arithmetic operations with fractions, which of these types of information have more potential to be helpful?
- norm-referenced* information: describes each student's performance relative to a other students in a group (e.g. percentile ranks, stanines), or
- criterion-referenced* information: describes each student's performance in terms of status on specific learning outcomes (e.g., number of items correctly answered for each specific objective)
- a. Norm-referenced information.
 - b. Criterion-referenced information.
 - c. Both types of information are equally useful in helping to plan for instruction.
 - d. Neither, test information is not useful in helping to plan instruction.
21. The best ways to find out the variation of students' grades is to calculate the
- a. mean.
 - b. median.
 - c. total score for each student.
 - d. standard deviation.

22. Elementary school teachers in the Baker School system collectively designed and developed a new curriculum in Reading, Mathematics, and Science that is based on locally developed objectives and objectives in state curriculum guides. The new curricula were not matched directly to the content of the fourth grade standardized test. A newspaper reports the fourth grade students in Baker Public Schools are among the lowest scoring districts in the State Assessment Program. Which of the following would **invalidate** the comparison between Baker Public Schools and other schools in the state?
- a. The curriculum objectives of the other districts may more closely match those of the State Assessment.
 - b. Other school systems did not design their curriculum to be consistent with the State Assessment test.
 - c. Instruction in Baker schools is poor.
 - d. Other school systems have different promotion policies than Baker.

Standard 5: Using assessment in grading

23. Of the following, which choice typically provides the most reliable student-performance information a teacher might consider when assigning a unit grade?
- a. Scores from a teacher-made test containing two or three essay questions related directly to instructional objectives of the unit.
 - b. Scores from a teacher-made 20 item multiple-choice test designed to measure the specific instructional objectives of the unit.
 - c. Oral responses to questions asked in class of each student over the course of the unit.
 - d. Daily grades designed to indicate the quality of in-class participation during regular instruction.
24. A teacher gave three tests during a grading period and she wants to weight them all equally when assigning grades. The goal of the grading program is to rank order students on achievement. In order to achieve this goal, which of the following should be closest to equal?
- a. Number of items.
 - b. Number of students taking each test.

- c. Average scores.
 - d. Variation (range) of scores.
25. When a parent asks a teacher to explain the basis for his or her child's grade, the teacher should:
- a. explain that the grades are assigned fairly, based on the student's performance and other related factors.
 - b. ask the parents what they think should be the basis for the child's grade.
 - c. explain exactly how the grade was determined and show the parent samples of the student's work.
 - d. indicate that the grading scale is imposed by the school board and the teachers have no control over grades.
26. Which of the following grading practices results in a grade that **least** reflects students' achievement?
- a. Mr. Jones requires students to turn in homework; however, he only grades the odd numbered items.
 - b. Mrs. Brown uses weekly quizzes and three major examinations to assign final grades in her class.
 - c. Ms. Smith permits students to redo their assignments several times if they need more opportunities to meet her standards for grades.
 - d. Miss Engle deducts 5 points from a student's test grade for disruptive behavior.
27. During the most recent grading period Ms. Johnson graded no homework and gave only one end-of-unit test. Grades were assigned only on the basis of the test. Which of the following is the major criticism of how she assigned the grades?
- a. The grades probably reflect a bias against minority students that exists in most tests.
 - b. Decisions like grade assignment should be based on more than one piece of information.
 - c. The test was too narrow in curriculum focus.
 - d. There is no significant criticism of this method providing the test covered the unit's content.

28. What kind of data can be used to carry out statistical analysis? the MOST correct answer is:

- a. Qualitative and quantitative.
- b. Quantitative.
- c. Qualitative.
- d. None of the above.

Standard 6: Communicating assessment results.

29. Many states are revising their school accountability programs to help explain differences in test scores across school systems. Which of the following is **NOT** something that needs to be considered in such a program?

- a. The number of students in each school system.
- b. The average socio-economic status of the school systems.
- c. The race/ethnic distribution of students in each school system.
- d. The drop-out rate in each school systems.

30. The purpose of administering PISA in Saudi Arabia is:

- a. for the teacher to get a more complete image of the students' achievement.
- b. to collect information on the efficacy of educational policy.
- c. to generate competition for acclaim internationally.
- d. to diagnose teaching and learning deficiencies.

31. Mrs. Brown wants to let her students know how they did on their test as quickly as possible. She tells her students that their scored tests will be on a chair outside of her room immediately after school. The students may come by and pick out their graded test from among the other tests for their class. What is wrong with Mrs. Brown's action?

- a. The students can see the other students' graded tests, making it a violation of the students' right of privacy.
- b. The students have to wait until after school, so the action is unfair to students who have to leave immediately after school.
- c. Mrs. Brown will have to rush to get the tests graded by the end of the school day, hence, the action prevents her from using the test to identify students who need special help.

- d. The students who were absent will have an unfair advantage, because her action allows the possibility for these students to cheat
32. Mr. Klein bases his students' grades mostly on graded homework and tests. Mr. Kaplan bases his students' grades mostly on his observation of the students during class. A major difference in these two assessment strategies for assigning grades can best be summarized as a difference in:
- a. formal and informal assessment.
 - b. performance and applied assessment.
 - c. customized and tailored assessment.
 - d. formative and summative assessment.

Standard 7: modern assessment. Or (relevant knowledge and skills for assessment).

33. A teacher uses rubrics to assess students' performance in developing and delivering a speech in front of their classmates.
- a. There is nothing wrong with this approach.
 - b. The teacher should have used direct observation rather than recording in order to fully observe the student.
 - c. The student can assess himself by asking his classmates about his speaking competence and informing the teacher.
 - d. Rubrics are only used in summative assessment.
34. A seventh-grade teacher develops students' assessment skills to support self-regulation. We can clearly state that:
- a. the student's self-regulation is concerned with emotional aspects of learning only.
 - b. developing the student's assessment skills increases self-regulation.
 - c. developing assessment skills is for the teacher, not for the student.
 - d. self-regulation of the student contradicts his independence in learning.
35. After the end of the unit on the history of ancient civilizations, the history teacher assigns each student a project that simulates these civilizations. The teacher should:
- a. assign all students the same project.
 - b. realise that project work is not an assessment tool.

- c. ask each student to design a project while adhering to unified criteria for grading.
 - d. None of the above
36. A teacher allows students to view and participate in building assessment criteria. Which of the following is MOST correct?
- a. Only the teacher should be aware of the assessment criteria, not the students.
 - b. The teacher and parents should be aware of and participate in building criteria.
 - c. The students' participation in building the assessment criteria could improve their learning outcomes.
 - d. Assessment criteria are NOT usually used in the classroom.

Appendix H: Invitation to School Principal

INVITATION TO SCHOOL PRINCIPAL TO ENCOURAGE TEACHERS TO PARTICIPATE IN THE RESEARCH



School of Education and Social Work

ABN 15 211 513 464

Chief Investigator
Professor Jim Tognolini

Room 333
Education Building A.35
The University of Sydney
NSW 2006 AUSTRALIA
Telephone: +61 2 86275205
Email: jim.tognolini@sydney.edu.au
Web: <http://www.sydney.edu.au/>

Dear Sir/Madam,

Re: Invitation to participate in the PhD research entitled, ‘An inquiry into teacher assessment literacy: the development of an assessment literacy framework and instrument to measure the assessment literacy of teachers in Saudi Arabia’

We are writing to invite your school teachers to participate in the above research. The research will be conducted by Professor Jim Tognolini, as Chief Researcher and Dr Jon Twing who is not from the University of Sydney, he is a member in Educational Measurement Interest group at the University of Sydney, and he is senior VP of Pearson Assessments, and Mr. Alzhrany Abdullah, as Co Researchers, from the School of Education and Social Work at the University of Sydney. This research is part of Mr. Alzhrany's PhD dissertation.

Mr Alzhrany is a Saudi citizen and a teacher from the Ministry of Education who is currently a Saudi government scholarship holder. The broad aim of his research is to develop a national framework for developing and validating a teacher assessment literacy questionnaire which can help measure the assessment literacy of Saudi teachers in public schools.

More specifically the study aims to:

- develop a framework that aligns the assessment knowledge and skills demanded of teachers (as described in the KSA Professional Teaching Standards) with increasing performance levels of assessment literacy;
- evaluate the efficacy of a current assessment literacy questionnaire (Teacher Assessment Literacy Questionnaire (TALQ) for measuring teacher performance in relation to the assessment literacy framework;
- use the framework to design an assessment literacy questionnaire that builds the TALQ and can be used to assess KSA teachers' assessment skills and knowledge in relation to the assessment literacy framework; and
- construct and validate an assessment literacy questionnaire for use with KSA teachers.

The proposed framework will specify what it is Saudi teachers need to know and be able to do to demonstrate proficiency in relation to the assessment standards prescribed in the Saudi National Professional Standards for Teachers (NPST). These standards encompass three areas:

An inquiry into teacher assessment literacy
Version [2 - 29112020]

INVITATION TO SCHOOL PRINCIPAL TO ENCOURAGE TEACHERS TO PARTICIPATE IN THE RESEARCH

Professional knowledge, Professional practices (assessment is the third component of this level) and Professional values and responsibilities (NPST).

In light of Saudi's educational reforms under the Saudi 2030 Vision, there is an urgent need to articulate a national assessment literacy framework and support the development of teachers' assessment literacy. What is required of the participants in this study is detailed below for your consideration.

Who will be involved and how will they be involved in the research

All teacher participants (who have volunteered to be part of the study) from public schools will be first involved in responding to 1 online questionnaire. Following this, a sub-sample of the participants will then be involved in 3 focus groups facilitated by Mr Alzhrany. These focus groups are aimed to get participants' views on the results of the questionnaire. The completion of this process will assist the researchers in the initial articulation of an assessment literacy framework.

Once the framework has been developed, a small sample of teachers in the focus groups will then be invited to consider each item from a number of existing questionnaires (tests) and indicate which elaboration (cell in the framework), if any, the item is assessing. This exercise provides evidence of the content validity of the existing items in the proposed assessment literacy framework.

Further items will also be constructed by the researchers to maximise the coverage of the framework. The focus groups will be invited to review the new items to see if they are really assessing the elaboration that they are supposed to be assessing and then come together in the online focus group to provide the feedback from their own review.

The times and places for focus group sessions and the completion of online questionnaires will be negotiated with you to fit these with your schedule.

Important Considerations

With the participants' permission, focus group sessions will be audio-recorded to ensure an accurate record of their comments during these sessions is produced. When the recordings have been transcribed, participants can request a copy of the transcript to be provided, so that they can verify that the information is correct and/or request deletions.

Ethical Considerations

We intend to protect participants' anonymity and the confidentiality of their responses to the fullest possible extent, within the limits of the law. Participants' name and contact details along with any data that you supply will be kept in a separate, password-protected computer file. In the final report, participants will be referred to by a pseudonym. We will remove any references to personal information that might allow someone to guess participants' identity.

Once the research has been completed, a brief summary of the findings will be available to you and the participants. It is likely that the results will be presented at academic conferences and

An inquiry into teacher assessment literacy
Version [2 - 29112020]

Appendix I: Letter to Ministry of Education

LETTER TO MINISTRY OF EDUCATION, SAUDI ARABIA TO REQUEST RESEARCH CONSENT



School of Education and Social Work

ABN 15 211 513 464

Chief Investigator
Professor Jim Tognolini

Room 333
Education Building A.35
The University of Sydney
NSW 2006 AUSTRALIA
Telephone: +61 2 86275205
Email: jim.tognolini@sydney.edu.au
Web: <http://www.sydney.edu.au/>

Dear Sir/Madam,

Re: Consent to conduct the PhD research entitled, An inquiry into teacher assessment literacy: the development of an assessment literacy framework and instrument to measure the assessment literacy of teachers in Saudi Arabia

We are writing to you to seek your consent in conducting the above research. The research will be conducted by Professor Jim Tognolini, as Chief Researcher and DR Jon Twing Researcher and Dr Jon Twing who is not from the University of Sydney, he is a member in Educational Measurement Interest group at the University of Sydney, and he is senior VP of Pearson Assessments and Mr. Alzhrany Abdullah, as Co Researchers, from the School of Education and Social Work at the University of Sydney. This research is part of Mr. Alzhrany's PhD dissertation.

Mr Alzhrany is a Saudi citizen and a teacher from the Ministry of Education who is currently a Saudi government scholarship holder. The broad aim of his research is to develop a national framework for developing and validating a teacher assessment literacy questionnaire which can help measure the assessment literacy of Saudi teachers in public schools.

More specifically the study aims to:

- develop a framework that aligns the assessment knowledge and skills demanded of teachers (as described in the KSA Professional Teaching Standards) with increasing performance levels of assessment literacy;
- evaluate the efficacy of a current assessment literacy questionnaire (Teacher Assessment Literacy Questionnaire (TALQ) for measuring teacher performance in relation to the assessment literacy framework;
- use the framework to design an assessment literacy questionnaire that builds the TALQ and can be used to assess KSA teachers' assessment skills and knowledge in relation to the assessment literacy framework; and
- construct and validate an assessment literacy questionnaire for use with KSA teachers.

An inquiry into teacher assessment literacy
Version [1 - 30092020]

LETTER TO MINISTRY OF EDUCATION, SAUDI ARABIA TO REQUEST RESEARCH CONSENT

The proposed framework will specify what it is Saudi teachers need to know and be able to do to demonstrate proficiency in relation to the assessment standards prescribed in the Saudi National Professional Standards for Teachers (NPST). These standards encompass three areas: Professional knowledge, Professional practices (assessment is the third component of this level) and Professional values and responsibilities.

In light of Saudi's educational reforms under the Saudi 2030 Vision, there is an urgent need to articulate a national assessment literacy framework and support the development of teachers' assessment literacy. What is required of the participants in this study is detailed below for your consideration.

Who will be involved and how will they be involved in the research

All teacher participants (who have volunteered to be part of the study) from public schools will be first involved in responding to 1 online questionnaire. Following this, a sub-sample of the participants will then be involved in 3 focus groups facilitated by Mr Alzhrany. These focus groups are aimed to get participants' views on the results of the questionnaire. The completion of this process will assist the researchers in the initial articulation of an assessment literacy framework.

Once the framework has been developed, a small sample of teachers in the focus groups will then be invited to consider each item from a number of existing questionnaires (tests) and indicate which elaboration (cell in the framework), if any, the item is assessing. This exercise provides evidence of the content validity of the existing items in the proposed assessment literacy framework.

Further items will also be constructed by the researchers to maximise the coverage of the framework. The focus groups will be invited to review the new items to see if they are really assessing the elaboration that they are supposed to be assessing and then come together in the online focus group to provide the feedback from their own review. The times for all focus group sessions and the completion of online questionnaires will be negotiated with participants to fit in with their schedule.

Important Considerations

With the Ministry's and participants' permission, focus group sessions will be audio-recorded to ensure an accurate record of your comments during these sessions is produced. When the recordings have been transcribed, participants can request for a copy of the transcript to be provided, so that they can verify that the information is correct and/or request deletions.

Participants' responses to the online questionnaire will be recorded to assist the researchers in the development of an assessment literacy framework and tool to measure teacher assessment literacy.

Ethical Considerations

We intend to protect participants' anonymity and the confidentiality of their responses to the fullest possible extent, within the limits of the law. Participants' name and contact details along with any data that you supply will be kept in a separate, password-protected computer file. In

LETTER TO MINISTRY OF EDUCATION, SAUDI ARABIA TO REQUEST RESEARCH CONSENT

the final report, participants will be referred to by a pseudonym. We will remove any references to personal information that might allow someone to guess participants' identity.

Once the research has been completed, a brief summary of the findings will be available to the Ministry and participants. It is likely that the results will be presented at academic conferences and seminars, and published in educational journals. The data will be kept securely at the University of Sydney for five years from the date of publication, before being destroyed.

Please be advised that teacher participation in this study is entirely voluntary. Should they wish to withdraw at any stage, or to withdraw any unprocessed data they have supplied, they will be free to do so without prejudice. Participants will be requested to indicate that they have read and understood this information by signing the consent form and returning it to the researchers.

The project has already received ethics approval from the University of Sydney and at this stage, we kindly seek your consent to conduct this important research with teachers from public schools under the Ministry of Education.

We anticipate that the outcomes of this research will be of benefit to Saudi's education and thus look forward to receiving the necessary approval before commencing the research.

Should you require further information, please feel free to contact:

- Mr Alzhrany Abdullah (Co-Researcher), PhD Candidate – School of Education, University of Sydney; telephone: +61 4[REDACTION]; email: aalz2938@uni.sydney.edu.au
- We are deeply grateful for your crucial support and look forward to gaining the relevant consent to conduct this important study.
- please note that DR Jon Twing a member in Educational Measurement Interest group at the University of Sydney. And he is senior VP of Pearson Assessments

Thank you.

Kind regards,

On behalf of the research team

Mr Alzhrany Abdullah, PhD Candidate

School of Education, University of Sydney;

Telephone: +61 4[REDACTION]

Email: aalz2938@uni.sydney.edu.au

Appendix J: Participant Consent



School of Education and Social Work

ABN 15 211 513 464

Chief Investigator
Professor Jim Tognolini

Room 333
Education Building A.35
The University of Sydney
NSW 2006 AUSTRALIA
Telephone: +61 2 86275205
Email: jim.tognolini@sydney.edu.au
Web: <http://www.sydney.edu.au/>

An inquiry into teacher assessment literacy: the development of an assessment literacy framework and instrument to measure the assessment literacy of teachers in Saudi Arabia

PARTICIPANT INFORMATION STATEMENT

(1) What is this study about?

You are invited to take part in a research study about developing and validating a teacher assessment literacy questionnaire which can help measure the assessment literacy of Saudi teachers in public schools and provide a source of evidence for developing professional development programs aimed at enhancing the assessment literacy of teachers.

More specifically the study aims to:

- develop a framework that aligns the assessment knowledge and skills demanded of teachers (as described in the KSA Professional Teaching Standards) with increasing performance levels of assessment literacy;
- evaluate the efficacy of a current assessment literacy questionnaire (Teacher Assessment Literacy Questionnaire (TALQ) for measuring teacher performance in relation to the assessment literacy framework;
- use the framework to design an assessment literacy questionnaire that builds on the TALQ and can be used to assess KSA teachers' assessment skills and knowledge in relation to the assessment literacy framework; and
- construct and validate an assessment literacy questionnaire for use with KSA teachers.

The proposed framework will specify what it is Saudi teachers need to know and be able to do to demonstrate proficiency in relation to the assessment standards prescribed in the Saudi National Professional Standards for Teachers (NPST). These standards encompass three areas: Professional knowledge, Professional practices (assessment is the third component of this level) and Professional values and responsibilities (NPST).

In light of Saudi's educational reforms under the Saudi 2030 Vision, there is an urgent need to articulate a national assessment literacy framework and support the development of teachers' assessment literacy.

You have been invited to participate in this study because you are a teacher who fits the research participant criteria and, therefore, we are of the view that you could contribute to the study in a significant manner with your participation.

This Participant Information Statement tells you about the research study. Knowing what is involved will help you decide if you want to take part in the research. Please read this sheet carefully and ask questions about anything that you don't understand or want to know more about.

Participation in this research study is voluntary.

By giving your consent to take part in this study you are telling us that you:

- ✓ Understand what you have read.
- ✓ Agree to take part in the research study as outlined below.
- ✓ Agree to the use of your personal information as described.

You will be given a copy of this Participant Information Statement to keep.

(2) Who is running the study?

The study is being carried out by the following researchers:

- Professor Jim Tognolini (Principal Researcher), School of Education, University of Sydney
- Dr Jon Twing (Co-Researcher), Who is not from the University of Sydney, he is a member in Educational Measurement Interest group at the University of Sydney, and he is senior VP of Pearson Assessments.
- Mr Alzhrary Abdullah (Co-Researcher), PhD Candidate – School of Education, University of Sydney

(3) What will the study involve for me?

You will be first involved in responding to 1 online questionnaire. This questionnaire you will seek your views on the extent to which you think items related to assessment skills and knowledge are essential in a proposed assessment literacy framework. You will also be invited to participate in 3 focus groups facilitated by Mr Alzhrary. These focus groups are aimed to get your views, as a teacher, on the results of the questionnaire. The completion of this process will assist the researchers in the articulation of an assessment literacy framework.

Once the framework has been developed, you, together with another small sample of teacher participants in the focus groups, will then be invited to consider each item from a number of existing questionnaires (tests) and indicate which elaboration (cell in the framework), if any, the item is assessing. This exercise provides the evidence of the content validity of the existing items in the proposed assessment literacy framework.

Further items will also be constructed by the researchers to maximise the coverage of the framework. As focus group participants, you will be invited to review the new items to see if they are really assessing the elaboration that they are supposed to be assessing and then come together in the online focus group to provide the feedback from their own review.

That's mean it will be three focus groups meetings.

The times and places for focus group sessions and the completion of online questionnaires will be negotiated with you to fit these with your schedule.

With your permission, focus group sessions will be audio-recorded to ensure an accurate record of your comments during these sessions. When the recordings have been transcribed, you can request a copy of the transcript to be provided, so that you can verify that the information is correct and/or request deletions.

Your responses to the online questionnaire will be recorded to assist the researchers in the development of an assessment literacy framework and tool to measure teacher assessment literacy.

(4) How much of my time will the study take?

Your participation in the online questionnaire will take a maximum of 90 minutes of your time.

Your participation in the focus groups will take up to 120 minutes of your time.

(5) Who can take part in the study?

The following are the selection criteria of teacher participants in the study:

- Possess KSA teaching qualifications – this criterion aims to include teacher participants who are familiar with the KSA teacher training given that the research focus on KSA education
- Possess the minimum trained teaching experience of 3 years – this criterion aims to exclude teacher participants who are considered beginning teachers who have yet to acquire sufficient experience and expertise working in the KSA educational system.
- Male or female – this criterion aims to include both genders
- Saudi citizen – this criterion aims to include teacher participants who are locals and are familiar with the KSA cultural norms and the local education system given that the research focus on KSA education

(6) Do I have to be in the study? Can I withdraw from the study once I've started?

Being in this study is completely voluntary and you do not have to take part. Your decision whether to participate or subsequently withdraw from the study will not affect your current or future relationship with the researchers or anyone else at the University of Sydney or the school you are working in and the KSA ministry of education.

If you decide to take part in the study and then change your mind later, you are free to withdraw at any time. You can do this by informing any of the researchers via email or telephone using the contact information provided. Should your withdrawal occur after data has been collected (for example, online questionnaire or focus groups), then your data will be excluded from the study. Your data will also remain private confidential until it is permanently deleted after 5 years upon the completion of the study.

Submitting your completed questionnaire is an indication of your consent to participate in the study. You can withdraw your responses any time before you have submitted the questionnaire. Once you have submitted it, your responses cannot be withdrawn because they are anonymous and therefore we will not be able to tell which one is yours.

If you take part in a focus group, you are free to stop participating at any stage or to refuse to answer any of the questions. However, it will not be possible to withdraw your individual comments from our records once the group has started, as it is a group discussion.

If you decide to withdraw from the study, we will not collect any more information from you. Please let us know at the time when you withdraw what you would like us to do with the information we have

collected about you up to that point. If you wish your information will be removed from our study records and will not be included in the study results, up to the point that we have analysed and published the results.

(7) Are there any risks or costs associated with being in the study?

Aside from giving up your time, we do not expect that there will be any risks or costs associated with taking part in this study.

(8) Are there any benefits associated with being in the study?

We cannot guarantee that you will receive any direct benefits from being in the study.

(9) What will happen to information about me that is collected during the study?

By providing your consent, you are agreeing to us collecting personal information about you for the purposes of this research study. Your information will only be used for the purposes outlined in this Participant Information Statement unless you consent otherwise. Your information will be stored securely and your identity/information will be kept strictly confidential, except as required by law. Study findings may be published, but you will not be individually identifiable in these publications.

(10) Can I tell other people about the study?

Yes, you are welcome to tell other people about the study.

(11) What if I would like further information about the study?

When you have read this information, Mr Alzhrany Abdullah] will be available to discuss it with you further and answer any questions you may have. If you would like to know more at any stage during the study, please feel free to contact Mr Alzhrany Abdullah (Co-Researcher), PhD Candidate – School of Education, University of Sydney; telephone: +61 [REDACTION] email: aalz2938@uni.sydney.edu.au

(12) Will I be told the results of the study?

You have a right to receive feedback about the overall results of this study. You can tell us that you wish to receive feedback by ticking the relevant box in the consent form. This feedback will be in the form of a one-page lay summary of study findings. You will receive this feedback after the study is finished.

(13) What if I have a complaint or any concerns about the study?

Research involving humans in Australia is reviewed by an independent group of people called the Human Research Ethics Committee (HREC). The ethical aspects of this study have been approved by the HREC of the University of Sydney. As part of this process, we have agreed to carry out the study

according to the *National Statement on Ethical Conduct in Human Research (2007)*. This statement has been developed to protect people who agree to take part in research studies.

If you are concerned about the way this study is being conducted or you wish to make a complaint to someone independent from the study, please contact the university using the details outlined below. Please quote the study title and protocol number.

The Manager, Ethics Administration, University of Sydney:

- **Telephone:** +61 2 8627 8176
- **Email:** human.ethics@sydney.edu.au
- **Fax:** +61 2 8627 8177 (Facsimile)

Appendix K: Teacher Participant Plain Language Statement

TEACHER PARTICIPANT PLAIN LANGUAGE STATEMENT



Project Title: An inquiry into teacher assessment literacy: the development of an assessment literacy framework and instrument to measure the assessment literacy of teachers in Saudi Arabia

Principal Researcher: Professor Jim Tognolini

Co-Researchers: Dr Jon Twing & Mr Alzhrany Abdullah

Dear

You are invited to participate in the above research, which is being conducted by Professor Jim Tognolini and Dr Jon Twing and Mr Alzhrany Abdullah from the School of Education and Social Work at the University of Sydney. The project has ethics approval from the University of Sydney and the approval to collect data from schools from the Ministry of Education, Saudi Arabia. This research is part of Mr. Alzhrany's PhD dissertation.

Mr Alzhrany is a Saudi citizen and a teacher from the Ministry of Education who is currently a Saudi government scholarship holder. The broad aim of his research is to develop a national framework for developing and validating a teacher assessment literacy questionnaire which can help measure the assessment literacy of Saudi teachers in public schools and provide a source of evidence for developing professional development programs aimed at enhancing the assessment literacy of teachers.

More specifically the study aims to:

- develop a framework that aligns the assessment knowledge and skills demanded of teachers (as described in the KSA Professional Teaching Standards) with increasing performance levels of assessment literacy;
- evaluate the efficacy of a current assessment literacy questionnaire (Teacher Assessment Literacy Questionnaire (TALQ) for measuring teacher performance in relation to the assessment literacy framework;
- use the framework to design an assessment literacy questionnaire that builds on the TALQ and can be used to assess KSA teachers' assessment skills and knowledge in relation to the assessment literacy framework; and
- construct and validate an assessment literacy questionnaire for use with KSA teachers.

The proposed framework will specify what it is Saudi teachers need to know and be able to do to demonstrate proficiency in relation to the assessment standards prescribed in the Saudi National Professional Standards for Teachers (NPST). These standards encompass three areas: Professional knowledge, Professional practices (assessment is the third component of this level) and Professional values and responsibilities (NPST).

In light of Saudi's educational reforms under the Saudi 2030 Vision, there is an

An inquiry into teacher assessment literacy

Version [1 - 29112020]

urgent need to articulate a national assessment literacy framework and support the development of teachers' assessment literacy. What is required of the participants in this study is detailed below for your consideration.

Participation as a teacher

You will be involved in three focus group sessions and respond to an online questionnaire. The purpose of the focus group sessions is to gather participants' views on the items that are considered essential in the articulation of a national assessment literacy framework and the design of a teacher assessment literacy questionnaire.

The purpose of responding to the online questionnaire is to understand the efficacy of the questionnaire as a tool to measure teachers' assessment literacy that is based on the articulated assessment literacy framework. The times for focus group sessions and the completion of online questionnaires will be negotiated with you to fit in with your schedule.

Important Considerations

With your permission, focus group sessions will be audio-recorded to ensure an accurate record of your comments during these sessions. When the recordings have been transcribed, you can request a copy of the transcript to be provided, so that you can verify that the information is correct and/or request deletions.

Your responses to the online questionnaire will be recorded to assist the researchers in the development of an assessment literacy framework and tool to measure teacher assessment literacy.

Ethical Considerations

We fully intend to protect your anonymity and the confidentiality of your responses to the best of our ability, within the limits of the law. Your name and contact details along with any data that you supply will be kept in a separate, password-protected computer file. In the final report, you will be referred to by a pseudonym. We will remove any references to personal information that might allow someone to guess your identity.

Once the research has been completed, a summary of the findings will be available to you. The results will likely be presented at academic conferences, seminars and be published in educational journals. The data will be kept securely at the University of Sydney for five years from the date of publication, before being destroyed.

Please be advised that your participation in this study is entirely voluntary. Should you wish to withdraw at any stage, or to withdraw any unprocessed data you have supplied, you are free to do so without prejudice. If you would like to participate, please indicate that you have read and understood this information by signing the accompanying consent form and returning it in the envelope provided.

An inquiry into teacher assessment literacy

Version [1 - 29112020]

For more information

For more information about this project participants should feel free to contact:

- Professor Jim Tognolini (Principal Researcher), School of Education, University of Sydney; telephone +61 2 86275205
- email: jim.tognolini@sydney.edu.au
- Dr Jon Twing (Co Researcher), who is not from the University of Sydney, he is a member in Educational Measurement Interest group at the University of Sydney, and he is senior VP of Pearson Assessments.
- telephone: email: jon.s.twing@pearson.com
- Mr Alzhrary Abdullah (Co-Researcher), PhD Candidate – School of Education, University of Sydney; telephone: +61 4[REDACTION]
- email: aalz2938@uni.sydney.edu.au

Should you have any concerns about the conduct of the project, you are welcome to contact the Executive Officer, Human Research Ethics, University of Sydney, on telephone: +61 2 86278176, or fax: +61 2 8627 8177 (Facsimile)

Email: human.ethics@sydney.edu.au

Thank you very much for reading this information. We are deeply grateful for your crucial support and look forward to your participation in this important study.

Appendix L: The Ministry of Education Approval

الرقم : ٣/١ / ١٤٥

التاريخ : ٢١ / ٥ / ١٤٤٢ هـ

المرفقات :



وزارة التعليم
Ministry of Education

المملكة العربية السعودية

وزارة التعليم
(٢٨٠)

الإدارة العامة للتعليم بمحافظة جدة
إدارة التخطيط والمعلومات - البحوث والدراسات

رؤيتنا : متعلم .. معتر بدينه .. منتم لوطنه .. منتج للمعرفة .. منافس عالمياً .

إفادة بتطبيق البحث

الاسم	عبدالله احمد علي الزهراني		
الاسم	السجل المدني	١٠٥٢٦٢٠٤٨٩	
الجوال	البريد الإلكتروني	winner11221122@gmail.com	٠٥٥٦٦١٦٢١٧
الجهة المشرفة على البحث	جامعة سبني باستراليا	التخصص	القياس والتقويم التربوي
الدرجة العلمية	دكتوراه	عينة البحث	معلمي المدارس الابتدائية والمتوسطة والثانوية
عنوان البحث	دراسة لبناء إطار ومقياس لقياس مهارات المعلمين ومعلوماتهم في القياس والتقويم التربوي		
الموضوع بشأن	تسهيل مهمة الباحث بتطبيق بحثه		

إلى : من يهمة الامر .

السلام عليكم ورحمة الله وبركاته

بناء على إفادة الملحق الثقافي في سفارة المملكة العربية السعودية في كانبيرا بأن الباحث (الموضح بياناته أعلاه) أحد الطلبة الدارسين خارج المملكة، واستجابة لرغبته في تزويدكم بالموافقة بتطبيق أدوات بحثه و جمع البيانات المتعلقة بعينة الدراسة .

نفيدكم أنه لا مانع لدينا بالموافقة بعد دراسة أداة البحث ؛ وذلك تشجيعاً للبحث العلمي وبما يعود على الوطن بمثل هذه البحوث الميدانية ؛ شاكرين ومقدرين تعاونكم واهتمامكم.

والسلام عليكم ورحمة الله وبركاته

مدير إدارة التخطيط والمعلومات بتعليم جدة

[REDACTION]

خليل بن فراج الوافي

هاتف ٦٤٤٤٣٠٥ - فاكس ٦٤٣٤٠٤٠ - الرمز البريدي : ٢١١٥٨

Appendix M: The Safety Protocol



ABN 15 211 513 464

School of Education and Social Work

Chief Investigator
Professor Jim Tognolini

Room 333
Education Building A.35
The University of Sydney
NSW 2006 AUSTRALIA
Telephone: +61 2 86275205
Email: jim.tognolini@sydney.edu.au
Web: <http://www.sydney.edu.au/>

An inquiry into teacher assessment literacy: the development of an assessment literacy framework and instrument to measure the assessment literacy of teachers in Saudi Arabia

Safety Protocol

- The researcher will be gathering data such as focus group interviews and online surveys in his home country, Saudi Arabia.
- He will namely conduct focus group interviews alone and will not require a local interpreter as he is a Saudi national who speaks the native language. His supervisor considers the safeguards provided in this safety protocol are sufficient to manage any safety risks.
- The researcher will discuss interview safety and perform practice interviews with his supervisor prior to travelling overseas.
- The researcher will take any further advice from the following local institution regarding the safety of the planned research process: Ministry of Education, Saudi Arabia.
- Being a local and Saudi national, the researcher is very accustomed to the country and the areas in which the research will be conducted. As such, any risk in this regard is mitigated.
- The researcher will provide a schedule to his supervisor of the date, time and place of all interviews.
- He will maintain contact with his supervisor at least twice a week by email to update on the progress and issues that may arise.
- The researcher will conduct the interviews in schools of the teacher participants. These are considered safe, public places and being a local, the researcher is familiar with the country, the areas and places.
- The researcher has confirmed there are currently no travel warnings from the Department of Foreign Affairs and Trade for Saudi Arabia. He has subscribed to the travel advice to receive email updates each time the travel warning is reissued.
- The researcher will obtain the relevant local ethics and research approval as required before commencing research in his home country.
- The researcher undertakes to follow The University of Sydney guidelines on Fieldwork Safety Standards

This safety protocol has been agreed and accepted by the researcher and the supervisor.

Student's signature

[REDACTION]

Abdullah Alzhrany

Supervisor's signature

[REDACTION]

Professor Jim Tognolini

An inquiry into teacher assessment literacy

Version [2 -29112020]

Page 1 of 1