Bhutanese Elementary Teachers' Assessment Beliefs, Assessment Practices and Assessment Literacy

A thesis submitted in partial fulfilment of the requirements for the

degree of **Master of Education**

Hemlata Karki

(a1737492)

School of Education,

Faculty of Arts

The University of Adelaide

November 2020

DECLARATION

I certify that this dissertation has no material which has been accepted for an award of diploma or degree against my name, in any tertiary institution or other university. Further, to the best of my knowledge and belief, no previously published or written material by another person has been included, except where due acknowledgement has been made . I also certify that without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint award of this degree, no part of this work will be submitted in my name for diploma or degree in other tertiary institution and university.

I give consent to this copy of my dissertation, when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.
I also give permission for the digital version of my dissertation to be made available on the web, via the University's digital research repository, the Library Search and also through web search engine, unless permission has been granted by the University to restrict access for a period of time.

Signature :

Date: 6 November 2020

Acknowledgement

The accomplishment of writing this dissertation would have not been possible without the support, guidance, nurture and encouragement from my principal supervisor **Dr Lynda Maree MacLeod.** I would like to extend my deepest gratitude and gratefulness to her for making the journey of writing this dissertation a hopeful and engaging one. My deepest appreciation is for her patience to do a detail correction and provide feedback in the draft which significantly improved my report.

Equally, I am deeply indebted to Dr I Gusti Ngurah Darmawan, a senior lecturer and a professor to teach quantitative approaches, for happily guiding me through handling and analysing the data. Without his guidance, I would not have successfully analysed the results for this dissertation. I would like to extend my special appreciation for his unconditional support. Additionally, I would like to extend my thankfulness to a Ph D student Khusaini who mentored me and provided feedback in improving result sections with appropriate tables and figures.

Likewise, I would like to express my gratefulness to Australia Awards Scholarship program for their generous grants to edit my writing through an external source. In connection to this, I would like to thank Niranjala Seimon, Student Contact Officer for approving the grant and editing assistance proposal. I am also grateful to Dr Owen Lindsay for editing my writing.

I would also like to extend my sincere gratitude to the principals, local contact persons and teacher participants of sample schools as well as members in the Facebook closed group for participating, creating network and encouraging more teachers to participate in this study. I must not forget to extend my deepest appreciation to Dr Kezang Sherab, a lecturer in Paro College of Education for initiating to submit my research proposal application to the Ministry of Education. I also would like to give my sincere thanks to the Ministry of Education (Bhutan) for approving my proposal.

Words are not enough to acknowledge my husband Robin Chhetri and sister Kumari Karki for their genuine support, love and encouragement throughout this journey. I also remember my loving children, beloved parents and two caring sisters for showering their blessings and love and having confidence on my endeavour.

Table of Contents

DECLARATION I

ACKNOWLEDGEMENT II

LIST OF TABLES VII

LIST OF FIGURES IX

LIST OF ABBREVIATION X

ABSTRACT XII

CHAPTER ONE: INTRODUCTION 1

1.1 INTRODUCTION 1

1.2 BACKGROUND AND RATIONALE 1

1.3. ASSESSMENT SYSTEM IN BHUTAN 3

1.4. STATEMENT OF THE PROBLEM 8

1.5. AIMS AND OBJECTIVES 10

1.6.RESEARCH QUESTIONS 11

1.7. SIGNIFICANCE OF THE STUDY 11

1.8. OUTLINE OF SUBSEQUENT CHAPTERS 13

1.9. DEFINITION OF KEY TERMS 13

CHAPTER TWO: LITERATURE REVIEW 14

2.1 INTRODUCTION 14

2.2 SUMMATIVE ASSESSMENT 14

2.3 FORMATIVE ASSESSMENT 15

2.4 INTEGRATING FORMATIVE ASSESSMENT AND SUMMATIVE ASSESSMENT 16

2.5 TEACHERS' BELIEFS/CONCEPTIONS ABOUT ASSESSMENT 16

2.6 ASSESSMENT PRACTICES 23

2.7 ASSESSMENT LITERACY 26

CONCLUSION 30

CHAPTER THREE: METHODOLOGY 31

3. 1. INTRODUCTION 31

3.2 CONCEPTUAL FRAMEWORK 31

3.3 HYPOTHESIS 34

3.4. RESEARCH QUESTIONS 34

3.5. RESEARCH APPROACH: QUANTITATIVE APPROACH 35

3.6. RESEARCH DESIGN: THE SURVEY DESIGN 37

3.7. POPULATION AND SAMPLE 38

3. 8. INSTRUMENTATION 39

3.8.1.Section One: Demographic information 39

3.8.2. Section Two: Conceptions of Assessment III (COA-III) 40

3.8.3. Section Three: Self-reported assessment practice 40

3.8.4. Section Four: Classroom Assessment Literacy Inventory (CALI) 41

3.9. VARIABLES IN THE STUDY 44

3.10. ETHICS 46

3.11. DATA COLLECTION PROCESS 46

3.12. DATA ANALYSIS AND INTERPRETATION 47

3.12.1 Overview 47

3.12.2 Data preparation 47

3.12.3 Data analysis techniques 47

CHAPTER FOUR: RESULTS 50

4.1. INTRODUCTION 50

4.2. DESCRIPTIVE RESULT OF DEMOGRAPHIC DATA 50

4.2.1. Gender 50

4.2.2. Highest Teacher Qualification 51

4.2.3. Number of years in teaching/teaching experience 52

4.2.4. GRADE TAUGHT 52

4.2.5. SUBJECT TAUGHT 53

4.2.6. ASSESSMENT EDUCATION/TRAINING 54

4.3 CONSTRUCT VALIDITY OF CONCEPTION OF ASSESSMENT (COA-III A) 54

4.3.2 EVALUATING THE FACTOR LOADINGS AND SQUARED CORRELATION 56

4.3.3 ASSESSMENT BELIEFS MODEL FIT 57

4.3.4 CORRELATION AMONG THE FOUR CONSTRUCTS OF ASSESSMENT BELIEFS 58

4.4. CRONBACH ALPHA; TESTING THE RELIABILITY 59

4.5 QUESTION 1. 60

4.6 ASSESSMENT PRACTICES 62

4.6.1 CONSTRUCT VALIDITY OF ASSESSMENT PRACTICES ITEMS 62

4.6.2 CONFIRMATORY FACTOR ANALYSIS 63

4.6.3. EVALUATING THE FACTOR LOADINGS AND SQUARED CORRELATION (R2) 63

4.6.4 ASSESSMENT PRACTICE MODEL FIT 65

4.6.5 CORRELATION AMONG THE THREE LATENT VARIABLES 65

4.7 CRONBACH ALPHA; TESTING THE RELIABILITY 66

4.8. QUESTION 2 66

4.9 QUESTION 3. 69

4.10. QUESTION FOUR. 71

4.10.1 DEMOGRAPHIC INFORMATION AND ASSESSMENT BELIEFS 72

4.10.2. DEMOGRAPHIC INFORMATION AND ASSESSMENT PRACTISES 74

4.10.3. DEMOGRAPHIC INFORMATION AND ASSESSMENT LITERACY LEVEL 76

4.11. QUESTION 5 77

4.11.1 STRUCTURAL EQUATION MODEL (SEM) 78

4.11.2. DIRECT, INDIRECT, AND TOTAL EFFECTS ON ASSESSMENT PRACTISES 79

4.11.3. THE GOODNESS OF MODEL FIT INDICES 82

CHAPTER FIVE: DISCUSSION, RECOMMENDATIONS, AND CONCLUSION 83

5.1 INTRODUCTION 83

5.2. ASSESSMENT BELIEFS 83

5.3. ASSESSMENT PRACTICES 86

5.4. ASSESSMENT LITERACY 89

5.5 SIGNIFICANT DIFFERENCES OF INDEPENDENT VARIABLES ON DEPENDENT VARIABLES 90

5.6. EFFECTS ON ASSESSMENT PRACTICES 93

5.7. RECOMMENDATIONS 95

5.8. SCOPES AND LIMITATION AND FUTURE RESEARCH 98

5.9. CONCLUSION 100

REFERENCES 101

APPENDICES 118

APENDIX 1 ETHICS APPROVAL 118

APPENDIX 2 APPROVAL LETTER FROM THE MINISTRY OF EDUCATION 120

APPENDIX 3 PARTICIPATION INFORMATION SHEET 121

APPENDIX 4 ONLINE SURVEY QUESTIONNAIRE 125

APENNDIX 5: APPLICATION TO THE PRINCIPALS 136

APPENDIX 6: INTRODUCTORY MESSAGE AND CONSENT TO TEACHERS 137

APPENDIX 7 ACKNOWLEDGEMENT AND THANKING EMAIL TO THE PRINCIPALS 138

APPENDIX 8 SAMPLE SCHOOLS 139

APPENDIX 9 CFA READING 140

List of Tables

Table 1. Summary of assessment practice and weighting in % across all grades 6

Table 2 Current assessment practice in elementary classes, (Royal EducationCouncil, 2019) . 7

Table 3 Continuous formative assessment techniques and tools matrix, (REC, 2019). 8

Table 4 Assessment beliefs continuum (formative assessment to summative assessment) .. 22

Table 5 Summary of the survey questionnaires 43

Table 6. Summary of variables in the study 45

Table 7 Summary of data analysis techniques 49

Table 8 Participants by gender in percent 51

Table 9 Factor Loadings and Squared Coefficient 57

Table 10 Goodness-of- model fit indices 58

Table 11 Correlation between four constructs of assessment 59

Table 12 Cronbach Alpha of each construct 60

Table 13 Descriptive statistics for assessment beliefs 61

Table 14 Factor loadings and squared correlation (R^2) 64

Table 15 Goodness-of-model fit indices 65

Table 16 Correlation among three constructs of assessment practices 66

Table 17 Cronbach Alpha for assessment practices 66

Table 18 Descriptive Statistics for Frequency , Mean and Percent for value of Assessment Practice 68

Table 19 Descriptive Statistics for three constructs of assessment practice 69

Table 20 Scores for classroom assessment literacy 70

Table 21 Levels of classroom assessment literacy 71

Table 22 One -way analysis of variance (ANOVA) results of significant difference on assessment beliefs (school accountability) by teaching experience 72

Table 23 Post-Hoc Tests (Bonferroni) results of significant difference on assessment belief (school accountability) by teaching experience 73

Table 24 One -way analysis of variance (ANOVA) results of significant difference on assessment beliefs (student accountability) by teacher education level. 73

Table 25 Post-Hoc Tests (Bonferroni) results of significant difference on assessment belief (student accountability) by teacher education level. 73

Table 26 One -way analysis of variance (ANOVA) results of significant difference on assessment beliefs (assessment is irrelevant) by subject taught 74

Table 27 Post-Hoc Tests (Bonferroni) results of significant difference on assessment belief (assessment is irrelevant) by subject taught 74

Table 28 Independent sample t-test showing significant difference on assessment practises (Formative assessment) by gender 75

Table 29 One -way analysis of variance (ANOVA) results of significant difference on assessment practices (assessment design) by level of teacher education 76

Table 30 Post-Hoc Tests (Bonferroni) results of significant difference on assessment practices (assessment design) by teacher education 76

Table 31 One -way analysis of variance (ANOVA) results of significant difference on assessment literacy by level of teacher education 77

Table 32 Post-Hoc Tests (Bonferroni) results of significant difference on assessment literacy by teacher education 77

Table 33 Direct, Indirect and Total effects of independent and exogenous variables on dependent, endogenous variable (Assessment Practices) 81

Table 34 Summary of Goodness-of-fit Indexes for SEM model. 82

LIST OF FIGURES

Figure 1 Conceptual framework 33

Figure 2 A Hypothesised Model 34

Figure 3 Quantitative Methodology Framework with Survey Research Design 36

Figure 4 The levels of teacher education of elementary teachers 51

Figure 6. Pie chart comparing participants by the grade taught 53

Figure 7 Pie chart comparing teachers by subject they taught 53

Figure 8 Intensity of assessment education/training received by teachers 54

Figure 10 Correlated Four Factor Model (Unstandardised) 55

Figure 11 Error Plot showing mean scores for assessment beliefs 61

Figure 12 Correlated Three Factor Model (Standardised) 62

Figure 13 Correlated Three Factor Model (Unstandardised) 63

Figure 14 Error Plot comparing mean scores for assessment practice 69

Figure 15 . SEM Path Model (Standardised) 78

Figure 16. SEM Path Model (Unstandardised) 79

LIST OF ABBREVIATION

ANOVA : Analysis of Variance

AFT : American Federation of Teachers

B Ed : Bachelor of Education

BCSEA: Bhutan Council for School Examination and Assessment

CALI: Classroom Assessment Literacy Inventory

CAPSD: Curriculum and Professional Support Division

CBAT: Competency Based Assessment Test

CERI: Centre for Educational Research and Innovation

CFA : Confirmatory Factor Analysis

CFA: Continuous Formative Assessment

CFI : Comparative Fit Index

CoA-III : Conceptions of Assessment

COA-III A : Conception of Assessment Abridged version

CSA: Continuous Summative Assessment

MOE: Ministry of Education

NAPE: New Approach to Primary Education

NBIP: Nationally Based In-Service Programs

NCME : National Council on Measurement in Education

NEA : National Education Association

NEA: National Education Assessment

OECD: Organisation for Economic Co-operation and Development

PGDE : Post Graduate Diploma in Education

PISA-D : Programme for International Student Assessment for Development

PP: Pre-Primary

PTC : Primary Teacher Certificate

REC: Royal Education Council

RGOB: Royal Government of Bhutan

RMSEA : Root Mean Square Error of Approximation

SA: Summative Assessment

SEM : Structural Equation Modeling

SPSS: Statistical Package for Social Science

TALQ: Teacher Assessment Literacy Questionnaire

TCoA: Teacher Conception of Assessment

TLI: Tucker-Lewis Index

Abstract

Teachers' assessment conceptions/beliefs, assessment practices, and assessment literacy level are important as they inform classroom assessment decision making of teachers. The purpose of this study was to investigate Bhutanese elementary teachers' conceptions/beliefs about assessment, assessment practices, and level of assessment literacy to get an insight into teachers' implementation of formative assessment. The assessment conceptions/beliefs were measured using the abridged version of the CoA-III inventory proposed by Brown (2006). The CoA includes four pre-determined assessment conceptions: assessment for improvement, assessment for student accountability, assessment for school accountability, and assessment as an irrelevance. The assessment practices were measured under the themes of formative assessment, summative assessment, and assessment design adapted from the instrument designed by McMillian et al (2002). The level of assessment literacy was measured using the Classroom Assessment Literacy Inventory (Mertler, 2003). The study collected data from 112 elementary teachers, teaching any subject from Pre-Primary to Garde 6, using an online, web-based survey in Survey Monkey. The findings showed that teachers agreed more with the improvement purpose of assessment, however, they did not disagree with the irrelevance conception. There was a strong correlation between improvement and student accountability conception (r=.71) and a positive association between improvement and irrelevance conception (r=.14). The teachers valued all assessment practices almost equally, formative assessment (M)=3.94, summative assessment (M)=3.71, assessment design (M)=3.75. The average score in assessment literacy was 8.82 on 21 points (42%). 62.2% of teachers scored at a low level, 30.8% scored at a medium level and 6.95% scored at a high level. It was concluded that teachers have positive beliefs about assessment, but they often experience an assessment practice dilemma between improvement and accountability purposes of assessment. These findings will be useful for policymakers, school leaders, and teacher educators in formulating policies, conducting professional development programs, and improving teacher education programs to support effective implementation of formative assessment.

Keywords: Assessment conceptions/beliefs, assessment practices, assessment literacy, improvement, accountability, irrelevance, formative and summative assessment.

CHAPTER ONE: INTRODUCTION

1.1 INTRODUCTION

This study investigates Bhutanese elementary teachers' assessment beliefs, practices, and literacy levels, to gain better understandings of formative assessment practices so that school leaders and policymakers can support the successful implementation of recent assessment reforms introduced in elementary classes. This chapter briefly presents the significance of formative assessment and the challenges of effective implementation through the lens of teachers' assessment beliefs, practices, and literacy. It further gives an overview of the assessment system in Bhutan. The chapter also provides problem statements, research questions, and the significance of this study. Finally, this chapter concludes by presenting the structure of the thesis and definitions of key terms.

1.2 BACKGROUND AND RATIONALE

Numerous studies indicate that the use of formative assessment by teachers in motivating and providing feedback enhances student achievement (Black & Wiliam, 1998; OECD, 2006). Formative assessment is a concept that encompasses different methods of using assessment to support student learning (Van der Kleij, Vermeulen, Schildkamp, & Eggen, 2015; Briggs, Ruiz-Primo, Furtak, Shepard, & Yin, 2012). The introduction of formative assessment by teachers into their classroom practices can bring considerable learning benefits such as increased academic performance, self-regulated learning, and self-efficacy (Black & Wiliam, 1998; Sadler 1989; Hattie, 2009; Kingston & Nash 2011; Panadero & Jonsson, 2013). However, the literature reveals the difficulties of implementing formative assessment or assessment for learning (Black & Wiliam, 2009; Brown, Lake, & Matters, 2011).

One way of gauging the difficulties of implementing formative assessment can be to consider the assessment beliefs and practices of teachers. The theories and research into these beliefs show that teachers' beliefs about assessment influence their assessment practices (Brown, 2008; Barnes, Fives & Dacey, 2015; Fives & Buehl, 2012). According to Brown and Gao (2015, p.3), the interpretation, implementation, and response to a system of assessment are largely determined by teachers' beliefs about assessment. This implies that for any assessment practice to be successful, policymakers and education leaders must understand what teachers

believe about assessment. Brown and Gao (2015), argue that gaining a better understanding of teachers' beliefs allows policymakers and leaders to support the implementation of new curricula and assessment systems. Remesal (2007) argues that to understand an assessment practice, it is crucial to understand the assessment beliefs; especially those of teachers, as when reforms are made, teachers are the last step in the change process.

Another attribute that may contribute to the successful implementation of formative assessment is the knowledge and expertise possessed by teachers. According to Churchil et al (2011), teachers must possess a sound knowledge of the effective application of assessment, among other educational concepts. However, the empirical evidence on teachers' assessment literacy suggests a low level of literacy in assessment practices (Mertler, 2003; 2009;) an inadequate preparation of pre-service teachers in assessment (Plake, Impara & Fager, 1993; Perry, 2013); and a difficulty among teachers in defining formative assessment, despite having consensus on the construct of formative assessment (Balck & Wiliam, 1998; Sadler, 1998; Stiggins, 1999). Meanwhile, few studies have been conducted into the assessment literacy of teachers (Leighton, Gokiert, Cor, & Heffernan, 2010). This is particularly true in Bhutan, where no studies have yet been conducted on teachers' assessment beliefs, practices, and literacy, despite Bhutan placing paramount importance on student learning assessment, and introducing several assessment reforms, and yet unable to achieve its desired educational outcomes.

It has been evident in the researcher's experience as a practising teacher and through anecdotal observations, that Bhutanese teachers seem to emphasise summative assessment practices. The Royal Education Council (REC) (2019) found that, despite the mandate to implement continuous formative assessment particularly in primary education, in practice, "teaching to the test" is predominant as opposed to supporting learners to acquire essential competencies. According to Wiliam, Lee, Harrison, & Black (2004), the conventional practice of high-stakes mandatory state tests that exists in most countries makes the successful implementation of formative assessment difficult. Conducting public and other major examinations in Bhutan, therefore, seems to be detrimental to the effective practice of formative assessment. Meanwhile, teachers appear less competent in practising formative assessment. The Bhutan Council for School Examination and Assessment- BCSEA (2015) found

that, even after more than a decade since the introduction of formative assessment, teachers showed only a little or moderate understanding of formative assessment. Teachers further reported facing challenges in implementing formative assessment due to having insufficient or no knowledge about formative assessment (BCSEA, 2015). The introduction of formative assessment did not alter traditional assessment practices used by teachers, and teachers viewed formative assessment as negatively impacting learning (Utha, 2014). Thus, fully adopting continuous formative assessment in primary education could be problematic without examining the assessment beliefs and knowledge of teachers, to align professional support and development for teachers.

Given these issues, there have been calls to improve the assessment system at the school level (MOE, 2014). This call recognises the critical roles of teachers and raises concerns regarding their assessment beliefs and assessment literacy, which have direct bearings on assessment practices and educational outcomes. Therefore, this study seeks to investigate Bhutanese elementary teachers' assessment beliefs, practices, and literacy in supporting the effective practice of formative assessment and improving the overall quality of education in Bhutan.

1.3. ASSESSMENT SYSTEM IN BHUTAN

Bhutan places paramount importance on student learning assessment and recognises assessment as an integral part of the teaching and learning process. Such considerations are evident in several policy documents. For instance, the *Bhutan Education Blueprint 2014-2024*, a working document for the government to follow (Thinley, 2016), clearly identifies and emphasises "revamping of the assessment system". The Blueprint recommends reviewing and strengthening formative assessment. Further, it states that "(a)ssessment practices are based on discovering the talents and potentials of each child and facilitate effective feedback to promote learning. All students are highly motivated to learn." (MOE, 2014, pp. 77-78). However, the document does not provide specific strategies on how the revamping of the assessment system will take place.

Similarly, the national education policy (draft, August 2019) states that "(a)ssessment of student learning shall be based on learning standards or outcomes prescribed in the school

curriculum and shall include standardised holistic school level assessment and national examinations. It shall incorporate formative and summative forms and emphasise competencies." (RGOB, 2019, p.10). Meanwhile, recognising the challenges of ensuring the quality and depth of students' learning in the face of their diverse needs and those of society, the Curriculum and Technical Advisory Board, recommends strengthening the continuous formative assessment process, and abolishing written examinations beginning from lower primary education (i.e from Pre-Primary – Garde 3) from 2020. These policies and reforms represent the aspirations of the Bhutanese education assessment system and consequently, various assessment practices have been adopted to realise them. Nonetheless, the effectiveness of these practices in the Bhutanese context remains unclear.

The Bhutanese education system is assessed at two levels, the school level, and the national level, following national and international standards (BCSEA, 2019). The BCSEA is responsible for conducting national level examinations and assessments (BCSEA, 2015). The BCSEA carries out examinations and assessments in four key stages of student learning at Grades 3, 6, 10, and 12. A Competency-Based Assessment Test (CBAT) is conducted at the end of Grades 3 and 6, for which question papers, model answers, marking schemes, and examination standards are provided to schools by BCSEA, and test administration and evaluations are performed at the school level. The BCSEA also conducts high stake examinations at Grade 10 and 12 on its own. Besides these examinations BCSEA also periodically conducts a National Education Assessment (NEA) in Bhutan, using standardised test instruments (BCSEA, 2019; Maxwell, Rinchen, & Cooksey, 2010). So far, the NEA has conducted two rounds of examinations from 2004 to 2013 for Garde 6 and 10 (BCSEA, 2019). However, the results of the NEA presented several concerns about the quality of education in Bhutan. Therefore, BCSEA (2019) developed a robust National Education Assessment Framework (draft) to improve the conduct of the NEA based on experience, mitigate challenges, and also present underlying theories related to NEA design (p. 15). The effectiveness of this framework is, however, yet to be determined.

Assessment at the school level has evolved through several reforms. Before the introduction of the New Approach to Primary Education (NAPE) in 1986, schools conducted entirely summative examinations. However, with the introduction of NAPE, the assessment system was reformed (Max et al, 2010; Utha, 2014) to eventually include both summative and

formative assessments, at least in primary education. The concept of 'ongoing evaluation'- meaning regular evaluation through observation of students' behavioural, social, and academic skills was emphasised. School-level assessment involves term examinations (summative assessments) and continuous assessment (formative assessments) which are conducted by the schools. Ongoing evaluation or continuous assessment involves classwork, homework, and project work constituting 50% of the total 100 marks in elementary classes (CAPSD, 1999; 2008). Scores from these examinations and assessments are used to determine student learning achievements and provide opportunities for interventions to improve the system. However, it remains little explored if the evidence is being used for this purpose.

Formative assessments have also been introduced in secondary classes, though with less weighting than in primary education. Table 1 shows the assessment practices and weighting across all grades from PP to 12. From Grade 7 to 10, formative assessment accounts for 20% and is generally obtained by assessing homework, classwork, and project work. For Grade 11 and 12 there is no formative assessment component. From 1986, the on-going evaluation was implemented for all subjects in elementary classes (PP to 5) with a weighting of 30% in the final evaluation for promotion to the next grade. Thus, the introduction of ongoing evaluation became the practice of formative assessment. Mid-year examinations (half-yearly examinations) and end-year examinations (annual examinations) comprised of both oral and written tests, with a weighting of 30% (mid-year) and 40% (end-year), with the remainder of students' grades comprised of ongoing evaluation. Beginning in 1999, the same practice was extended to Grade 6. By this time the term "ongoing evaluation" was referred to as "continuous assessment". Although "ongoing evaluation" or "continuous assessments" are meant to be formative and come in the form of descriptive feedback, in practice, they have taken the form of grades or scores (30%) appearing in students' report cards. To date, the effectiveness of formative assessment therefore remains dubious.

YEAR	GRADE	ONGOING EVALUATION	MID-YEAR EXAMINATION	END-YEAR EXAMINATION
1986	PP- 5	30%	30%	40%
		CONTINUOUS ASSESSMENT (FORMATIVE ASSESSMENT)		EXAMINATION (SUMMATIVE ASSESSMENT)
1996- 2019	PP-6 (PRIMARY)	50 %		50%
1996-	7 -10 (LOWER & SECONDARY)	20%		80%
1996-	11-12	0%		100%

TABLE 1. SUMMARY OF ASSESSMENT PRACTICE AND WEIGHTING IN % ACROSS ALL GRADES

(adapted from Max et al., 2010; Utha,2014).

There are three types of assessments currently adopted in elementary classes: Continuous Formative Assessment (CFA), Continuous Summative Assessment (CSA), and Summative Assessment (SA) with slight variations in weighting according to subject and grade. Table 2 shows the current assessment practices across the elementary grades. In mathematics, science, and social studies, from Grade 4 to 6, continuous formative assessment, continuous summative assessment, and summative assessment are implemented. Continuous formative assessments are implemented using different assessment tools and techniques such as portfolios, checklist, anecdotal records, quizzes, debates, and presentations. Continuous summative assessments occur through unit/chapter tests, graded homework, and classwork; while summative assessment is conducted through written examinations. From Grade 4 to 6 in these subjects, 50% of students' total score comes from the continuous summative assessment with the remaining 50% derives from the midterm and annual examinations. For English and Dzongkha subjects scores are divided by 60% and 40% respectively, between continuous and summative assessments (examinations).

Grades	Subjects	CFA	CSA	SA
4-6	Maths, Science and Social Studies		50%	50%
	English and Dzongkha		60% (CFA+CSA)	40%
PP-3	Dzongkha, English, and Maths		100% (CFA)	

TABLE 2 CURRENT ASSESSMENT PRACTICE IN ELEMENTARY CLASSES, (REC, 2019)

On the other hand, as shown in Table 2, from Grades PP to 3 written examinations were removed from 2020. Consequently, there are no summative assessments for these grades, and all assessments occur through the tools and techniques of continuous formative assessment in all subjects: Dzongkha, English, and Mathematics. Table 3 further describes the continuous formative assessment practices from Grade PP to 3. There are a wide variety of assessment tools that attempt to capture all learning domains including cognitive, psychomotor, and affective (REC,2019). The approaches to execute the given assessment tools and techniques are integrated into subject textbooks and teacher manuals. Therefore, the latest development in the assessment practices for primary education is the removal of major written examinations from Grades PP to 3 in 2020 and from Grades 4 to 6 likely in 2021. This represents a major shift towards abolishing summative assessment and fully embracing continuous formative assessment- a change that raises concerns about teachers' competencies, readiness, and beliefs.

Assessment tools	Learning Domains		
	Cognitive	Psychomotor	Affective
Suggested Techniques	Interview, project work, portfolios, anecdotal, audio-visual, concept map, conferencing, test, etc.	Project work, field trip, observation, portfolios, scrapbook, anecdotal, etc.	Journal, scrapbook, self-assessment, observation, game- based assessment, field trip, portfolios, audio-visual, anecdotal, concept map, conferencing, etc.
Suggested Tools	Checklist, rubrics	Checklist, rubrics	Rating scale, rubrics
Approach	Integrated into the subject textbooks and guidebooks	Integrated into the subject textbooks and guidebooks	Integrated into the subject textbooks, guidebooks and as part of co-curricular activities
Weighting	20%	30%	50%

TABLE 3 CONTINUOUS FORMATIVE ASSESSMENT TECHNIQUES AND TOOLS MATRIX, (REC, 2019).

1.4. STATEMENT OF THE PROBLEM

The Bhutanese education system has made several essential shifts such as nationalising the curriculum, making the pedagogical shift from teacher-centered to student-centered education, and revising assessment practice from a system of 100% summative assessment in the form of examinations to a system of formative and summative assessment with a balanced weighting in elementary classes. Despite these essential shifts, however, the role of teachers as "Sage on the Stage" and teachers as the source of knowledge, has not changed much. According to Utha (2014), this practice is deeply rooted in the Bhutanese culture of respecting elders, particularly teachers. She explains that teachers' authority and knowledge are placed at the highest level in a classroom, which has a great influence on the culture of classroom teaching (p.5). Similarly, Keller and Utha, (2017) find that the education system in Bhutan in practice is not as modern as the word suggests, largely because teachers adopt teacher-centered pedagogies and stress to students the importance of acquiring knowledge

from textbooks to ultimately reproduce in the examination. Therefore, this indicates that traditional teaching methods with teacher-centered pedagogies are impediments to the effective implementation of formative assessment. Despite the recent transformations in the Bhutanese education system, teachers' beliefs about teaching, learning, and assessment therefore continue to impact their practices.

Additionally, empirical studies conducted on the examinations and assessment systems in Bhutan have noted gaps between current and expected levels of student learning. The REC (2019) reveals that the learning outcomes of students are below their grade levels, that students are unable to apply learning to real-life situations, that students are better at recalling rather than comprehending and applying higher-order thinking. Similarly, BCSEA (2015) reports that the NEA results are indicative of the poor quality of education. The results of the PISA-D assessment survey conducted in 2017 show the performance of 15-year-old students significantly below the OECD average (National Project Centre, 2019). These findings in the context of Bhutan indicate a low quality of education which may be significantly associated with the quality of teachers and their assessment beliefs, practices, and knowledge.

Another factor affecting the effective implementation of formative assessment in Bhutan could be the higher visibility of summative assessment. BCSEA (2015), observes that CSA and SA are more prevalent in assessment practices. It also reveals that the introduction of CFA did not alter the traditional assessment practices by the teachers. Further, the teachers viewed that the CFA impacted learning negatively (Utha, 2014). Meanwhile, the implementation of CFA is viewed as a challenge by many teachers according to the BCSEA (2015) which states that the teachers did not have clear concepts of CFA and its integration in the teaching and learning process. Therefore, the inadequate knowledge of assessment among teachers also needs to be addressed.

Above all, the assessment practices currently used in elementary classes in Bhutan are going through a transition phase by removing the examinations and fully adopting formative assessment in lower elementary grades (PP-3) in 2020 with the goal by 2021 to remove the same in upper elementary grades (4-6) (Palden, 2018). Following this change, the REC in collaboration with the Ministry of Education (MOE), organised a training of trainers (TOT)

event for 201 teachers who in turn then train teachers at the district level to ensure that elementary teachers can implement CFA effectively("Schools will implement," 2020). However, this cascading model of providing professional development to teachers has the limitation of not being able to fulfill teachers' needs about their specific environments and situations. A Cascading model (Ning et al., 2010.p.67) in this context refers to the practice of a few teachers attending Nationally -Based In-Service Programs (NBIP) then conduct the same workshop in their schools and districts. The evidence shows that this model has not been successful in preparing teachers professionally (Ning et al., 2010). Despite phasing out major examinations being one step forward towards the implementation of formative assessment, concerns, therefore, remain about teachers' competencies, readiness, and beliefs. Further, assessment training given through the cascading model seems to have not fully prepared elementary teachers as expected.

1.5. AIMS AND OBJECTIVES

The purpose of this study is to investigate Bhutanese elementary teachers' assessment beliefs, practices, and literacy levels, to gain a better understanding of formative assessment practices so that school leaders and policymakers can support the successful implementation of recent assessment reforms introduced in elementary classes. This purpose is guided by the following aims and objectives.

- Investigate the current beliefs about assessment held by elementary teachers in Bhutan using the Conception of Assessment (COA) Inventory developed by Brown (2006).
- Investigate Bhutanese elementary teachers' preferences about assessment types and practices.
- Examine the level of assessment literacy among Bhutanese elementary teachers as measured by Classroom Assessment Literacy Inventory (CALI).
- Analyse how teachers' assessment beliefs, practices, and literacy levels are affected by their gender, teaching experience, teacher education level, grade taught, subject taught and assessment education/training received.

- Explore the effects of teacher characteristics such as gender, teacher education, teaching experience, assessment training/education, teachers' and the two constructs, assessment beliefs, and teachers' assessment literacy on assessment practices.

1.6.RESEARCH QUESTIONS

The study is directed by the following questions:

1. What are the Bhutanese elementary teachers' beliefs/conceptions about assessment?
2. What assessment practices do Bhutanese teachers value?
3. What is the level of assessment literacy of elementary teachers in Bhutan as measured by the CALI?
4. Do the independent variables of gender, teacher education level, years of teaching experience, grade taught, subject taught and assessment education/training received to make a significant difference in teachers' assessment beliefs, assessment practice, and assessment literacy level?
5. How is the dependent plus endogenous variable of teachers' assessment practice affected by the independent variables of gender, years of teaching experience, level of teacher education, grade taught, subject taught and assessment education/training received and by assessment beliefs and assessment literacy?

1.7. SIGNIFICANCE OF THE STUDY

This study is significant as it addresses the need to investigate teachers' assessment beliefs practices and literacy level to understand and support the effective implementation of formative assessment in elementary grades. Accordingly, this study provides an insight into the connection between Bhutanese elementary teachers' assessment beliefs and literacy levels and their assessment practices. Thus, the information, results and recommendations presented by this study could potentially inform policymakers, teacher educators, school leaders, and the teachers themselves about the nature and connection between assessment beliefs and assessment literacy level and assessment practices. This information may be utilised to facilitate interventions and conduct sustained professional development programs

for the effective implementation of formative assessment practices to increase the quality of education in Bhutan and elsewhere.

A further significance is that there are few studies within the Bhutanese context and even fewer quantitative studies on assessments. To date, Bhutanese studies that have focused on assessments/formative assessment (e.g Max et al, 2010; Utha, 2014) have employed qualitative methodologies that are not generalisable to all Bhutanese teachers. This research adopts a quantitative methodology with previously and currently validated survey tools, and so is anticipated to generate unbiased results that can be generalised to all Bhutanese elementary teachers. Therefore, the results of this study will inform Bhutanese stakeholders about assessment beliefs, practices, and literacy, to support teachers in implementing the new assessment system introduced in elementary classes.

Teachers' assessment literacy level will inform teacher practice and improve their competencies in designing/executing formative assessment activities, sharing learning goals/ criteria, establishing dialogues, giving feedback, employing peer and self-assessment with the learners, and using the evidence to improve teaching/learning. The process of completing the survey questionnaires provided an opportunity for self-reflection by the respondents (teachers). Meanwhile, the results of this study may make teachers more aware of their own assessment beliefs, practices, and literacy levels.

To the knowledge of the researcher, to date, no studies have been conducted that examine together the relationships of the bigger constructs of assessments such as assessment beliefs, assessment practices, and assessment literacy. Further, investigating the effects of demographic and teacher characteristics on these three constructs will confirm or reject the findings of similar studies in the literature. Thus, by investigating the assessment beliefs, assessment practices, and assessment literacy level of elementary teachers, and exploring their relationships, this study will add an extra dimension to the existing knowledge in the literature about assessment.

1.8. OUTLINE OF SUBSEQUENT CHAPTERS

The following chapters include a literature review (Chapter 2), research methodology (Chapter 3), results (Chapter 4), and discussion including recommendations, limitations, and conclusion (Chapter 5).

1.9. DEFINITION OF KEY TERMS

- *Assessment:* A formal process of judging by determining a learners' status to certain predetermined educational variables (Popham, 2010)
- *Assessment beliefs/conceptions:* An indication of agreement or disagreement by teachers in response to self-reported beliefs and understanding statements of intentions or purposes of assessments. (Conceptions of Assessment III abridged version) (Brown, 2008)
- *Assessment literacy*: The expertise that teachers hold about the basics of measuring activities that occur in the classroom (Popham, 2009).
- *Assessment Practices:* The construction and design of a variety of assessment tasks/techniques and use of these techniques in teaching and learning processes by teachers (McMillan et al., 2002, p. 203) or implementation of assessment practices designed and published in guidebooks and manuals.
- *Formative assessment*: A process through which teachers and students in classroom teaching elicit information about students' learning through assessment and use this information as feedback, whereby instruction is modified towards improving the quality of performance (Black & Wiliam, 1998; 2009; Shepard, Hammerness, Darling-Hammond & Rust, 2005; McManus, 2008; Cauley & McMillan, 2010; Kippers, Wolterinck, Schildkamp & Poortman, 2018).
- *Summative assessment:* A "point-in-time judgment" made on students' learning tasks, evident from ongoing formal and informal assessment (Collbert & Cumming, 2014)

CHAPTER TWO: LITERATURE REVIEW

2.1 INTRODUCTION

Assessment is increasingly recognised as an inseparable entity of the teaching and learning process and a point of leverage for educational change (Allal, 1988 cited in Black & Wiliam, 2018; Shavelson et al., 2008; Van Staden & Motsamai, 2017). Assessment is significant in influencing the success and quality of education (Black & Wiliam 1998), in ascertaining the effectiveness of educational activities (Popham,2009), and in providing information at the classroom, program, and institution levels (Stiggins, 2009). Numerous studies on assessment are focused on how assessment can support teaching and learning. Notably, a meta-analysis review by Black and William in 1998 with 250 research articles on formative assessment has led to a widespread interest in the topic (Black & Wiliam, 2018). However, it may be argued that effective implementation of formative assessment largely depends on teachers' assessment beliefs, practices, and assessment literacy. Therefore, this literature review explores these factors by presenting the definitions of formative and summative assessment, their integrating relationship, assessment beliefs held by teachers, assessment practices valued by teachers, and level of assessment literacy of teachers.

2.2 SUMMATIVE ASSESSMENT

Summative assessment is a "point-in-time judgment" made on student's learning tasks, evident from ongoing formal and informal assessment (Collbert & Cumming, 2014). In the summative assessment, the evidence records only the current student achievement (Cauley & McMillan, 2010). This form of assessment is employed at the end of instruction to gather and interpret evidence of learning and then to grade as well as report learning (Dunn & Mulvenon, 2009;). Popham (2009) states that summative assessment is making use of evidence-based assessment to decide the worthiness or effectiveness of completed instructions or courses. It is intended to assist in decision making based on the success of a final version instructional program (Popham, 2009). Summative assessments make use of traditional assessment tools such as multiple-choice, true/false, completion of items, and matching questions, and consequently, this form of assessment is known as conventional or traditional assessment and these tools do not stimulate high order thinking that support quality learning for students (Kennedy, Chan, Fok and Yu, 2008).

2.3 FORMATIVE ASSESSMENT

Formative assessment is a concept that encompasses different methods of using the assessment in supporting student learning (Van der Kleij et al., 2015; Briggs, Ruiz-Primo, Furtak, Shepard, & Yin, 2012). It is a process through which teachers and students in classroom teaching elicit information about students' learning through assessment and use this information as feedback, with instruction modified to improve the quality of performance (Black & Wiliam, 1998; 2009; Shepard, Hammerness, Darling- Hammond & Rust, 2005; McManus, 2008; Cauley & McMillan, 2010; Kippers, Wolterinck, Schildkamp & Poortman, 2018). Wiliam (2009), asserts that an assessment can be formative if it evokes evidence that is interpreted about student learning and then applied to make adjustments in meeting learning needs. According to OECD (n.d, p.1), "(I)n classroom, formative assessment refers to frequent, interactive assessments of student progress and understanding to identify learning needs and adjust teaching appropriately." Berry and Kennedy (2008, p.49) mention that formative assessment consists of teachers and students conducting a series of activities to enhance learning, with the resulting provisional outcomes from these actions serving to inform them to adjust teaching and learning activities, with students directly engaged in these activities.

Definitions of formative assessment include three major components: a continuous gathering of information, using the information (feedback) to make instructional decisions, and using the information to support learning. Therefore, an assessment process without one of these components is not a formative assessment. Meanwhile, Formative assessment as a process involves four major components as identified in the literature, (a) explaining learning objectives and success criteria, (b) increasing the quality of dialogue, (c) increasing the quality of marking/feedback/record keeping, and (d) using self and peer-assessment (Black & Wiliam, 1998; Black, Harrison, Lee, Marshall & Wiliam, 2003; Wiliam & Thompson, 2007; Centre for Educational Research and Innovation (CERI), 2008; Wiliam, 2011; Swaffield, 2011)

2.4 INTEGRATING FORMATIVE ASSESSMENT AND SUMMATIVE ASSESSMENT

Many researchers argue that formative assessment and summative assessment can be used by integrating them for effective implementation of formative assessment (Black et al 2010; 2011; Looney, 2011). Besides supporting learning, formative assessment can be utilised to summatively assess the understanding of the learners (Bennett, 2011; Kingston & Nash, 2011). According to Birenbaum et al (2006), a learner's understanding may be evaluated at the terminals (end of a chapter, unit, term, year, of a course) by using the evidence collected during the instruction through formative assessment. The judgment then can be used to adapt subsequent instruction (Black & Wiliam, 2009; Perie, Marion & Gong, 2009). Masters (2013) states that a single assessment is useful for both the purposes of monitoring the progress of learning individually or in a group (summative assessment) and for identifying the point of intervention for future action (formative assessment). Dunn and Mulvenon (2009) distinguish assessment as an instrument and evaluation as a usage. Thus, they propose that an assessment may be designed to be either formative or summative and the data obtained by administering either of them may be utilised formatively or summatively. Bennett (2011) argues that the relationship between formative and summative assessment is too simplified when their relationship is indeed complex. The complex relationship implies that the primary purpose of documenting students' learning must be fulfilled through summative assessment as well as the secondary purpose of supporting learning. Bennett (2011) and Black and Wiliam (2018) recognise that it is essential to focus on the idea, that formative assessment is *assessment* if the formative assessment is to be helpful. It is conclusive that correct assessment design with its appropriate usage for summative or formative purpose calls for teachers' expertise and knowledge, for which assessment literacy and assessment practicing skills are crucial. Exploring teachers' assessment beliefs, assessment practices, and assessment knowledge, thus, is basic in clarifying the integrated relationship between formative and summative assessments.

2.5 TEACHERS' BELIEFS/CONCEPTIONS ABOUT ASSESSMENT

A series of studies were conducted by Brown from 2002 to investigate and capture the beliefs about assessment that teachers possess. Beliefs are the lenses that people use to make sense

of people, events, and interactions, whose meanings are connected to mental construct shaped by environment and culture (Thompson, 1992. Cited in Brown, 2004). Though Brown used the term conception to describe the beliefs of teachers, this study uses the term conception or beliefs interchangeably hereafter. Based on the literature, Brown identifies three purposes of assessment- assessment is used to improve teaching and learning, to hold students accountable, and to hold teachers and schools accountable. Further, Brown includes the conception that assessment is irrelevant to education arguing that assessment is irrelevant to teachers' life and work and students as well because assessment could be bad, easily ignored, or inaccurate. This series of studies by Brown developed and modified a four-factor tool called Conception of Assessment (CoA-III) with 50 items (Brown, 2004) and Conception of Assessment Abridged version (CoA-III A) with 27 items (Brown, 2006) to capture teachers' beliefs about assessment. This survey tool was subsequently validated in several contexts such as in New Zealand and Australia (Brown, 2006; 2008), Hong Kong (Brown, Kennedy, Fok, Ahan &Yu, 2009), Egypt (Gebril & Brown, 2014), China (Brown & Gao, 2015), and India (Brown et al.2015).

Brown (2008) conducted quantitative research through a self-administered survey with 525 primary school teachers in New Zealand, 784 primary school teachers, and 614 secondary teachers in Queensland using the CoA four-factor tool. The result showed that primary teachers from both countries agreed more with the conceptions of assessment as improving teaching and learning and strengthening student accountability, rather than the conceptions of assessment as strengthening school accountability or being irrelevant. Improvement of teaching and learning was positively correlated with school accountability (Queensland r = .20, New Zealand r = .21) and negatively correlated with irrelevancy (Queensland r = -.38, New Zealand r = .36). On the other hand, irrelevancy was positively correlated with student accountability(Queensland r = .68, New Zealand r = .36) (Brown, 2008, p. 109). More than the secondary school teachers, primary school teachers in both contexts agreed with views that assessment is irrelevant. Possibly because in primary schools in both contexts, assessments are used not necessarily for grading, while in secondary classes teachers use assessment for grading and making students accountable (Brown, 2011). It may be understood that assessment is an applicable means of improving teaching and learning at the same time keeping the students accountable. This also means that improved teaching and learning is an

aspect of quality schooling (Brown, 2008). Therefore, this study confirmed the pre-identified four major beliefs that teachers hold about assessment- assessment for accountability of students; assessment for accountability of schools and teachers, assessment for improving teaching and learning, and assessment is irrelevant to education (Brown, 2008. p. 153).

Subsequently, Brown, et al., (2009) researched in Hong Kong employing a cross-sectional survey with nearly 300 teachers from 14 primary and secondary schools, the study proposed to test the validity of the CoA instrument, develop and validate an inventory of assessment practices, examine the linkages between beliefs and practice of assessment as reported by participants. The results showed a strong positive relationship (r=.91) between the conception of improvement and that of student accountability (p.354), the highest correlation compared to previous results from New Zealand and Queensland. This finding suggests that Hong Kong teachers understood improvement and accountability to be inseparable. The most agreed-upon belief about assessment in three (New Zealand, Queensland, and Hong Kong) contexts was the conception of improvement. The belief that assessment was related to student accountability was agreed more than its relationship with school accountability, and all three groups rejected the idea that it was irrelevant (Brown et al, 2009). The findings in Hong Kong are significant because they highlight the importance of cross-cultural research. Hong Kong is a more examination driven society than New Zealand and Australia (Brown et al., 2015), and Hong Kong teachers seem to strongly emphasis students' accountability (evaluation) and believe that the more students are accountable for their learning, the more they can improve their learning and performance.

A consistent finding was discovered in a study conducted by Brown et al (2011) in South China province. There was a strong positive correlation (r=80) between student accountability and improvement demonstrated in the assessment beliefs of Chinese teachers (p.42). In the context of long-standing practices of examination-driven decisions (Brown et al., 2011), both the Hong Kong and Chinese teachers seemed to believe that students' learning can be improved by examining and making them accountable. These two studies suggest that Bhutanese teachers may hold similar conceptions as public examinations (for Grades 10 and 12) are given strong emphasis as pathways to public schools and university education.

With similar purposes but a different approach, Brown and Gao (2015) studied Chinese teachers' conceptions of assessment for and of learning. The study synthesised eight interview and survey studies in which the nature and purpose of assessment were described by a diverse sample of Chinese teachers, using inductive analysis and factor analysis. The study identified six major constructs ranging from positive to negative roles of assessment. The roles of assessment identified were:

- assessment develops personal qualities as humans,
- assessment develops learning abilities through motivations,
- assessment diagnoses the effectiveness of teaching and allows alteration towards improvement,
- the assessment allows teachers to confirm that students have achieved pre-set learning targets,
- assessment is for inspection and control of schools, teachers, and students
- assessment is inaccurate and comes with errors hence it is viewed negatively (Brown & Gao, 2015, p.2).

The roles of assessment range from individual development through institutional management and control. As in previous studies, the findings here are dichotomous, with conceptions that assessment can be for improvement (development) and accountability (management and control). While the study established a conceptual framework to understand the beliefs about assessment among Chinese teachers, it is not highly credible as the data was sourced from studies by graduate students, including two masters-level dissertations (Brown &Gao, 2015). Dissertations are not reliable academic sources for high-quality research and are normally conducted by beginning researchers. Nevertheless, the study's identification of the improvement-accountability continuum of the purpose of assessment still resonates.

In addition to Brown and his colleagues' research, Remesal (2011) conducted a study to explore practising teachers' conceptions of the functions of assessment in basic education in Spain. However, as the CoA survey tool (Brown, 2008) did not suit Spanish teachers, she developed her continuum of assessment purposes. The study adopted a qualitative approach by interviewing 30 primary schools and 20 compulsory secondary mathematics teachers,

focusing on four dimensions of assessment namely the learning and teaching processes, accreditation of learning, and teachers' professional accountability (Remesal, 2011, p. 475). The study established a two-pole continuum with pedagogical and societal functions at either end, placing the four dimensions of assessment in between. Further, Remesal (2011) articulated these four dimensions into pure and mixed conceptions on the continuum, arguing that these dimensions are not distinct and that together they form a conception and so can hardly be separated. Nevertheless, Remesal (2011) similar to Brown (2008) discovered that assessment has two dimensions- pedagogical (improving learning and teaching) and societal (accountability).

Similarly, Barnes et al. (2015) investigated the empirical research on pre-service and in-service teachers' beliefs about assessment. Their study selected 28 peer-reviewed, empirical articles published after 2000 that investigated pre-service and in-service teachers' beliefs about assessment. The study gave an overview of the research conducted on teachers' beliefs about assessment and organised the conceptions reported in these studies on a continuum of purposes, from pedagogical conceptions to accounting conceptions (Barnes et al. (2015). The review presented three findings on teachers' beliefs about assessment. First, as claimed by Brown and Harris (2009), belief systems about assessment formed by teachers are shaped by the legal frameworks and sociocultural priorities of their societies. Second, the review noted that some studies have made a distinction between knowledge of assessment and beliefs about assessment in their research into assessment conception. The review, therefore, cautions that asking teachers about the conceptions about assessments may reflect their knowledge rather than their beliefs. Consequently, it appears that differentiating teachers' assessment knowledge and beliefs may indicate distinct explanations of assessment practices, an aspect that is analysed in this research project through investigation of both assessment beliefs and knowledge (literacy). Third, Barnes et al (2015) conclude that assessment can either enhance learning or be used to punish and control students, teachers, and schools. Thus, here too the study identifies two purposes of assessment; improvement and accountability.

Deviating from the above studies, Postareff, Virtanen, Katajavuori, and Lindblom-Yla¨nne (2012), without using the CoA survey tool (Brown, 2006), analysed teachers' beliefs about the

purpose of assessment using the data collected from semi-structured interviews asking open-ended questions with 28 pharmacy teachers. From the analysis, Postareff et al. (2012) categorised the beliefs of the pharmacy teachers in a continuum ranging from reproductive beliefs with stress on reproducing correct information to transformational beliefs emphasising students' constructive thinking and understanding. Thus, the continuum has reproduction on one end and transformation on the other end. Although this study is conducted in higher education, this finding suggests that teachers may hold assessment beliefs based on their education level and the disciplines being taught.

Based on the literature on beliefs about assessment purposes, these beliefs can be presented on a continuum of assessment with two ends, as demonstrated in Table 4 below. At one end is a formative assessment while at the other end is a summative assessment.

TABLE 4 ASSESSMENT BELIEFS CONTINUUM (FORMATIVE ASSESSMENT TO SUMMATIVE ASSESSMENT)

Formative ⟷ Summative

Authors	Formative			Summative		
Black & Wiliam(1998a)	Assessment for learning			Assessment of learning		
Brown (2002;2004;2008)	Improvement of teaching	Improvement of learning	Student accountability	Teacher/school accountability		Irrelevant
Remesal (2007; 2011)	Pedagogical	Mixed pedagogical	Mixed undefined application	Mixed societal	Societal	
Postaerff et al (2012)	Transformational			Reproductive		
Brown et al (2015)	Personal quality	Ability development	Facilitation diagnostic	Institutional target	Management & Instruction	Negativity
Barnes et al (2015)	Extreme pedagogical	Mixed		Extreme Accounting		

Black and Wiliam, in their series of studies on the assessment, have stressed on formative and summative functions of assessment, and suggest an integration of the two to ensure that assessments are effective. Likewise, the studies of Brown and his colleagues present four major beliefs about assessment; that it improves teaching and learning, enhances the accountability of students, enhances the accountability of teachers or that assessment is irrelevant. The belief that assessment is for improving teaching and learning is a formative purpose, while the beliefs related to accountability are summative purposes. Likewise, Brown et al (2015) placed the purposes of assessments from positive to negative, which may be viewed as formative purposes (positive; development of personal quality and abilities) and summative purpose (negative; meeting institutional targets, management, and instruction). Remesal (2011) places pedagogical purposes at one end depicting the formative function of assessment and societal purpose at the other end depicting summative functions. Similar to Remesal (2011), Barnes et al (2015) categorise the purposes of assessments under the categories extreme pedagogical (formative), and extreme accounting (summative), with any belief in between these extremes placed between the two poles. Postaerff et al., (2012) on the other hand, identify two major purposes for assessment: transformational (formative) and reproductive (summative). It may be, therefore, concluded that there could be commonly believed purposes such as improvement (formative) and accountability (summative) of assessment, but there could be other overlapping beliefs of assessment that teachers may hold due to various factors such as culture, policies, frameworks, subjects, and levels taught.

2.6 ASSESSMENT PRACTICES

The beliefs that people hold and the norms they are bound with, greatly shape the behaviour type and practices (Brown, 2008). Teachers' beliefs affect their teaching more than the socioeconomic status of a school and their teaching experience (Griffiths, Gore, and Ladwig, 2006). The studies on assessment beliefs have variously identified two general purposes of assessment: improvement and accountability (Brown, 2008; Brown et al. 2015) pedagogical and societal functions (Remesal, 2011); reproductive and constructive functions (Postareff et al., 2012) assessment for learning and assessment of learning (Azis, 2015). The literature,

therefore, suggests that assessment has either formative functions (i.e its function is to improve learning) or summative functions (i.e its function is to evaluate students).

Based on these overarching purposes, classroom teachers seem to employ a wide range of assessment practices to serve either or both of these major purposes. One of the key studies exploring the assessment practices of teachers was conducted by McMillan et al (2002). Their study is significant for the current research as the survey questionnaires to investigate the assessment practice of the respondents have been adopted from their study. McMillan et al (2002) investigated the assessment practices and grading systems in mathematics and language arts, with 901 third to fifth-grade teachers in Virginia. There were three major types of assessments used: constructed-responses, such as projects, essays, and presentations; objective assessments including multiple choice and short answers; and teacher-constructed major examinations (McMillan et al., 2002). The findings showed that the participants adopted an array of tools to assess student performances in mathematics and language arts. The most frequently practised assessments were quizzes (Maths M= 3.93 and language arts M= 3.80) followed by objective assessments for both the subjects (Maths M=3.82 and Language arts M=3.75). Performance assessment was practised more in language arts (M= 3.43) than in maths (M=2.84), while authentic assessments scored almost the same mean (Maths M=2.95, Language arts M= 2.89) in both the subjects. This result also demonstrated that teacher-made assessments (Maths M= 3.63 and language arts M= 3.90) were utilised more than the publisher supplied assessments (Maths M= 3.54 and language arts M= 3.22) (McMillan et al., 2002, p.207).

Combining CoA-III (Brown, 2006; 2008) and McMillan et al (2002) questionnaires, Calveric (2010) in her dissertation used quantitative methods to examine the assessment beliefs and practices of 79 elementary teachers teaching from third through fifth grades in the Commonwealth of Virginia. In terms of assessment beliefs, assessment for improvement received the highest mean (M=4.18), and assessment as irrelevant received the lowest (M=3.43). The author adapted and utilised the survey tool designed by McMillan et al (2002) to examine which assessment types the participants valued. The teacher participants (51%) valued authentic assessments (real-world examples) more than other types of assessments. They felt publisher assessments and major exams were *"Not Important"* (11.5% and 6.1%

respectively). However, the participants recognised the following assessment types as valuable in classroom assessment; assessments designed by self, performance quizzes, objective assessments such as multiple choice and matching, short answer assessments, performance assessments, authentic assessments, and oral presentations, none of the types of assessment received the rating of *"1"-Not Important* in the Likert scale of 1 to 5. Publisher assessments received the lowest mean (M=2.69), while performance and authentic assessments received the highest (M=4.01 and 4.32) respectively. Similarly, a mean of 3.8 (M=3.8) was found for assessment designed by the teachers and short answer assessments. Although the beliefs and types of assessments the participants preferred are consistent, these results do not guarantee that the teachers studied practised the types of assessments that they reported. Self-reported and real-time assessment practices may not be the same.

Additionally, Azis (2015) conducted a study, employing a mixed-method methodology, using the Teacher Conception of Assessment (TCoA, Hong Kong version) survey (Brown et al., 2010), with 107 English junior high teachers in one of the regions of Indonesia. Unlike the results revealed in the study of Calveric (2010), the result of this study revealed an inconsistency between the participants' assessment conceptions and practices. The teachers expressed a stronger conception of the purpose of assessment for improvement (M=4.99) more than assessment for accountability (M=4.66), and almost all of the participants disagreed with the conception that it is irrelevant (M=1.94). However, their assessment practices predominantly used the traditional assessments, such as paper-pencil tests including the test items such as multiple-choice, matching, fill-in/item completion, short answer, and essay type questions. This study strengthens the argument that self-reported practices and real-time practices may not be consistent.

Along with assessment beliefs, Brown et al., (2009) identify five factors of assessment practice employing the Practice of Assessment Inventory which was administered for the first time in a Confucian society, Hongkong. The five factors were "practices that diagnose student learning needs, practices that use assessment to prove school quality, practices that prepare students for high-stakes examinations, practices that improve, change or adapt teaching in response to assessment information and practices that ignore or treat as irrelevant assessment information." (Brown et al. 2009, P. 353). Further, the study found that beliefs of

assessment influence assessment practices. The assessment practices of diagnosing student learning needs were aligned with the conception that assessment is for improvement. Similarly, preparing students for examination was aligned with the conception of assessment as making students accountable. Using examination to evaluate the quality of school was aligned with the conception of school accountability and the idea that assessment is irrelevant (Brown et al. 2009). These results are, therefore, indicative of a strong prediction of assessment practices through the assessment belief system.

2.7 ASSESSMENT LITERACY

Teachers' assessment literacy is considered essential for quality assessment. Mellati and Khademi, (2018) assert that, among other factors, teachers' knowledge of student assessment is the central factor in student learning. Popham (2009) asserts that inadequate knowledge of classroom assessment on the part of teachers can greatly harm the quality of education students receive and reports that most teachers know little about educational assessment. Webb (2002, p. 4) defines assessment literacy as the knowledge of resources that assess what students know and can do, knowing how to interpret the results generated from these assessments, and having the knowledge of applying the results in improving students' learning and the effectiveness of programs. Popham (2009) refers to assessment literacy as the expertise that teachers hold about the basics of measurement of the activities that occur in the classroom. Stiggins (1995, p. 240) provides a short but practical definition of assessment literacy: "Assessment literates know the difference between sound and unsound assessment". Englsen and Smith (2014) elaborate on this conception with a definition that assessment literacy is reflected in the quality of assessment and on the choices made about assessment. Therefore, they assert that quality of assessment literacy is not represented by the scores in theoretical tests, but by the practicing of assessment based on informed decisions (Englsen & Smith, 2014).

Englsen and Smith (2014) argue that knowledge of assessment is necessary, but that this knowledge must be supported by pedagogical knowledge of learning and assessment so that teachers' actions are not directed by technical prescriptions, but rather by their informed decision from such knowledge. Popham (2009) asserts that since assessment literate teachers can make better classroom and accountability assessment decisions, and such decisions are

subsequently visible on how the students are taught, teachers must obtain greater assessment literacy. Assessment literate teachers are aware of a variety of assessment choices and can create more appropriate assessments. Therefore, all teachers necessarily should possess assessment literacy so that their classroom practices are considered effective. For the well-being of teachers themselves and their students' educational well-being, assessment literacy is like a "commodity" for teachers (Popham, 2009). Further, Stiggins (2014) shares that nearly a third of teachers' professional time is utilised on assessment-related activities. Thus, to have assessment literacy/knowledge is a requirement for teachers (Popham, 2009) and as such Popham (2013, p.13) appeals for every teacher to be a “skilled user of” formative assessment.

One of the commonly used frameworks to measure the assessment literacy level is the Seven Standards developed by AFT, NCME, and NEA (1990). Given the importance of assessment in good teaching AFT, NCME and NEA (1990, pp.2-5) developed seven assessment principles known as "standards" that teachers are required to possess to enable sound assessment practice. These standards correspond to the principles proposed by Rowntree (1987) and Stiggins (1999a). The seven standards include (1) choosing appropriate assessment methods, (2) developing appropriate assessment methods, (3) administering, scoring, and interpreting the results of assessments, (4) using assessment results to make decisions, (5) developing valid grading procedures, (6) communicating assessment results and (7) recognising unethical or illegal practices.

Some studies have been carried out to investigate the assessment literacy level of various stakeholders in education such as pre-service and in-service teachers, principals, district officers, and policymakers. However, most of these studies have been conducted about teachers (Englsen & Smith, 2014). Stiggins (2014) argues that vibrant assessment is possible only if teachers, educational leaders, and policymakers are all assessment literates. Stiggins (2018) further asserts that if dependable evidence has to be elicited through assessments to inform correct educational decisions and prevent the failures of assessment purposes, practitioners and policymakers should not lack "assessment know-how" (p.18). Stiggins (2014; 2018) and Englsen and Smith (2014), argue that assessment literacy should be

discussed not only about teachers but concerning all stakeholders in education including students, school leaders, policymakers, and parents.

Stiggins (2014) observes low levels of assessment literacy among in-service teachers and leaders. Several studies are carried out using the standards developed by AFT, NCME, and NEA (1990), that support Stiggins (2014) observation. For instance, one of the earliest and foundational studies was conducted by Plake, Impara, and Fager (1993). Their sample included 555 teachers and 286 administrators. The raw scores that the participants obtained from the survey instrument were considered the level of assessment literacy with the results showing a mean score of teachers as 23.20, with nearly 66% of the responses correct. This result was below the minimum benchmark of 70%. It was also found that teachers demonstrated better literacy related to Standard 3 (administering, scoring, and interpreting the results of assessments) and poor literacy related to Standard 6, (communicating assessment results). Similarly, Mertler (2005), conducted a parallel study to compare the level of assessment literacy of 67 pre-service teachers and 101 in-service teachers. The pre-service teachers on average scored a little less than 19 out of 35 items correctly, while the in-service teacher scored a little lower than 22 out of 35 items. The results for both groups were not substantially different from those found by Plake et al. (1993), with scores of less than 70%. Additionally, the pre-service teachers were more literate in Standard 1-(choosing appropriate assessment methods) but less literate in Standard 5- (developing valid grading procedures). Meanwhile, in-service teachers were found to be more literate in Standard 3- (administering, scoring, and interpreting), confirming the findings of Plake et al (1993). They were, however, also minimally literate in Standard 5- (developing valid grading procedures), which was unlike the result found by Plake et al (1993). Both studies are significant as the current Classroom Assessment Literacy Inventory (CALI) to measure teachers' assessment literacy is derived from these studies.

Yamtim and Wongwanich (2014), investigated the level of classroom assessment literacy of primary school teachers using the AFT, NCME, and NEA's (1990), Seven Standards in Thailand. The participants included 19 primary school teachers who completed the Classroom Assessment Literacy Questionnaire (Mertler, 2003; 2005) and 8 teachers in the focus group. Thus, the study employed a mixed method of descriptive statistics (quantitative) and content

analysis (qualitative). The results revealed a low level of assessment literacy among the primary teachers in Thailand. The mean score was 17.11 points, with a standard deviation of 3.62 (total score of 35). Standard 1 received the highest mean score by the in-service teachers, unlike in the studies conducted by Plake et al (1993) and Mertler (2005). Standard 5 received the lowest mean score similar to the finding by Mertler (2005) about in-service teachers. According to the data from the focus group, Yamtim and Wongwanich (2014) suggest that promoting cooperative learning and teamwork among practitioners with someone who can coach them may improve the assessment literacy of primary school teachers. Rosas (2014) also revealed a low level of assessment literacy among the elementary teachers and principals in Central Valley Public Schools in the USA. However, principals obtained higher mean scores than that of the teachers (teachers 19.03 and principals 23.14). Likewise, Hailaya (2014) who conducted a study on assessment literacy of teachers in the Philippines also showed similar results where both elementary and secondary teachers demonstrated relatively low levels of assessment literacy.

In a study conducted by Ogan-Bekiroglu and Suzuk (2014), pre-service physics teachers showed gaps between assessment theory and practice, despite having quite a high level of assessment literacy. Therefore, the authors suggest that teacher education programs should not only teach assessment theories but also the types of evaluation, validity, and reliability and provide opportunities to engage in reflecting, practising, and revising methods of both traditional and performance-based assessments. Masters (2013) also observes that current teacher education courses treat assessment at a low level, by teaching concepts and distinction from 20th-century textbooks, most of which hamper thinking about assessment and further transmit simple and inaccurate views of assessment. Together, these studies demonstrate drawbacks in giving assessment education to pre-service teachers. However, Popham (2009) suggests that, until pre-service teachers are adequately provided with meaningful assessment literacy, professional development programs must offer quality assessment literacy programs. Therefore, professional development programs are identified as crucial interventions to change assessment beliefs, practices, and literacy (Cumming & Kleij, 2016; Song & Koh, n.d; Koh, Burke, Luke, Gong, and Tan, 2018). On the other hand, Deneen & Brown, (2016) in their study with 32 in-service and pre-service teachers noted that teacher education programs increased the assessment knowledge (i.e literacy) but not

conceptions. The study suggests that higher assessment literacy level does not necessarily change assessment beliefs, but both may influence teachers' assessment practices. Therefore, this study provides scope for further research on teachers' assessment beliefs and assessment knowledge to determine their practices.

CONCLUSION

This literature review draws four major themes such as formative and summative assessment, teachers' assessment beliefs, assessment practices, and assessment literacy. The roles of formative assessment on teaching and learning are well acknowledged. At the same time, it is noted that the successful implementation of formative assessment has remained a challenge in many education systems including Bhutan. Black and Wiliam (1998b) state that altering existing assessment practices is not straight forward. Further, the literature reveals the difficulty of implementing formative assessment or assessment for learning (Black & Wiliam, 2009; Brown et al., 2009). According to Wiliam et al (2004), the conventional practice of high stakes and mandatory state tests that exist in most countries make the successful implementation of formative assessment even more difficult. Owing to these challenges, this study argues that investigating teachers' assessment beliefs, practices, and assessment literacy would give an insight into the implementation of formative assessment in Bhutan and elsewhere. The literature review identified four teachers' assessment beliefs, several assessment practices that may be categorised into formative and summative assessment, and the level of assessment literacy of teachers. Thus, the investigation for this study is based on these assessment variables and their relationship to better understand Bhutanese teachers' implementation of formative assessment practices in the Bhutanese education system, particularly in elementary grades.

CHAPTER THREE: METHODOLOGY

3. 1. INTRODUCTION

This study, first in the Bhutanese context is conducted to investigate Bhutanese elementary teachers' beliefs, practices, and literacy about assessment, to gain better understandings of teachers' beliefs and practices of formative assessment. Through the findings of this study, it is anticipated that school leaders and policymakers can support the successful implementation of recent assessment reforms introduced in elementary grades. Specifically, the findings of this study may provide empirical evidence for policymakers to formulate guiding policies for formative assessment, for the district and school leaders to plan and organise professional developments on formative assessment, and for teachers themselves to improve their implementation of formative assessment. The elementary teachers investigated in this study are those who teach any subject (Dzongkha, English, Mathematics, Science, and Social Studies) in grades from Pre-primary (PP) to Grade 6.

3.2 CONCEPTUAL FRAMEWORK

A conceptual framework in a quantitative study is a visual representation in the form of a diagram that illustrates the relationships between the identified variables of a study (Mugizi, 2019, p. 75). The variables and their relationships for this research project are determined from past studies. This study identifies independent variables, mediating variables, and dependent variables for analysis. Independent variables are those that would influence an outcome (Creswell, 2014. P. 52). In this study, the independent variables are demographic information such as gender, level of teacher education, years of experience in teaching, subject taught, grade taught, and the assessment education/training received. These independent variables had been chosen to explore any significant differences that may appear in shaping the identified dependent variables. These independent variables were determined considering previous relevant studies on teachers' beliefs and practices about teaching, learning, and assessment in elementary grades, and from the studies on the assessment literacy of the teachers and principals. Brown (2004, p. 104) in his study examined

the influence of the variables "teacher sex, ethnicity, years of teaching experience, role in the school, years of teacher education, school type, and types of teacher education" on assessment conceptions of New Zealand teachers. Similarly, Clevair (2010, p. 41) identified "years of experience, grade level assignment, level of education and assessment training" as independent variables to explore their relationship with assessment beliefs and practices of teachers in Virginia, USA. Likewise, Rosas (2014, p.5), recognised "years of teaching experience, grade level assignment, level of education and intensity of assessment literacy training" as the independent variables that would influence the assessment beliefs, practices, and literacy of principals and teachers of Central Valley, California.

Dependent variables are the outcomes of the influence of independent variables (Creswell, 2014, p. 52). Based on the existing literature and the aims of this project, the constructs of assessment beliefs, assessment practices, and levels of assessment literacy were determined to be the dependent variables. Additionally, the current study chose to explore the relationships of all the variables with assessment practices. Therefore, the dependent variables of assessment beliefs and assessment literacy were identified as mediating variables. Mediating variables are those that mediate the effects of independent variables on dependent variables by standing in between the independent and dependent variables (Creswell, 2014, p. 52).

Having identified the independent, mediating, and dependent variables, a conceptual framework (Figure 1) was designed to illustrate the relationship between the two sets of independent and dependent variables. The construct of assessment beliefs is measured by four dimensions of assessment: assessment for improving learning/teaching, assessment for student accountability, assessment for school accountability, and assessment is irrelevant (Brown, 2004; 2006; 2008,). The construct of assessment practice is measured by formative assessment and summative assessment practices (Clevair 2010; Rosas 2014) and assessment design. Lastly, the construct of assessment literacy is measured by The Seven Standards for Teacher Competence in the Educational Assessment of Students (AFT, NCME, & NEA, 1990) of America (Rosas, 2014; Ryan, 2018).

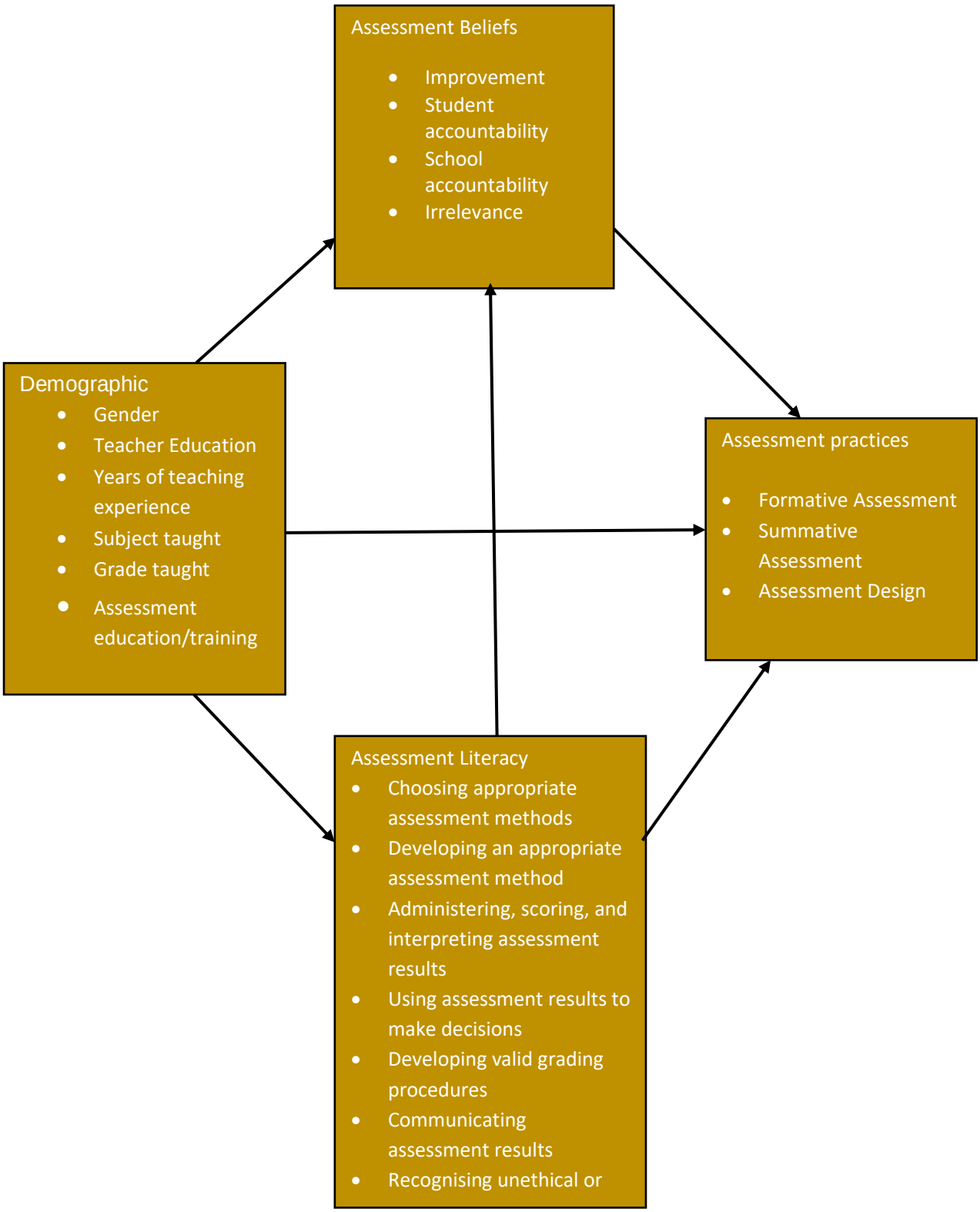

FIGURE 1 CONCEPTUAL FRAMEWORK

3.3 HYPOTHESIS

A hypothesised model, Figure 2 was created from the conceptual framework (Figure 1) to represent the hypothesis of this study. However, to avoid redundancy (Creswell, 2014), hypothesis statements are not written but represented by the research questions.

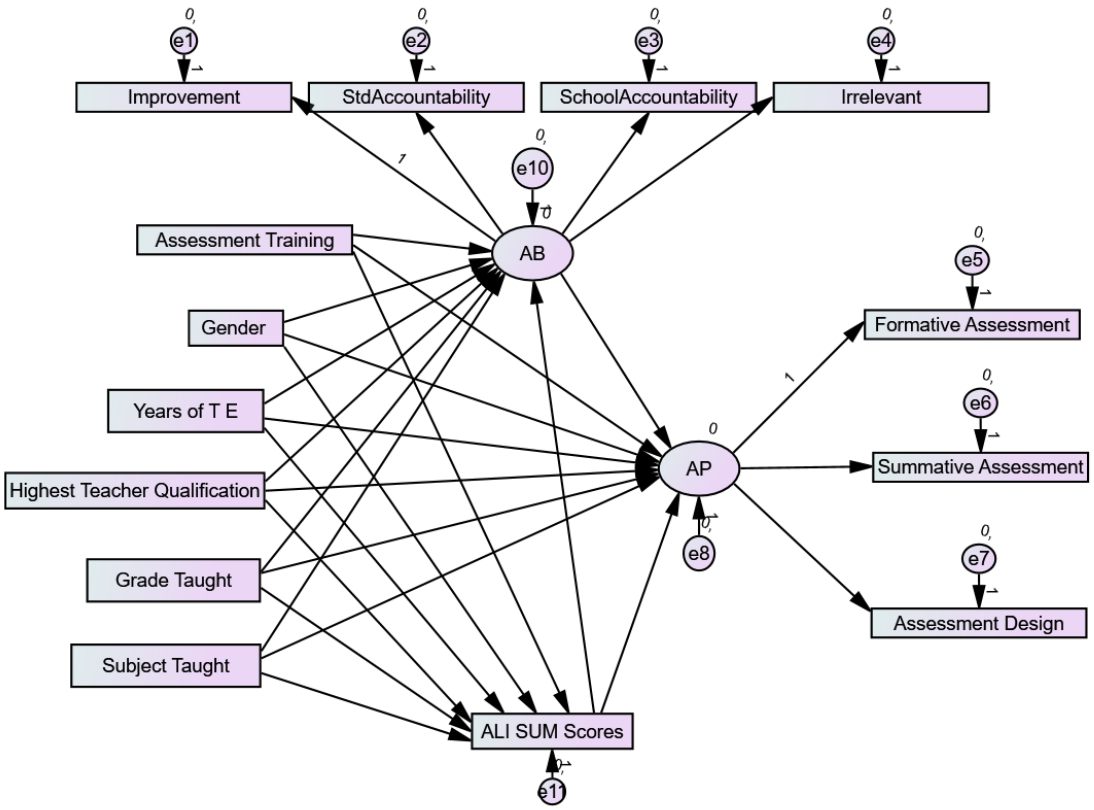

FIGURE 2 A HYPOTHESISED MODEL

3.4. RESEARCH QUESTIONS

The following research questions were developed to guide this study's investigation of Bhutanese elementary teachers' assessment beliefs, assessment practices, and assessment literacy, and to explore their relationship to gain better understandings of teachers' beliefs and practices of formative assessment to facilitate effective implementation of recent formative assessment policies.

1. What conceptions/beliefs of assessment do Bhutanese Elementary teachers have?

2. What assessment practices do Bhutanese teachers value?
3. What is the level of assessment literacy of elementary teachers in Bhutan as measured by the Classroom Assessment Literacy Inventory?
4. Do the independent variables of gender, teacher education level, years of teaching experience, grade taught, subject taught and assessment education/training have significant differences in teachers' assessment beliefs, assessment practices, and assessment literacy level?
5. How is the dependent plus endogenous variable of teachers' assessment practices affected by the independent variables of gender, years of teaching experience, level of teacher education, grade taught, subject taught, and assessment education/training; and by the exogenous variables of assessment beliefs and assessment literacy?

3.5. RESEARCH APPROACH: QUANTITATIVE APPROACH

This research employed a quantitative approach to investigate teachers' beliefs, practices, and literacy about assessment. A quantitative approach seeks to quantify and analyse variables to generate results by using numerical data and statistical techniques (Williams, 2011; Leavy, 2017). This research project analysed the attitude data of self-reported assessment beliefs and assessment practices, and the performance data of assessment literacy. Further, it explored the relationship between the independent variables of demographic information and dependent variables of assessment beliefs, practices, and literacy. Therefore, correlational design was used (Creswell, 2012). A quantitative approach was suitable for this study which provided opportunities for respondents to agree or disagree with self-reported assessment beliefs, rate the value of particular assessment practices, and undertake a cognitive test of their assessment literacy. Further, the quantitative analysis allowed the collection of data from a relatively large sample size (n=112), as well as facilitating the calculation of a generalisable result to the population of Bhutanese elementary teachers, and remote collection of data during the COVID-19 pandemic. Additional advantages were the lower costs of completing the research project within a limited time, and the quick collection and analysis of data (Carr, 1994 cited in Rahman, 2016)

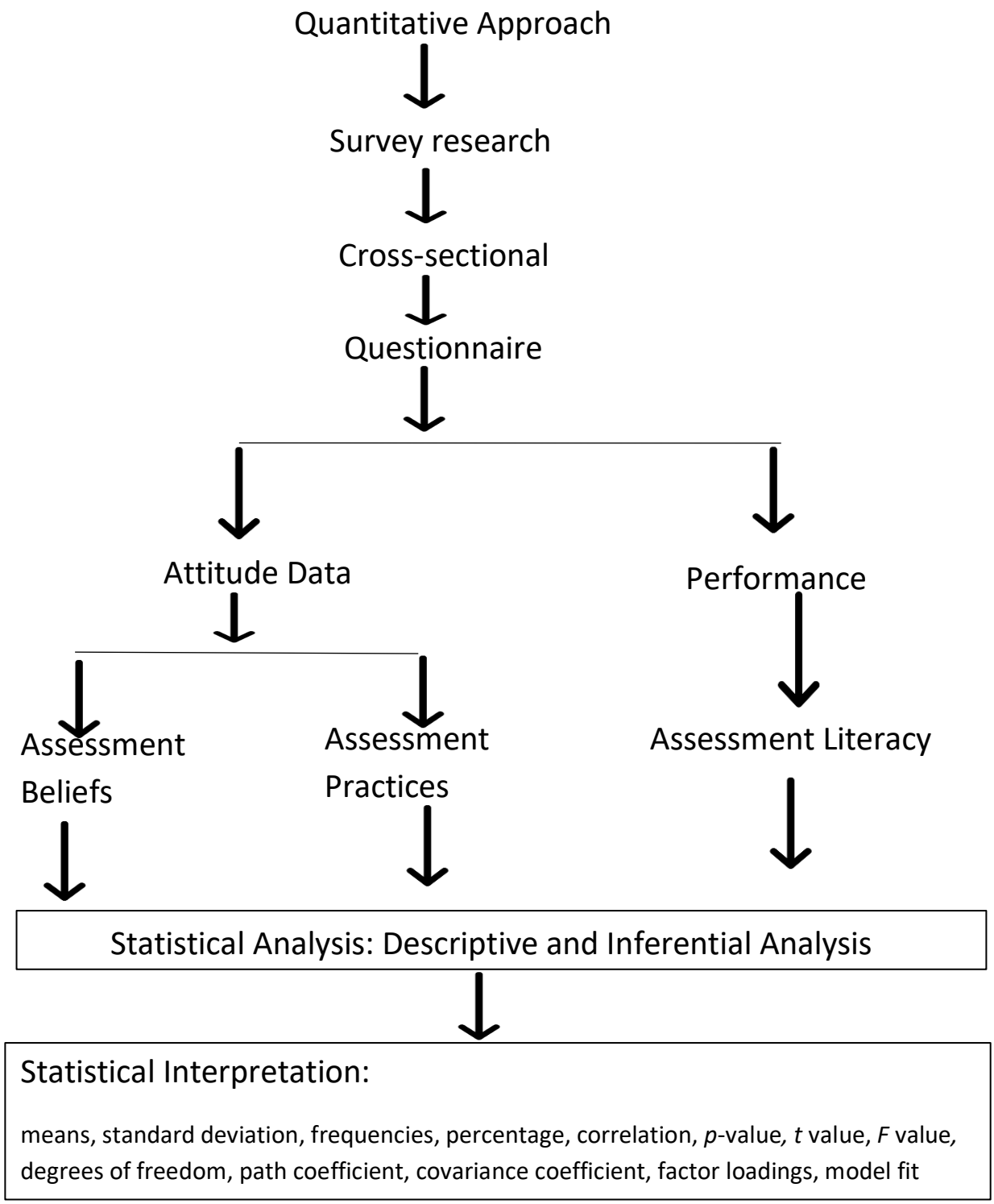

FIGURE 3 QUANTITATIVE METHODOLOGY FRAMEWORK WITH SURVEY RESEARCH DESIGN

3.6. RESEARCH DESIGN: THE SURVEY DESIGN

This study employed a survey approach as the quantitative research design, as shown in Figure 3. As it is not possible to monitor human thinking through non-experimental research, inferential procedures such as self-reporting in the form of discussions, interviews, or questionnaires can be used (Brown, 2008). According to Creswell (2014, p. 13), a survey provides a numeric or quantitative description of the attitudes of a population through the study of a sample. The purpose of the survey design in this study, therefore, was to capture teachers' beliefs, practices, and assessment literacy, and generalise from a sample of Bhutanese elementary teachers to its population to make inferences regarding their assessment beliefs, practices, and literacy. A survey is a method of collecting data through which information is elicited directly from people and can gather valid and reliable data in a systematic and structured form that facilitates efficient analysis and reporting (Parveen & Showkat, 2017). Further, a survey was the preferred collection procedure for this study because of its low cost and the fast turnaround in data collection (Creswell, 2014, p. 157).

According to Mertler (2019), the increasing use of the internet has greatly expanded the adoption of web-based surveys in collecting data in recent times. This study utilised a cross-sectional (one point of time) web-based online self-administered survey to collect data. A web-based survey is a survey that is typically shared through a website, with respondents directed to the website through initial contacts via email (Mertler, 2019) or social media. The web-based survey program Survey Monkey was chosen for this study. Several advantages were considered in choosing an online self-administered survey. Self- administered surveys are ideal in collecting demographic and personal information from large samples or populations (Fowler, 1993 cited in Brown, 2008). They allow for privacy and anonymity in responding to the survey items (Brown, 2008). Additional advantages of data collection from remote participants in a short time with a minimal economic expense, and to generalise the research findings to the broader elementary teacher population were also considered (Brown, 2008). As SurveyMonkey is known for having multiple layers of firewalls and security, and the data can be downloaded in multiple forms and transferred directly to SPSS, the

anonymity of the respondents can be strongly maintained, and the program is available at minimum cost (Cleviar, 2010).

3.7. POPULATION AND SAMPLE

The target population of this study was elementary teachers, working in government schools in Bhutan who teach any subject from Pre-Primary to grade 6 (Dzongkha [national language], English, Mathematics at Pre-Primary to Grade 3; and, in addition to these, Science and Social studies in Upper Primary). In 2019, 2,296 teachers were teaching in elementary grades, according to Annual Education Statistics (MoE, 2019. p.20). The study intended to have 200 participants, which would be approximately 10 % of the population. 10 % was decided based on selecting a fraction of the target population, although it might be misleading (Fowler, 2009).

The study employed cluster sampling of the schools, which means that sampling was conducted involving more than one stage. In a cluster sampling design, the researcher ascertains the organisation or groups (Creswell, 2014. p. 158). In the case of this study, a list of schools was obtained from the Ministry of Education. From this list, schools having elementary classes (PP to 6) were identified, resulting in 577 primary, lower secondary, middle secondary, or central schools being identified. (These categories of school, all have elementary grades in Bhutan, it is not only primary schools that have elementary grades.) From the 577 identified schools, 25 sample schools were selected through nonprobability (convenience) sampling. In non-probability/convenience sampling, the sample is chosen based on availability and convenience for respondents (Creswell, 2014, p. 158). However, in the current study convenience sampling was conducted not for respondents but the sample schools. Convenience sampling was chosen based on the availability of local contact persons who are probably known to the researcher and who would represent the researcher in the sample school and the availability of good internet facilities in the area where the sample school is located. The area of the schools sampled was considered because, during the data collection period, schools were closed due to the COVID-19 pandemic, so teachers were located at their homes near schools instead of being based at school campuses.

Selected schools' official email addresses were noted from the lists of schools. The principals were then contacted through these official email addresses, to request that they distribute the survey questionnaire with potential respondents in their respective schools by sharing the web link. Besides, the respondents were also invited informally through the Facebook page. Therefore, respondents for this study were found through convenience sampling (through sample schools and the Facebook page). Although convenience sampling is a less desirable sampling method (Creswell, 2014, p. 158), it is known for its low cost, lower time commitment, and ease of survey administration, which encourages greater participation and may generalise the result to parallel subjects (McMillan & Schumacher, 2006).

3. 8. INSTRUMENTATION

The online web-based survey consisted of four sections: Demographic information, Conceptions/beliefs about assessment, Self-reported assessment practices, and Classroom assessment literacy. These sections were adapted from previous studies.

.

3.8.1.Section One: Demographic information

This section of the survey sought to obtain participants':

- gender,
- years of teaching experience (0-15 years and more than 15 years),
- highest teacher qualification (PTC, BEd, PGDE, Masters, Ph.D.)
- grades taught (PP-6)
- subjects taught (Dzongkha, English, Mathematics, Science, Social studies)
- assessment education/training received (None, workshops provided by the school, the workshop provided by the MOE or REC, completed assessment course in undergraduate, graduate or postgraduate course)

3.8.2. Section Two: Conceptions of Assessment III (COA-III)

This section employed an abridged version of a survey with 27 items initially developed by Brown (2006) and later used by Calveric (2010) and Rosas (2014), to explore participants' beliefs about assessment. Items were scored on a scale from 1 to 5 (strongly disagree to strongly agree). Calveric (2010) and Rosas (2014) sought permission from the original author G.T.L. Brown (2006) to use the Conception of Assessment Abridged version (COA-III A) in their research projects with a slight modification on the Likert Scale. The current study borrowed the COA-III A from Calveric (2010) and Rosas (2014).

The abridged version of COA III (Brown, 2006, 2008) was developed and tested in New Zealand and Queensland with both primary and secondary teachers. Brown (2006), investigated whether an abridged version of COA III (50 items) would measure the same conceptual framework sufficiently. During the investigation, the data fit indices derived from confirmatory factor analysis demonstrated a good fit model to the data (x^2 =841.02; RMSEA=.057; TLI=.87). Brown's (2006) abridged model, therefore, showed similar interface correlation values and directions as the original version COA III. Next, the abridged version of COA-III was applied to teachers in Queensland and New Zealand. Independent Confirmatory Factor Analysis of these two jurisdictions, especially with the sample of primary teachers (n=692), was reported to have an acceptable fit (x^2 =1492.61; *p<.001;* RMSEA=.074; TLI=.80) with enough loadings of items on their respective factors and in the same direction (Brown, 2006, p. 169). Thus, Brown (2006, p.170) claims that the Confirmatory Factor Analysis of populations in different jurisdictions (Queensland and New Zealand) validated the abridged version of COA-III with 27 items, which can be used to study teachers' conceptions of assessment. Hence, the COA-III abridged version (COA-III A) was employed in the current research to measure the assessment beliefs of Bhutanese elementary teachers.

3.8.3. Section Three: Self-reported assessment practice

This section had 11 items that were originally borrowed and adapted by Calveric (2010) and Rosas (2014) from McMillan et al. (2002). Further, four more items were added by the researcher, making 15 items on the scale. The additional items were related to mechanisms

of formative assessment such as feedback, self-assessment, peer assessment, and questioning. These elements were added because this research project intended to understand the implementation of formative assessment by investigating the assessment beliefs, practices, and literacy of the teachers. The items for this research were scored on a scale from 1 to 5 ('not at all important' to extremely important'), unlike the scoring scale ('not important' to 'very important') used by Calveric (2010) and Rosas (2014).

McMillian et al (2002) included 47 items in the original version of their self-reported assessment practice instrument. After two subsequent pilot studies coupled with item review, McMillan reduced the items to 27. After a third survey and review of this scale, the author proposed an assessment practices measuring tool with 34 items that measured three constructs; namely, grading factors (19 items), types of assessment (11 items), and cognitive level of assessment (4 items). Given the way that the instrument was piloted and reviewed, the scale measuring assessment practices (11 items) was adopted for the current research project.

3.8.4. Section Four: Classroom Assessment Literacy Inventory (CALI)

This section consisted of items adapted by Mertler (2003) from the Teacher Assessment Literacy Questionnaire (TALQ) (Plake, Impara & Fager, 1993). TALQ explores levels of assessment literacy in relation to the Seven Standards for Teacher Competence in the Educational Assessment of Student (AFT, NCME, & NEA, 1990). Adapted from TALQ, the Classroom Assessment Literacy Inventory (CALI) by Mertler (2003) contains 35 items and is free and available online for researchers (Ryan, 2018, p. 54). For this research, CALI was modified and contextualised for Bhutanese teachers to include 21 multiple-choice questions. The resulting survey, therefore, had three questions per standard. The 21 multiple-choice questions were applied in contexts that required respondents to respond to factual knowledge items or read short classroom scenarios.

CALI is known for its continuous development, easy administration, less- time-consuming application, and sound reliability and validity (Ryan, 2018, p. 54). Mertler (2003) employed CALI with 197 (n=197) in-service teachers and 67 pre-service (n=67) teachers. The author observed an internal consistency reliability estimate of .57 for in-service teachers and .74 for pre-service teachers. The original instrument (TALQ) (Plake et al., 1993) on which Mertler's

CALI was based has therefore undergone robust validation and review, thereby ensuring validation of CALI in measuring levels of assessment literacy.

Table 5 summarises the items in the survey questionnaire. Meanwhile, the adapted version of the instrument is attached in Appendix 4.

Sections	Name	Purpose	No. of items	Scale	Adapted sources
One	Demographic Information	to test their significant differences on dependent variables	6		Calveric (2010); Rosas (2014)
Two	The conception of Assessment (COA-III) or Assessment beliefs	to allow the respondents to indicate the level of agreement or disagreement with the statements that measure the assessment beliefs	27	1-Strongly Disagree 2-Disagree 3-Neutral 4-Agree 5-Strongly Agree	Brown (2006) Calveric (2010) Rosas (2014)
Three	Self-Reported Assessment Practice	to allow teachers to rate the value against the assessment types assessment design	15	1-Not at all important 2-Low important 3-Moderately important 4-Very important 5-Extremely important	McMillan et al (2002) Calveric (2010); Rosas (2014
Four	Classroom Assessment Literacy Inventory (CALI)	to measure the level of assessment literacy.	21	MCQ with one correct response and 3 incorrect responses (Cognitive test) 0-incorrect response 1-correct response	Mertler (2003) Ryan (2018)

TABLE 5 SUMMARY OF THE SURVEY QUESTIONNAIRES

Although each section of the survey questionnaire had been validated by both the original authors and other researchers in their respective contexts, this study also established the instruments' validity and reliability. According to Creswell (2014, p. 160), it is necessary to re-establish the validity and reliability of a modified and adapted questionnaire for use in new studies. As such, this study conducted a confirmatory factor analysis to measure the construct validity of the items that measured the constructs of assessment beliefs, assessment practices, and assessment literacy, similar to Brown (2006, 2008, 2011). The results of the confirmatory factor analysis are presented in Chapter 4, the results section. The explanation of CFA and its characteristics are presented in Appendix 9. To test the item reliability, Cronbach's Alpha was utilised.

3.9. VARIABLES IN THE STUDY

According to Creswell (2014, p. 161), relating variables to research questions and items allows readers to view the relationship of data collection with a study's research questions or hypothesis. The independent variables of this study include the demographic information of respondents which include gender, years of teaching experience, level of teacher education, grade taught, subjects taught, and assessment education/training received. The dependent variables are assessment beliefs, assessment practices, and assessment literacy. Table 6 portrays the connection of variables, research questions, and items.

Variables	Research Questions	Items on Survey
Independent variables: **Demographic**	Descriptive item Check the appropriate box	Section 1 (Appendix 4) gender: teaching experience: level of teacher education: Subject taught: Grade taught: Assessment education/training:
Dependent variable 1 **Assessment beliefs**	Descriptive research question:1 What conceptions (beliefs) of assessment do Bhutanese Elementary teachers have?	Section 2 (Appendix 4) See questions 1- 27
Dependent variable 2 **Assessment practices**	Descriptive research question:2 What assessment practices do Bhutanese teachers value?	Section 3 (Appendix 4) See questions 1-15
Dependent variable 3 **Assessment literacy**	Descriptive research question 3 What is the level of assessment literacy of elementary teachers in Bhutan measured by the Classroom Assessment Literacy Inventory?	Section 4 (Appendix 4) See questions 1- 21
Relating independent variables to 3 dependent variables	Inferential question 4 Do the independent variables of gender, teacher education level, years of teaching experience, grade taught, subject taught and assessment education/training received to make a significant difference in teachers' assessment beliefs, assessment practice, and assessment literacy level?	section 1 related to section 2, 3,and 4 respectively
Mediating variables with independent and dependent variables	Inferential question 5How is the dependent plus endogenous variable of teachers' assessment practice affected by the independent variables of gender, years of teaching experience, level of teacher education, grade taught, subject taught and assessment education/training received and by assessment beliefs and assessment literacy?	Section 1 to 4

TABLE 6. SUMMARY OF VARIABLES IN THE STUDY

3.10. ETHICS

The study sought and received approval from the Department of School Education under the Ministry of Education in Bhutan, with approval number DSE/SPCD/SLCU(2.2)/2020/471 (Appendix 2). Subsequently, the proposal for this study was reviewed by the Human Research Ethics Committee (HREC) of the University of Adelaide. The proposal was approved and provided with an approval number H-2020-121 (Appendix 1). After the approval was granted by the HREC, the principals of 25 sample schools were sent emails seeking written approval and distribution of the online survey to the potential respondents in their schools (Appendix 5). The respondents were provided with Participation Information Sheets to allow them to make informed choices about their participation in the survey (Appendix 3).

3.11. DATA COLLECTION PROCESS

At the initial stage of the research project, all of the principals of 25 sample schools were sent an email (Appendix 8) seeking permission and asking them to distribute the survey to potential respondents in their respective schools. The email attached an introductory message (Appendix 6) for respondents and a Participation Information Sheet (Appendix 3) with the survey link included. Ten principals responded to the email, wherein four principals sent a written approval and others acknowledged the email stating that the survey was received and distributed to the potential participants.

There were two platforms created to collect data from the respondents: a weblink distributed to school principals (formal), and a Facebook page (informal). The Facebook page was created to increase the response rate, in light of the slow and low turnover of initial respondents due to the closure of the schools in the COVID-19 pandemic. The responses were automatically stored in Survey Monkey in the researcher's account. Eventually, the number of respondents increased. Finally, after a month when the survey was closed both formally and informally, there were 59 responses submitted through the web link and 53 responses submitted through the Facebook page.

3.12. DATA ANALYSIS AND INTERPRETATION

3.12.1 Overview

Data Were collected from 112 respondents from the targeted number of approximately 200 teachers, in the sample within the population of 2,296 elementary teachers in Bhutan. The response rate for the research project was therefore 56%, with responses from all 112 respondents included in the data for analysis. The missing responses were compensated for by the software used to analyse the results as described below.

3.12.2 Data preparation

After the closure of the link for the survey, the data in SurveyMonkey were extracted to a separate file in Microsoft Excel by filtering the data according to the survey questionnaire. The responses in the Excel were converted to numerical values, after which the data were transferred to the *Statistical Package for Social Science* (SPSS) software (v.26) (IBM Corp. 2019). In SPSS, the data were prepared for analysis by assigning Codes for the demographic information and responses to a scale that measured the constructs of assessment beliefs, practices, and assessment literacy.

3.12.3 Data analysis techniques

To answer the study's five research questions, the data were analysed using the quantitative data analysing software *Statistical Package for Social Sciences (SPSS version 26)* and *Analysis of Moment Structure (AMOS; AMOS graphic version 23) (Arbuckle, 2014)*. According to Abbott (2011, p.23), SPSS is the "most versatile and responsive program" for numerous statistical procedures. SPSS features a large spreadsheet that permits users to enter, manipulate, and analyse data in a wide range of variations (Abbott, 2011, p. 24). AMOS is an IBM SPSS segment designed for the analysis of covariance structure models such as structural equation modeling (SEM), path analysis, and confirmatory factor analysis (CFA) (Barnidge & De Zuniga, 2017, p.1). The authors claim that one of the advantages of AMOS is that it presents a user-friendly graphical interface, thereby enabling nonprogrammers to construct visual models.

Descriptive data statistical techniques were used to analyse the results for the descriptive data. This descriptive analysis computed the data's frequency, percentage, mean (the measure of central tendency and the average value/score), and standard deviation (the measure of dispersion/variability) using SPSS v. 26.0 (SPSS Inc., 2017). Before, analysing the results to respond to the study's research questions, the demographic information data were

run through descriptive analysis to obtain an overview of the respondents' characteristics. Further, before analysing the data to answer the research questions, structural level analysis, (CFA using AMOS) was conducted to evaluate the construct validity of the data with assessment beliefs and assessment practices data. Also, scale reliability was tested for internal consistency by calculating Cronbach's alpha. CFA and Cronbach's alpha analysis are explicitly presented in the results section (Chapter 4).

Following the CFA, the study's first three research questions were answered by analysing the corresponding data using descriptive analysis statistical techniques in SPSS. Table 7 presents a summary of the data analysis techniques used. The frequency, composite mean scores, standard deviations, and percentages of the responses were compared to determine the assessment beliefs, assessment practices, and the levels of assessment literacy demonstrated by the respondents. To answer research question 4, inferential analysis involving a t-test of the independent sample, and analysis of variance (ANOVA) using SPSS, were conducted to test the significant difference that the six independent variables from the demographic information had on the three dependent variables. The t-test of the independent sample was conducted in particular to determine the influence with significant differences that gender had on the three dependent variables. A series of One-Way ANOVA tests were run to identify any significant differences of the remaining independent variables that had more than two levels, on the three dependent variables. Structural equation modeling (SEM) in AMOS was used to examine the possible relationships among variables at a particular level, which answered research question 5.

Research Questions	Statistics	Data Analysis	Software
What conceptions (beliefs) of assessment do Bhutanese Elementary teachers have?	Descriptive	Frequencies, Means, Standard Deviation and Percentage	SPSS v 26
What assessment practices do Bhutanese teachers value?	Descriptive	Frequencies, Means, Standard Deviation and Percentage	SPSS v 26
What is the level of assessment literacy of elementary teachers in Bhutan measured by the Classroom Assessment Literacy Inventory?	Descriptive	Frequencies, Means, Standard Deviation, Minimum score, Maximum score, and percentages	SPSS v 26
Do the independent variables of gender, teacher education level, years of teaching experience, grade taught, subject taught and assessment education/training received have significant differences in teachers' assessment beliefs, assessment practice, and assessment literacy level?	Inferential	Analysis of variance (ANOVA), Independent sample *t*-test, and Post hoc analysis as required.	SPSS v 26
How is the dependent plus endogenous variable of teachers' assessment practice affected by the independent variables of gender, years of teaching experience, level of teacher education, grade taught, subject taught and assessment education/training received and by the exogenous variables of assessment beliefs and assessment literacy?	Inferential	Structural Equation Modelling (SEM), direct, indirect, and total effect, model fit indices.	AMOS v 23

TABLE 7 SUMMARY OF DATA ANALYSIS TECHNIQUES

CHAPTER FOUR: RESULTS

4.1. INTRODUCTION

This chapter presents the findings of the data collected from Bhutanese elementary teachers about their beliefs about assessment, their assessment practices, and level of assessment literacy. The results begin with teachers' demographic information followed by the construct validity of scales in the instrument and presenting the results for each research question in sequence.

4.2. DESCRIPTIVE RESULT OF DEMOGRAPHIC DATA

The following results represent the demographic information provided by 112 respondents. The demographic features included gender, highest teacher qualification, years of experience in teaching, grade taught, subject taught, and the types of education or training received on assessment education. The results are presented in the sequence of these variables.

4.2.1. Gender

The data was collected from teachers teaching any subject in PP through Grade 6 working in government schools. A total of 112 teachers took the survey. However, one respondent did not provide gender information. As shown in Table 8, from a total of 111 (n=111) who reflected their gender, 72.1% (F=80) were female and 27.9% (M=31) were male. The result showed that there were more female respondents than male respondents.

Gender

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid					
	Female	80	71.4	72.1	72.1
	Male	31	27.7	27.9	100.0
	Total	111	99.1	100.0	
Missing	System	1	.9		
Total		112	100.0		

TABLE 8 PARTICIPANTS BY GENDER IN PERCENT

4.2.2. Highest Teacher Qualification

The teacher qualification was classified into five categories of PTC, BEd, PGDE, master's degree, and Ph D. A total of 111 (n=111) had mentioned their teacher qualification of which 18.9% (21) had PTC qualification, 59.5% (66) had B Ed qualification, 8.1% (9) had PGDE, 13.5% (15) had a master's degree and none had Ph.D. qualification. It indicates that the majority of the respondents had the teacher qualification of bachelor's degree in education. Figure 4 illustrated the levels of teacher qualification.

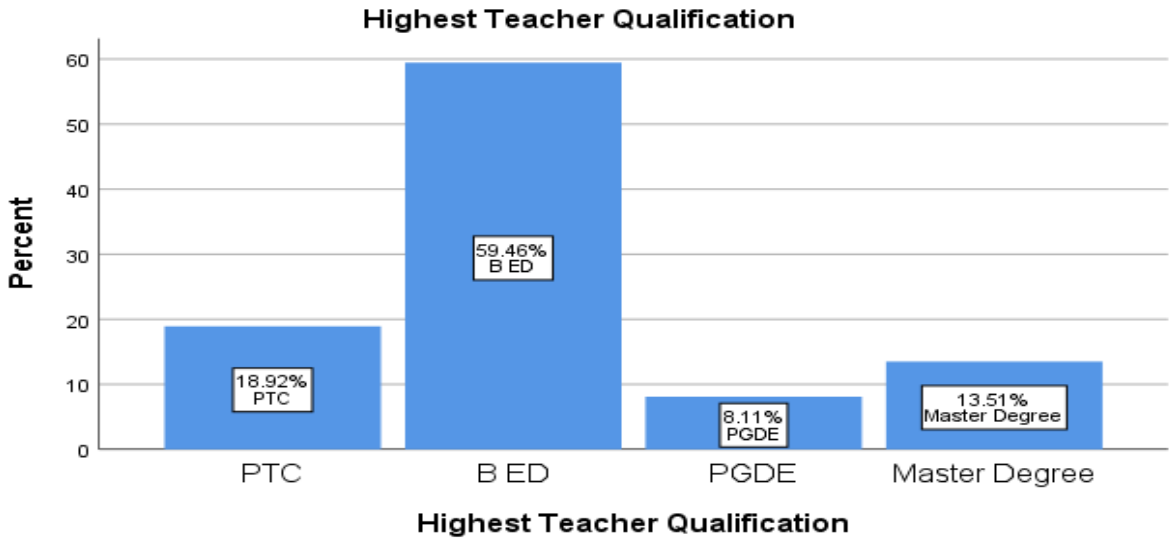

FIGURE 4 THE LEVELS OF TEACHER EDUCATION OF ELEMENTARY TEACHERS

4.2.3. Number of years in teaching/teaching experience

The number of years in teaching was categorised as 0-5 years, 6-10 years, 11-15 years, and more than 15 years. Figure 5 highlighted the level of teacher qualification of 111 (n=111) respondents who provided demographic information. As shown, 27.0% (30) respondents had teaching experience between 0-5 years, 6.3% (7) respondents had teaching experience between 6-10 years, 27.0% (30) respondents had teaching experience between 11-15 years and 39.6% (44) respondents had teaching experience of above 15 years. This result indicated that more senior teachers were teaching in elementary classes than the recent teachers.

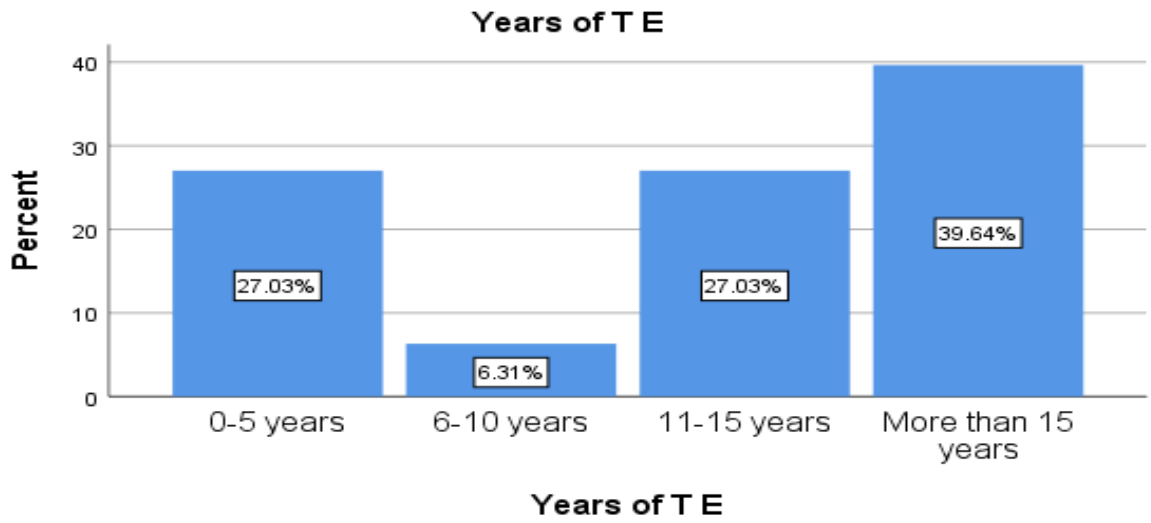

FIGURE 5 THE YEARS OF TEACHING EXPERIENCE OF ELEMENTARY TEACHERS

4.2.4. Grade taught

The demographic information sought information on grade taught by the respondents. The elementary grades included Pre-Primary, Grade 1, Grade 2, Grade 3, Grade 4, Grade 5, and Grade 6. As displayed in Figure 6, 14.7% (16) respondents taught in Pre-Primary, 12.8% (14) respondents taught in Grade 1, 10.1% (11) respondents taught in Grade 2, 11.9%(13) respondents taught in Grade 3, 11.0% (12) respondents taught in Grade 4, 13.8% (15) respondents taught in Grade 5 and the highest percent of 25.7%(28) respondents taught in Grade 6.

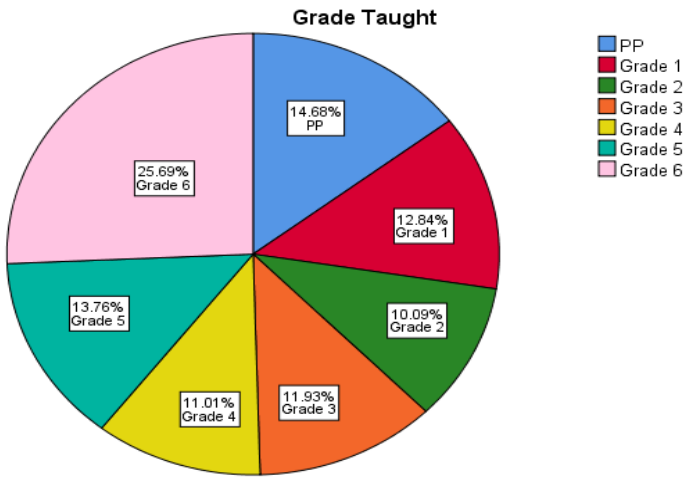

FIGURE 6. PIE CHART COMPARING PARTICIPANTS BY THE GRADE TAUGHT

4.2.5. Subject taught

The subjects included Dzongkha (Bhutanese national language), English, Mathematics, Science, and Social studies. Figure 7 displayed the result of the subject taught by the respondents. 10.8% (12) of the respondents taught Dzongkha, 46.8% (52) taught English, 25.2% (28) taught Mathematics, 9.9% (11) taught Science and 7.2% (8) taught Social Studies. The results portrayed that more respondents were teaching English during the survey.

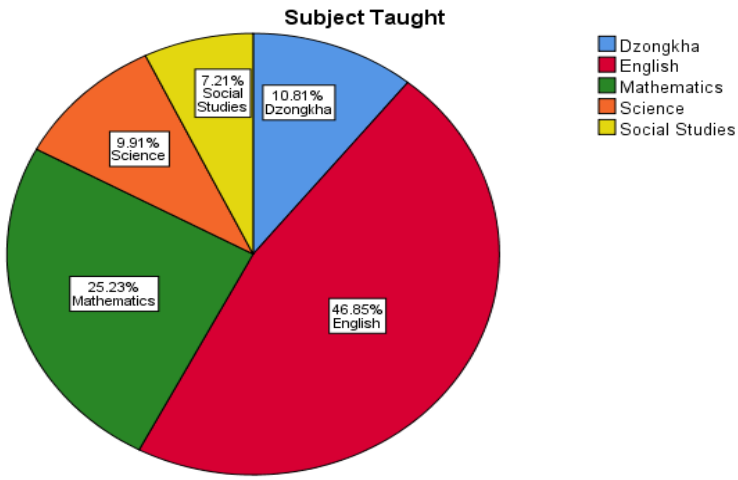

FIGURE 7 PIE CHART COMPARING TEACHERS BY THE SUBJECT THEY TAUGHT

4.2.6. Assessment education/training

Figure 8 represents the teachers categorised by assessment education/training received. A total of 111 respondents 1.8% (2) did not receive any assessment training or education. 21.6% (24) of them got their assessment education and training through school professional development programs. 37.8% (42) received their assessment training or education through national professional development programs conducted either by the MoE or REC. Likewise, 1.8% (2) got such education during their undergraduate assessment course, 25.2% (28) got in their graduate course, 11.7% (13) got from their post-graduate assessment course. It suggests that the majority of the respondents had received assessment education or training through professional development programs and during their teacher education courses.

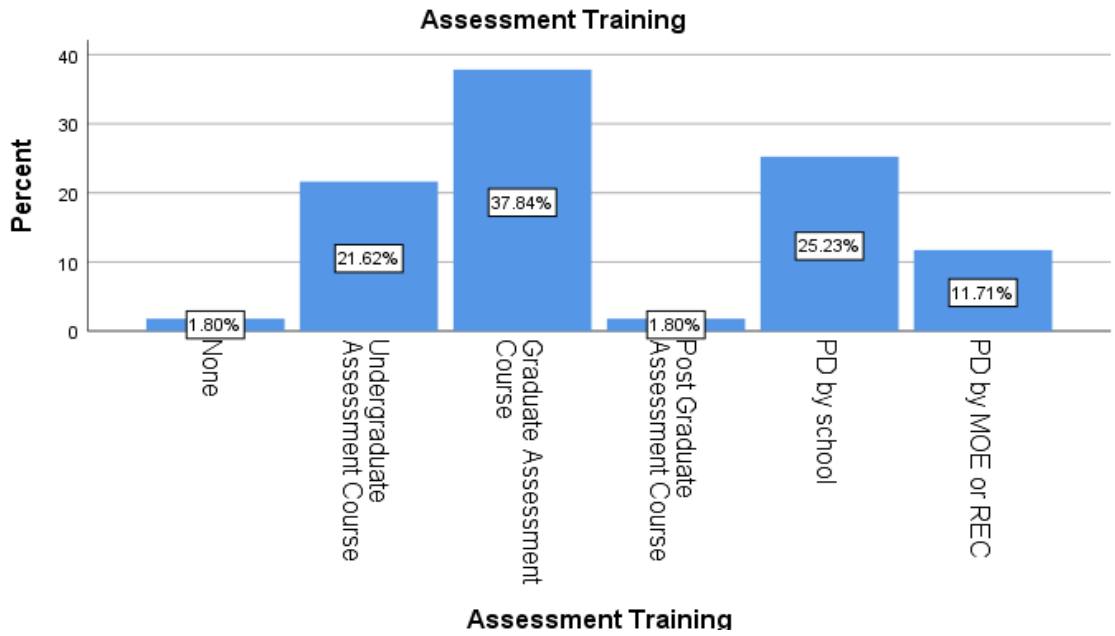

FIGURE 8 INTENSITY OF ASSESSMENT EDUCATION/TRAINING RECEIVED BY TEACHERS

4.3 CONSTRUCT VALIDITY OF CONCEPTION OF ASSESSMENT (COA-III A)

To test the construct validity of items measuring the assessment beliefs conceived by Bhutanese elementary teachers, the confirmatory factor analyses (CFA) were employed.

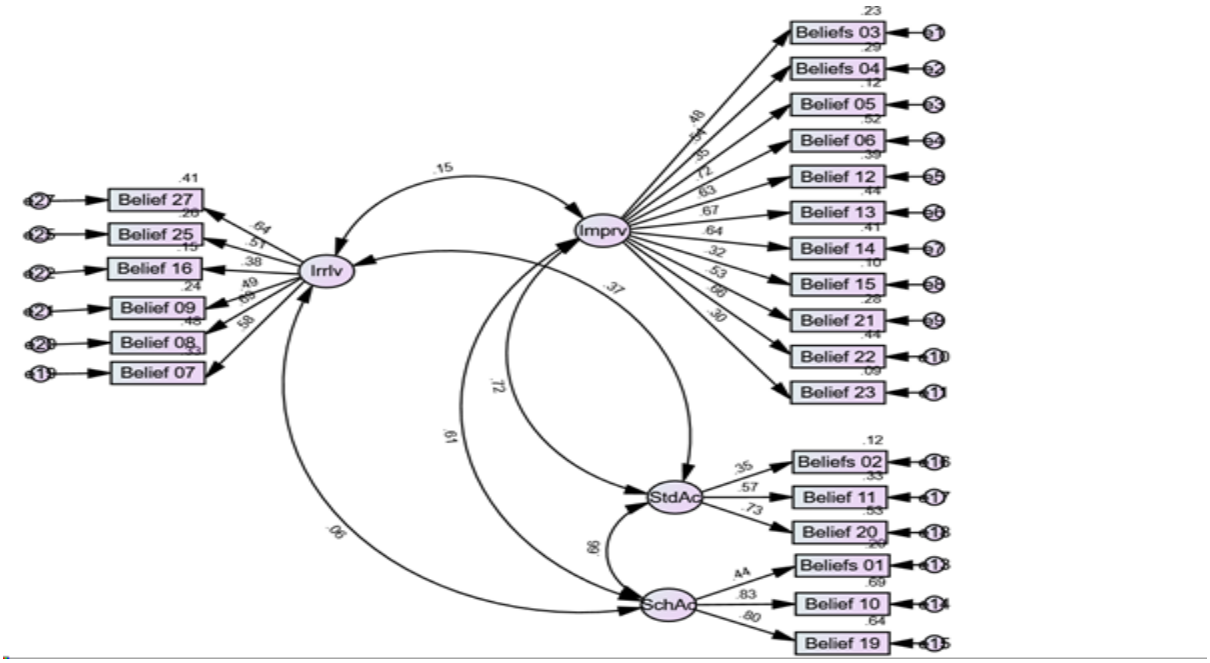

FIGURE 9 CORRELATED FOUR FACTOR MODEL (STANDARDISED)

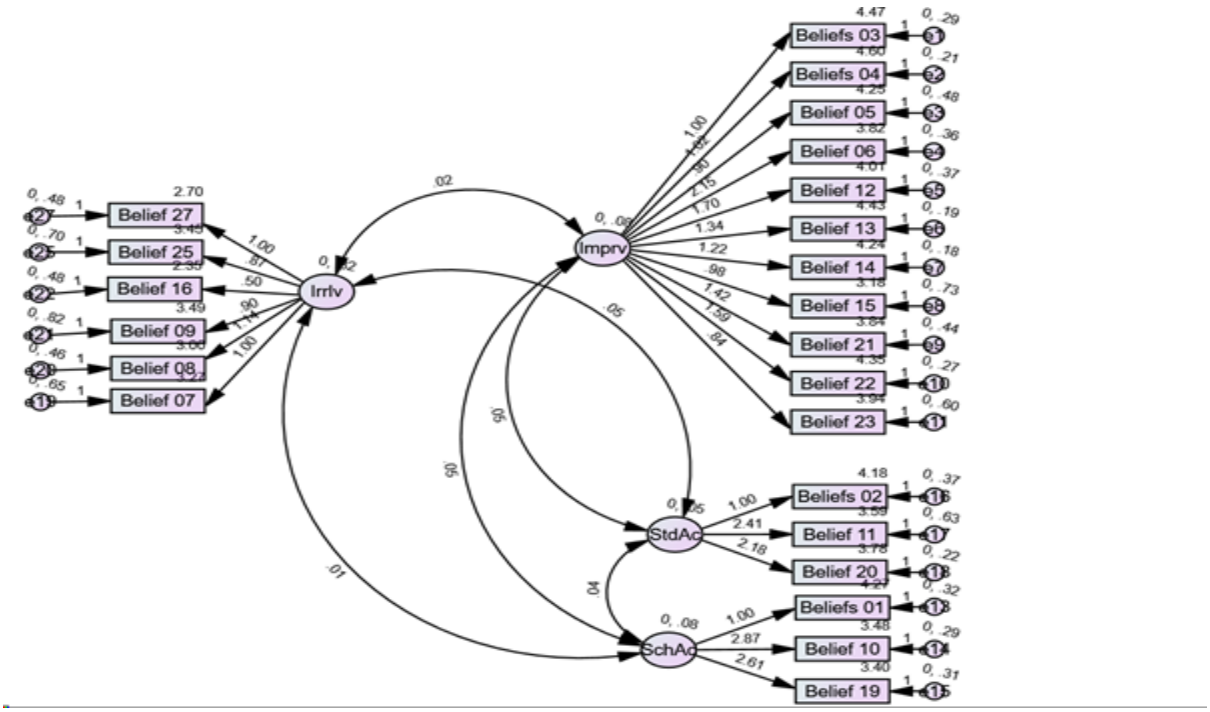

FIGURE 10 CORRELATED FOUR FACTOR MODEL (UNSTANDARDISED)

4.3.1 CONFIRMATORY FACTOR ANALYSIS

Figures 9 and 10 present the results of the path analysis of the Correlated Four Factor model with standardised and unstandardised estimates and R^2 values. The model was selected out of four alternative CFA models. The four unobserved variables in the models were assessment

for improvement (Imprv), Assessment for student accountability (StdAc), assessment for school accountability (SchAc), and assessment is irrelevant (Irrlv) with 23 items (initially 27 items) measuring assessment beliefs. The unobserved variable of Improvement (Imprv) had 11 items loaded on it, Student Accountability (StdAc) had 3 items, School Accountability (SchAc) had 3 items, and Irrelevance (Irrlv) had 6 items loaded. Further, each unobserved variable was correlated with each other.

4.3.2 Evaluating the factor loadings and squared correlation

The assessment beliefs (CoA) items were examined using the factor loadings to gauge whether or not the items reflected the factors that they were expected to represent. The scoring of the COA-III abridged version was a Likert scale of 0-5, from strongly disagree-0, disagree-1, neutral-3, agree-4, and strongly agree-5. As a threshold, a factor loading of 0.30 (Kline, 1994) was chosen. Table 15 showed the factor loadings for both models, standardised and unstandardised.

During the CFA run, items with the factor loading less than 0.3 (Kline, 1994) were eliminated one by one in each run to get a more parsimonious result/model [parsimonious – simple model yet powerful in explaining the data (Matzke, 2014, p.127)]. From 27 items 4 items were eliminated; item coded as Belief24 was removed from the construct "assessment for improvement" (Impv), and items coded as Belief17, Belief18 and Belief26 were removed from the construct "assessment is irrelevant" (irrlv), therefore resulted into 23 items. As noted in Table 9, rest of the items according to four conceptions/beliefs of assessment exhibited factor loadings above the threshold value as follows: Improvement conception-11 items (Belief03, Belief04, Belief05, Belief06, Belief12, Belief13, Belief14, Belief15, Belief21, Belief22, Belief23), Student Accountability Assessment conception-three items (Belief02, Belief11, Belief20), School Accountability Assessment Conception-three items (Belief01, Belief10, Belief19) and Assessment is irrelevant conception-6 items (Belief07, Belief08, Belief09, Belief16, Belief25, Belief27). Similarly, the R^2 ranged from 0.09 to 0.69. The item Belief23 had the least R^2 value of .09 which indicated that less than 10% of this variance is accounted to subjective norms. However, all other items exhibited more than 10% of the variances that are accounted to subjective norms.

Unobserved Variables	Observed variables	Standardised		Unstandardised	
		Factor Loadings	R^2	Factor Loadings	R^2
Improvement	Belief03	.465	.23	1.00	4.47
(Imprv)	Belief04	.538	.29	1.01	4.60
	Belief05	.349	.12	.90	4.25
	Belief06	.721	.52	2.14	3.82
	Belief12	.628	.39	1.70	4.01
	Belief13	.667	.44	1.34	4.43
	Belief14	.638	.41	1.21	4.24
	Belief15	.316	.10	.98	3.18
	Belief21	.530	.28	1.42	3.84
	Belief22	.663	.44	1.59	4.35
	Belief23	.300	.09	.83	3.94
Student Accountability	Belief02	.353	.12	1.00	4.18
(StdAc)	Belief11	.573	.33	2.41	3.59
	Belief20	.729	.53	2.17	3.78
School Accountability	Belief01	.444	.20	1.00	4.27
(SchAc)	Belief10	.833	.69	2.86	3.48
	Belief19	.799	.64	2.60	3.40
Irrelevance	Belief07	.576	.33	.99	3.27
(Irrlv)	Belief08	.692	.48	1.13	3.00
	Belief09	.493	.24	.89	3.49
	Belief16	.384	.15	.50	2.35
	Belief25	.507	.20	.18	3.45
	Belief27	.637	.41	1.00	2.70

TABLE 9 FACTOR LOADINGS AND SQUARED COEFFICIENT

4.3.3 Assessment beliefs model fit

In addition to factor loadings, model fit indices were examined to test the data fit. The model fit indices were compared among four CFA alternative models of One Factor Model, Orthogonal Four Factor Model, Correlated Four factor Model, and Hierarchical model. As shown in Table 10 the indices x^2, df, CMN/df, TLI, CFI, and RMSEA (Schreiber, Stage, Barlow &

King, 2006) improved significantly with the subsequent models from the one-factor model through the hierarchical model. Since the Correlated Model had the best indices comparatively, it was accepted for further analysis. The fit indices highlighted that the chi-square was (x^2= 347.475). Smaller the x^2 value, the better the model fit. However, x^2 and x^2 /df are sensitive to sample size (Hox & Bechger, n.d) The ratio of chi-square to degrees of freedom (x^2/df= 1.551) which may be considered acceptable as the value is almost 2 (Tabachnick & Fidell, 2007). The TLI and CFI are considered significant when they are closed to 1 such as 0.9 or 0.95 (Bentler, 2007; Matsunaga, 2010). The Comparative Fit Index (CFI) was 0.776. The Tucker-Lewis Index (TLI) was found 0.724 and they demonstrate an acceptable fit of the data to the chosen model. The RMSEA is known for excluding the influence of sample size as well as perform statistical tests on the values and therefore, considered one of the primary indicators for evaluating the goodness of fit of a model. Generally, RMSEA is acceptable with a 0.08 value, but less than 0.05 is better, and less than 0.01 is a perfect model (Kline, 2016). The (RMSEA) was found 0.050, therefore, indicated a good fit. Based on the factor loadings equal to or more than 0.3 and good fit indices, the set of data on conception/beliefs of assessment were taken for further result analysis that answered the research questions.

N0	Model	Chi-Square	df	CMN/df	TLI	CFI	RMSEA
1	1-factor model	600.067	324	1.852	.479	.553	.062
2	Orthogonal model	573.237	324	1.769	.529	.597	.059
3	*Correlated model*	*347.475*	*224*	*1.551*	*.724*	*.776*	*.050*
4	Hierarchical model	351.277	226	1.554	.723	.773	.050

TABLE 10 GOODNESS-OF- MODEL FIT INDICES

4.3.4 Correlation among the Four Constructs of Assessment Beliefs

The correlated four-factor model presented that the four constructs of assessment beliefs were correlated. The correlations between the four constructs of assessment beliefs were presented in Table 11. The results showed that the improvement construct and student accountability construct had a high positive correlation (r=.71; p=0.01). A positive moderate

correlation was noted between student accountability and school accountability (r=66; p=0.01). Likewise, assessment for improvement and school accountability also had a moderate positive correlation (r=.61; p=0.01). Interestingly, "assessment is irrelevant" was weakly but positively correlated to improvement (r=.15; p=0.05). Similarly, a positive association existed between assessment is irrelevant and student accountability (r=.37) and assessment is irrelevant, and assessment for school accountability (r=.06).

			Estimate
Irrelevance	<-->	Improvement	.147
Irrelevance	<-->	Student Accountability	.368
Irrelevance	<-->	School Accountability	.057
Improvement	<-->	School Accountability	.606
School Accountability	<-->	Student Accountability	.656
Improvement	<-->	Student Accountability	.716

TABLE 11 CORRELATION BETWEEN FOUR CONSTRUCTS OF ASSESSMENT

4.4. CRONBACH ALPHA; TESTING THE RELIABILITY

Cronbach's alpha coefficient and the number of items for four constructs had been determined in Table 12. Generally, the cut off value for the reliability test is Cronbach's alpha 0.7. Analysis of Cronbach's alpha coefficient produced acceptable value for two constructs; Assessment for improvement and assessment for school accountability (.791 and .716 respectively). However, for Assessment for student accountability and Assessment is Irrelevant produced alpha coefficient below 0.7 (.556 and .666). Assessment for improvement construct earned the highest alpha score of 0.784, whereas Assessment for student accountability showed the lowest alpha value of 0.556 yet acceptable for further analysis.

Constructs	Alpha Coefficient	No. of items
Assessment for improvement	.791	11
Assessment for student accountability	.556	3
Assessment for school accountability	.716	3
Assessment is irrelevant	.666	6

TABLE 12 CRONBACH ALPHA OF EACH CONSTRUCT

4.5 QUESTION 1.

What conceptions (beliefs) of assessment do Bhutanese elementary teachers have?

In answering the first research question, "What conceptions (beliefs) of assessment do Bhutanese elementary teachers have? a descriptive statistical analysis was made to obtain frequency (n), Mean (M), and Standard Deviation (SD). Based on the CFA correlated four-factor model (Figures 9&10) and findings in the existing literature on beliefs of assessment, in particular, that of Brown (2002, 2004, 2006, 2008) and Cleviar (2010), the 23 items scale was categorised into four assessment beliefs; assessment for improvement, assessment for student accountability, assessment for school accountability and assessment is irrelevant. Therefore, these four constructs were employed to determine the assessment beliefs held by Bhutanese elementary teachers.

Table 13, highlights the frequency (n), Means (M), and Standard Deviation (SD) of the four constructs measuring assessment beliefs. The mean score ranged from (M=4.13) the highest to (M=3.03) the lowest. The mean score above 4 denoted that the respondents agreed with a particular belief and a mean score above 3 denotes neutral with a particular belief and below 2 denotes disagreement. The construct assessment for improvement scored the highest mean (M=4.13) with (SD=40) suggesting that Bhutanese elementary teachers agree that assessment has improvement purposes. Meanwhile, the assessment for student accountability and assessment for school accountability scored relatively the same mean scores (M=3.85 with SD=50 and 3.72 with SD=.70 respectively). These Mean scores suggested that Bhutanese elementary teachers had strong neutral beliefs on assessment has the

purpose of making both student and school accountable. The lowest mean score of (M=3.03; SD=.72) achieved by the fourth construct of assessment is irrelevant suggested that Bhutanese elementary teachers had weak neutral belief on this construct.

Constructs	n	M	SD
Assessment for improvement	102	4.13	.40
Assessment for student accountability	102	3.85	.57
Assessment for school accountability	101	3.72	.70
Assessment is irrelevant	102	**3.03**	.72

TABLE 13 DESCRIPTIVE STATISTICS FOR ASSESSMENT BELIEFS

The error plot in Figure 11 further explained the assessment beliefs held by the Bhutanese elementary teachers. The mean for assessment for improvement is the highest against three other assessment beliefs. It indicated that teachers agree that assessment improved teaching and learning. However, they did not disagree that assessment is irrelevant as indicated by the neutral mean scored by this construct.

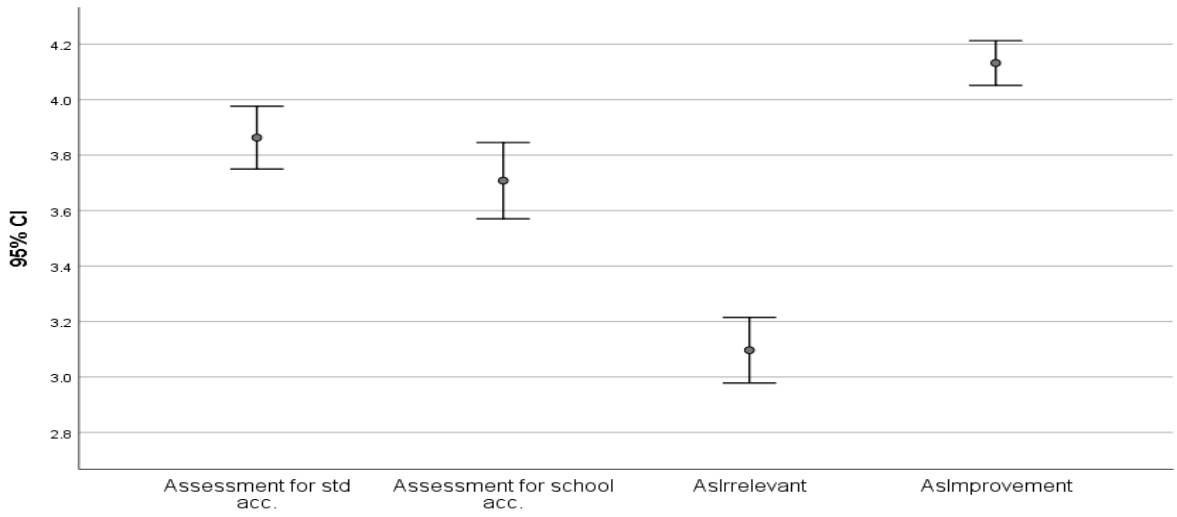

FIGURE 11 ERROR PLOT SHOWING MEAN SCORES FOR ASSESSMENT BELIEFS

4.6 ASSESSMENT PRACTICES

4.6.1 Construct Validity of Assessment Practices items

To test the construct validity of items measuring the values Bhutanese teachers placed on assessment practices including assessment design, a confirmatory factor analysis (CFA) was employed.

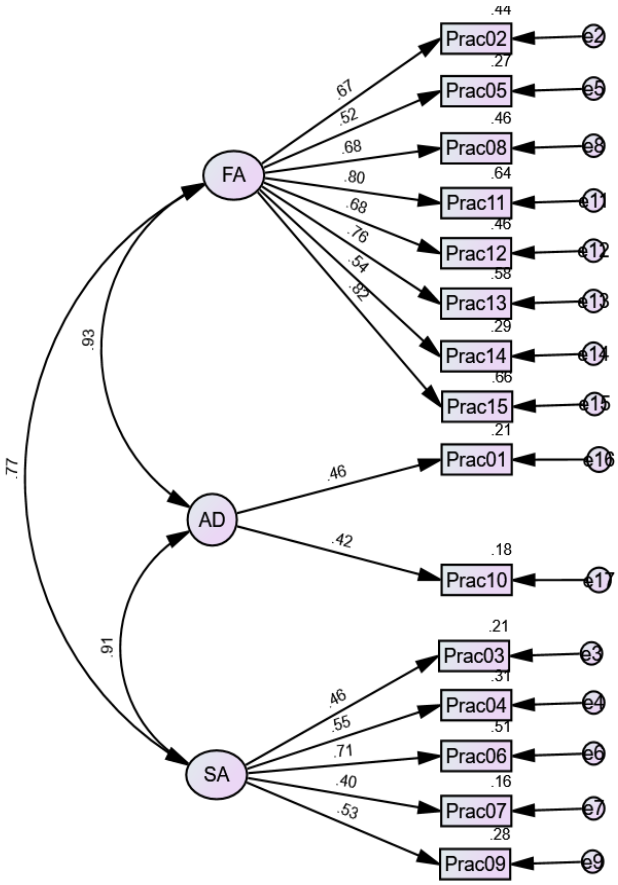

FIGURE 12 CORRELATED THREE-FACTOR MODEL (STANDARDISED)

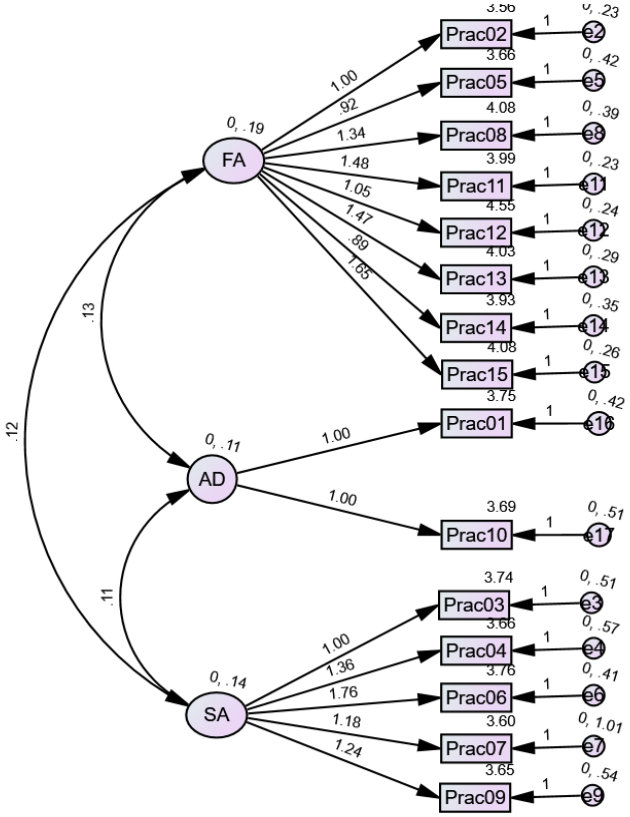

FIGURE 13 CORRELATED THREE-FACTOR MODEL (UNSTANDARDISED)

4.6.2 Confirmatory Factor Analysis

The two figures above presented the results of path analysis. The Correlated Three-Factor model was chosen among four alternative CFA models. Figures 12 and 13 show standardised and unstandardised estimates and R^2 values of the model. Three latent variables in the models were formative assessment (FA), summative assessment (SA), and assessment design (AD). The latent variable of formative assessment had 8 items loaded on it, 5 items loaded on summative assessment and 2 items loaded on assessment design. Further, the three latent variables were correlated.

4.6.3. Evaluating the factor loadings and squared correlation (R2)

The Assessment Practices items were examined using factor loadings to gauge whether or not the items reflected the factors that they were expected to represent. The scoring of the Assessment practices items was a Likert scale of 0-5, from Not at all important-0, Low important-1, Moderately important-3, Very important-4 to Extremely important -5. As a

threshold, a factor loading of 0.30 (Kline, 1994) was chosen. The factor loadings for both models, standardised and unstandardised are presented in Table 14.

As noted in the table, none of the items in the constructs of formative assessment, summative assessment, and assessment design loaded with estimates less than 0.3. It suggested that all items measured the intended constructs. Similarly, the R^2 ranged from 0.16 to 0.66. The items such as Prac07 and Prac10 had low scores for R^2 (.16 and .18), which indicated that only 16% and 18% of these variances were accounted for subjective norms.

Second-order Factor	**First Order Factors**	**Observed variables**	**Standardise**		**Unstandardised**	
Assessment practice			Factor Loadings	R^2	Factor Loadings	R^2
	Formative Assessment	Prac02	.666	.443	1.00	3.56
		Prac05	.523	.273	.916	3.66
		Prac08	.680	.462	1.343	4.08
		Prac11	.800	.640	1.479	3.99
		Prac12	.678	.460	1.052	4.55
		Prac13	.760	.578	1.466	4.03
		Prac14	.542	.294	.887	3.93
		Prac15	.815	.664	1.654	4.08
	Summative Assessment	Prac03	.459	.211	1.00	3.74
		Prac04	.553	.306	1.360	3.66
		Prac06	.711	.505	1.764	3.76
		Prac07	.397	.158	1.184	3.60
		Prac09	.528	.278	1.245	3.65
	Assessment Design	Prac01	.455	.207	1.00	3.75
		Prac10	.420	.176	1.00	3.69

TABLE 14 FACTOR LOADINGS AND SQUARED CORRELATION (R^2)

4.6.4 Assessment practice model fit

In addition to factor loadings, model fit indices were examined to test the data fit. The model fit indices were compared among four CFA alternative models of One Factor Model, Orthogonal Two Factor Model, Correlated Two-factor Model, and Hierarchical model. Table 15 showed the model fit indices (x^2, df, CMN/df, TLI, CFI, and RMSEA) in the correlated and hierarchical model were contending. However, considering the close association of formative assessment, summative assessment, and assessment design, the correlated model was found to be more relevant to the data as well as the literature. The fit indices highlight that the chi-square was (x^2=153.385). The ratio of chi-square to degrees of freedom (x^2/df= 1.743), smallest among other values, therefore, was considered acceptable. The Comparative Fit Index (CFI) was 0.812 which was acceptable but the Tucker-Lewis Index (TLI) was found 0.743 and demonstrated a poor but acceptable fit for the data. The (RMSEA) was found 0.058, therefore, indicated a good fit. Although the value of TLI demonstrates a poor fit of the data, given the RMSEA's significant value and that of CFI, the Correlated Factor Model was chosen. Considering the factor loadings of all items which were above 0.30 and the model fit, it was suggestive to take forward the set of data on assessment practice for further result analysis that answered research questions 2.

N0	Model	Chi-Square	df	CMN/df	TLI	CFI	RMSEA
1	1-factor model	167.864	90	1.865	.701	.776	.062
2	Orthogonal model	204.374	91	2.246	.569	.674	.075
3	*Correlated model*	*153.385*	*88*	*1.743*	*.743*	*.812*	*.058*
4	Hierarchical model	153.385	88	1.743	.743	.812	.058

TABLE 15 GOODNESS-OF-MODEL FIT INDICES

4.6.5 Correlation among the three latent variables

The correlated three-factor models (Figure 12 &13) and Table 16 depicted that formative assessment, summative assessment, and assessment design measured by the assessment practice items had a strong positive correlation. Assessment design had a strong positive

correlation with formative assessment (r=.93). It was also strongly correlated with summative assessment (r=.90). Formative assessment and summative assessment were highly correlated (r=.77).

Assessment Design	<-->	Formative Assessment	.931
Assessment Design	<-->	Summative Assessment	.905
Formative Assessment	<-->	Summative Assessment	.766

TABLE 16 CORRELATION AMONG THREE CONSTRUCTS OF ASSESSMENT PRACTICES

4.7 CRONBACH ALPHA; TESTING THE RELIABILITY

The Cronbach Alpha test was conducted to reaffirm the reliability of the items measuring three constructs of assessment practices. As shown in Table 17, scores of alpha coefficients of all measures ranged between .27 to .83. Formative assessment yielded the highest alpha value of .83 whereas assessment design represented the lowest alpha score of .27. The low alpha score was due to having only two items measuring assessment design. The summative assessment had an alpha coefficient of .63.

Constructs	Alpha Coefficient	No. of items
Formative Assessment	.830	9
Summative Assessment	.633	6
Assessment Design	.268	2

TABLE 17 CRONBACH ALPHA FOR ASSESSMENT PRACTICES

4.8. QUESTION 2

What assessment practices do Bhutanese elementary teachers value?

In response to the second research question, “What assessment practices are valued by the

Bhutanese elementary teachers?" a descriptive analysis was run to obtain the frequency(n), Mean(M), Standard Deviation (SD), and the percentage of respondents' assessment value by survey items. Table 18 summarised the descriptive values of each item. Further, a descriptive analysis was run for three unobserved variables: formative assessment, summative assessment, and assessment design to calculate the frequency, mean, and standard deviation.

As evident from Table 18, the highest mean score was in prompt constructive feedback (M=4.58). It suggested that the teachers considered giving prompt constructive feedback extremely important. Likewise, authentic assessment, oral presentation, self-assessment by students, and questioning received the next highest mean scores (M=4.13; 4.04;4.08;4.08). The scores suggested that the respondents consider these practices "very important". The rest of the assessment practices scored the means of not less than (M=3.6). Such scores indicated that all assessment practices were moderately important. Assessment designed by self and provided in the manuals or instructional guides were valued almost very important (M=3.78; M=3.72).

When looked at the percentages associated with the value of assessment practices, teachers' ratings suggested that 30.2% felt that "prompt constructive feedback" was "very important" and 64.6 % felt that it was "extremely important". For the rest of the assessment practices (items), the percentages were highest under the "very important" rating. Nearly 60% of the teachers rated these assessment practices as "very important" or "extremely important" except for major examination that was considered "very to extremely important" by slightly less than 55% of teachers. Interestingly, 4.2% rated that major exams were "not at all important" when 0% of the respondents rated the majority of the assessment practices as "not at all important". Meanwhile, about 1% expressed that performance quizzes, projects completed by individual students, authentic assessments, and projects completed by teams of students as "not at all important".

Items	n	M	SD	Not at all Impt. %	Low Impt %	Moderately Impt %	Very Impt %	Extremely Impt %
Assessment designed primarily by yourself	95	3.78	.70	0	3.2	28.4	55.8	12.6
Performance quizzes	94	3.60	.61	1.1	0	40.4	55.3	3.2
Objective assessments (eh MCQ, short answer, matching)	96	3.76	.79	0	5.2	30.2	47.9	16.7
Essay type questions	96	3.69	.88	0	9.4	31.3	40.6	18.8
Performance assessments	96	3.69	.73	0	6.3	28.1	56.3	9.4
Projects completed by individual students	96	3.80	.87	1.0	5.2	28.1	43.8	21.9
Major exams	96	3.63	1.1	4.2	8.3	31.3	35.4	19.8
Authentic assessments	96	4.13	.79	1.0	1.0	16.7	46.9	34.4
Project completed by teams of students	96	3.68	.85	1.0	7.3	29.2	47.9	14.6
Published Assessment	96	3.72	.77	0	4.2	35.4	44.8	15.6
Oral presentation	96	4.04	.72	0	2.1	17.7	54.2	26
Prompt constructive feedback	96	4.58	.62	0	1.0	4.2	30.2	64.6
Self -assessment by students	96	4.08	.76	0	3.1	15.6	51.0	30.2
Peer assessment by students	96	3.96	.67	0	1.0	21.9	57.3	19.8
Questioning	98	4.08	.88	0	2.0	14.3	53.1	30.6

TABLE 18 DESCRIPTIVE STATISTICS FOR FREQUENCY, MEAN AND PERCENT FOR ASSESSMENT PRACTICES

Additionally, 15 items of rating assessment practices (valuing/importance) were constructed into three major factors as portrayed in its CFA model, Figures 12 and 13. They were formative assessment, summative assessment, and assessment design. The descriptive analysis results in Table 19 and Figure 14 illustrated the composite mean of the three constructs. The teachers valued formative assessment more than summative assessment with a mean score of (M=3.94; SD= .73 and M=3.71; SD= .57) respectively. It meant the teachers considered that formative assessment was "moderately" to "very important". Likewise, they felt assessment design was "moderately" to "very important" in practising

effective assessment with a mean score of (M=3.75; SD= .56). Meanwhile, the standard deviation of each construct did not indicate greater variability in rating the importance.

	Formative Assessment	Summative Assessment	Assessment Design
Mean	3.94	3.71	3.75
Standard Deviation	.73	.57	.56

TABLE 19 DESCRIPTIVE STATISTICS FOR THREE CONSTRUCTS OF ASSESSMENT PRACTICE

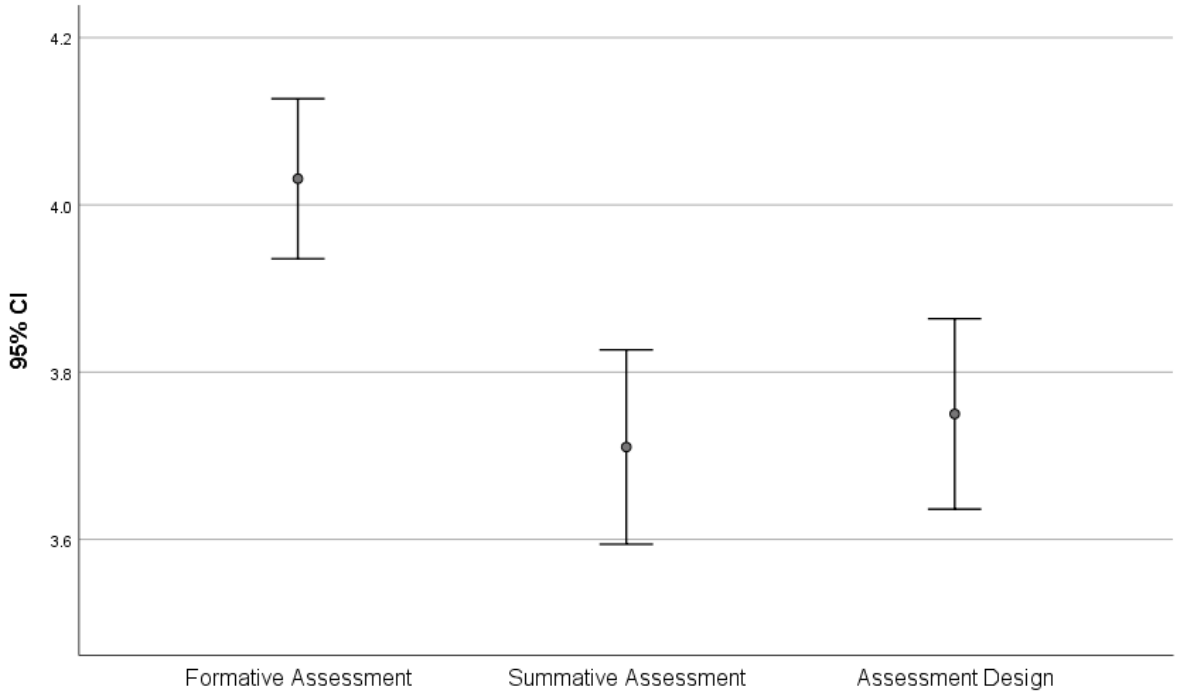

FIGURE 14 ERROR PLOT COMPARING MEAN SCORES FOR ASSESSMENT PRACTICE

4.9 QUESTION 3

What is the level of assessment literacy of elementary teachers in Bhutan as measured by the Classroom Assessment Literacy Inventory (CALI)?

The data to measure the level of assessment literacy was collected from a cognitive test. Therefore, the results presented were based on the raw scores of the test. The results were presented in three different ways. First, it presented the overall assessment literacy minimum, maximum, and average scores of Bhutanese elementary teachers. Second, it

presented the results measured against the Seven Standards of assessment literacy framed by AFT, NCME, and NEA, (1990). Lastly, the teachers have been categorised into three levels of assessment literacy, low, medium, and high based on the mean scores of each standard and overall raw score.

The results showed that, overall, the minimum total score of classroom assessment literacy was 1 point, while the maximum score was 13 points, out of the total score of 21 points, with a mean score of 8.82 and the standard deviation of 2.70. The average score was equivalent to approximately 42 %. When considering the scores by the seven standards for teacher competence in educational assessment of students, it was found that the mean scores ranged from .83 to 1.62, with a standard deviation of 0.65-.91. The standard that received the highest mean score was Standard 6: communicate assessment results (Mean = 1.62; SD = 0.81). On the other hand, the standard that had the lowest mean score was Standard 3: the ability to administer scores (Mean = .83; SD = 0.73), as depicted in Table 20.

Standards	N	Min (score)	Max (Score)	Mean	S. D
Choosing an assessment method	81	00	3.00	1.37	.81
Developing Assessment	80	00	3.00	.95	.69
Ability to administer score	80	00	3.00	.83	.73
Using assessment result	80	00	3.00	1.20	.84
Developing valid grading procedures	80	00	3.00	1.46	.91
Communicate assessment results	80	1.00	3.00	1.62	.81
Recognise unethical, illegal assessment methods	80	00	3.00	1.47	.65
Overall Raw scores		1.00	13	8.82	2.70

TABLE 20 SCORES FOR CLASSROOM ASSESSMENT LITERACY

The assessment literacy scores as measured by the Classroom Assessment Literacy Inventory could be categorised into three levels; low, medium, and high as presented in Table 21. On a scale of 0-3 (one standard has 3 items), the first category (low) scored either 0 or 1 out of 3, the second category (medium) scored 2 out of 3 and the last category (high) scored 3 out of

3. Overall, 62.25% of the teachers scored either 0 or 1 point, 30.8% of teachers scored 2 points and 6.95% scored 3 points.

While considering the scores for each Seven Standard, it was found that most teachers scored at a low level for all the seven standards. The highest percentage to score at low level was for Standard 3; ability to administer score (82.5%) followed by Standard 2; developing assessment methods (81.3%), Standard 4; using assessment result (65.1%), Standard 5; Developing valid grading procedures (55.1%), Standard 1; choosing an assessment method (54.3%), Standard 7; recognise unethical, illegal assessment methods (51.2%) and the lowest percentage being for Standard 6; communicate assessment results (46.3%). Overall, teachers were not more literate in any of the seven standards.

Seven Standards		**Classroom Assessment Literacy**					
		Low (0-1)		Medium (2)		High (3)	
	n	N	%	N	%	N	%
Choosing an assessment method	81	44	54.3	32	39.5	5	6.2
Developing Assessment	80	65	81.3	14	17.5	1	1.3
Ability to administer score	80	66	82.5	13	16.3	1	1.3
Using assessment result	80	52	65.1	23	28.7	5	6.3
Developing valid grading procedures	80	44	55.1	24	30	12	15
Communicate assessment results	80	37	46.3	31	38.8	12	15
Recognise unethical, illegal assessment methods	80	41	51.2	36	45	3	3.8
Overall Raw Scores	80	50	62.25	25	30.8	5	6.95

TABLE 21 LEVELS OF CLASSROOM ASSESSMENT LITERACY

4.10. QUESTION FOUR

Do the independent variables of gender, teacher education level, years of teaching experience, grade taught, subject taught and assessment education/training have a significant difference in teachers' assessment beliefs, assessment practices, and assessment literacy level?

Response to question 4 was to test if there were any significant differences the independent variables of gender, years of teaching experience, highest teacher education, grade taught,

subject taught and assessment education or training received would have on dependent variables of “assessment beliefs”, “assessment practices" and "assessment literacy". To determine a significant difference if any of the Gender on assessment beliefs, assessment practices, and assessment literacy, an Independent Sample t-test was employed. For other independent variables, One-Way ANOVA was employed to analyse the data selecting for Levene’s tests and post hoc tests. However, the results in the tables are consolidated only for those independent variables that had significant differences in various constructs of dependent variables.

4.10.1 Demographic information and Assessment Beliefs

There was no significant difference between gender (females and males), grade taught, and intensity of assessment education/training received on any of the constructs of assessment beliefs (Improvement, student accountability, school accountability, and Irrelevance). Nonetheless, years of teaching experience, level of teacher education and subject taught had statistically significant differences ($p<0.05$) on assessment belief constructs. The years of teaching experience as presented in Table 22 had significant difference on assessment for school accountability F(3, 97)=2.603, p = 0.058, 0-5years (M)=3.70 SD=.74), 6-10 years (M)=4.00 SD=.42, 11-15 years (M)= 3.52, SD=.53), and >15 years (M)=3.75 SD= 61 . The post hoc analysis in Table 23 showed that the significant difference was between years of teaching experience with 11-15 years and with more than 15 years, MD=.475, SE=.170, p=0.039.

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Assessment for school acc.	Between Groups	3.710	3	1.237	2.603	.056
	Within Groups	46.087	97	.475		
	Total	49.796	100			

TABLE 22 ONE -WAY ANALYSIS OF VARIANCE (ANOVA) RESULTS OF SIGNIFICANT DIFFERENCE IN ASSESSMENT BELIEFS (SCHOOL ACCOUNTABILITY) BY TEACHING EXPERIENCE

Comparison	Mean Difference	S.E	P-level at < 0.05
11-15 years VS more than 15 years	.475*	.170	.039

TABLE 23 POST-HOC TESTS (BONFERRONI) RESULTS OF SIGNIFICANT DIFFERENCE IN ASSESSMENT BELIEF (SCHOOL ACCOUNTABILITY) BY TEACHING EXPERIENCE

The level of teacher education as shown in Table 24 had significant difference on assessment for student accountability F(3, 97)=3.886, *p* = 0.011, PTC(M)=3.93, SD=.51, B Ed(M)= 3.91, SD=.53, PGDE(M)= 3.95, SD= .86, Masters(M)=3.38, SD=.50. The Post Hoc analysis in Table 25 illustrated that the difference was between PTC and master's teachers, MD=.556, SE=.195, p=.032, and between PGDE and master's teachers, MD=-.537, SE=.164, p=.009.

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Assessment for std acc.	Between Groups	3.585	3	1.195	3.886	.011
	Within Groups	29.832	97	.308		
	Total	33.417	100			

TABLE 24 ONE -WAY ANALYSIS OF VARIANCE (ANOVA) RESULTS OF SIGNIFICANT DIFFERENCE IN ASSESSMENT BELIEFS (STUDENT ACCOUNTABILITY) BY TEACHER EDUCATION LEVEL.

Comparison	Mean Difference	S.E	P-level at < 0.05
PTC vs Masters	.566*	.195	.032
B Ed vs Masters	-.537	.164	.009

TABLE 25 POST-HOC TESTS (BONFERRONI) RESULTS OF A SIGNIFICANT DIFFERENCE IN ASSESSMENT BELIEF (STUDENT ACCOUNTABILITY) BY TEACHER EDUCATION LEVEL.

Subject taught by the teachers as shown in Table 26 had significant difference on assessment is irrelevant F(4,98)=2.468, *p*=0.050, Dzongkha(M)=3.56, SD=.66, English(M)= 3.00, SD=.66, Mathematics(M)= 2.94, SD= .81, Science(M)=3.02, SD=.67 and Social Studies(M)=2.60,SD=.64. The post hoc analysis in Table 27 showed that the significant difference was between Dzongkha and Social Studies teachers, MD=.969, S.E= .347, p=.54.

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
As is Irrelevant	Between Groups	4.971	4	1.243	2.468	.050
	Within Groups	49.353	98	.504		
	Total	54.323	102			

TABLE 26 ONE -WAY ANALYSIS OF VARIANCE (ANOVA) RESULTS OF SIGNIFICANT DIFFERENCE IN ASSESSMENT BELIEFS (ASSESSMENT IS IRRELEVANT) BY SUBJECT TAUGHT

Comparison	**Mean Difference**	**S.E**	**P-level at < 0.05**
Dzongkha VS Social Studies	.969	.347	.054

TABLE 27 POST-HOC TESTS (BONFERRONI) RESULTS OF SIGNIFICANT DIFFERENCE IN ASSESSMENT BELIEF (ASSESSMENT IS IRRELEVANT) BY SUBJECT TAUGHT

4.10.2.Demographic information and Assessment Practises

Mean scores of assessment practices under three constructs formative assessment, summative assessment, and assessment design from 15 assessment practice items were compared for independent variables of gender, years of teaching experience, teacher education, grade taught, subject taught and assessment education /training received. Independent Sample t-test for gender and One-Way ANOVA was conducted for the remaining independent variables together with Levene's test and post hoc analysis.

There was no significant difference between years of teaching experience, levels of grade

taught, subject taught and intensity of assessment education/training received on any of the constructs of assessment practices (formative assessment, summative assessment, and assessment design). Nonetheless, gender and levels of teacher education had a statistically significant difference (p<0.05) on assessment practices constructs. Table 28 shows a significant difference between females and males on formative assessment, $t(95)=2.680$, p = .009, Females(M)=4.06, SD.48, Males(M)=3.62, SD 1.16. The standard deviation of 1.16 for males indicated that there was greater variability in their responses.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Formative Assessment	Equal variances assumed	6.788	.011	2.680	95	.009	.44771	.16707	.11603	.77939
	Equal variances not assumed			1.866	26.921	.073	.44771	.23999	-.04478	.94019

TABLE 28 INDEPENDENT SAMPLE T-TEST SHOWING THE SIGNIFICANT DIFFERENCE IN ASSESSMENT PRACTICES (FORMATIVE ASSESSMENT) BY GENDER

The level of teacher education as shown in Table 29 had statistically significant difference on assessment design, $F(3,91)=3.467$, p=0.019, PTC(M)=4.00, SD=.64, B Ed(M)= 3.75, SD=.51, PGDE(M)= 3.28, SD= .56, Masters(M)=3.57, SD=.44. The post hoc analysis in Table 30 showed that a significant difference was between PTC teachers and PGDE teachers, MD=.414, SE=.239, p=0.022.

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Assessment Design	Between Groups	3.019	3	1.006	3.467	.019
	Within Groups	26.413	91	.290		
	Total	29.432	94			

TABLE 29 ONE -WAY ANALYSIS OF VARIANCE (ANOVA) RESULTS OF SIGNIFICANT DIFFERENCE IN ASSESSMENT PRACTICES (ASSESSMENT DESIGN) BY THE LEVEL OF TEACHER EDUCATION

Comparison	**Mean Difference**	**S.E**	**P-level at < 0.05**
PTC vs PGDE	.714	.239	0.022

TABLE 30 POST-HOC TESTS (BONFERRONI) RESULTS OF SIGNIFICANT DIFFERENCE IN ASSESSMENT PRACTICES (ASSESSMENT DESIGN) BY TEACHER EDUCATION

4.10.3.Demographic information and Assessment Literacy level

To determine the significant difference of gender for assessment literacy, an Independent Sample t-test was conducted. For the remaining independent variables, One Way ANOVA, Leven's test, and post hoc analysis was run. There was no significant difference between gender, years of teaching experience, levels of grade taught, subject taught, and intensity of assessment education/training received on the raw scores of assessment literacy. However, as shown in Table 31, levels of teacher education had statistically significant difference (p<0.05) on assessment literacy scores, F (3,65) =3.282 p=0.026, PTC(M)=9.18, SD=.1.88, B Ed(M)= 8.47, SD= 2.43, PGDE(M)= 8. 50, SD= 3.69, Masters(M)=11.25, SD=1.28. The post hoc analysis in Table 32 showed a significant difference between B Ed teachers and PGDE teachers, MD=-2.77, SE=.89, p=0.017.

ANOVA

Assessment Literacy

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	53.864	3	17.955	3.282	.026
Within Groups	355.615	65	5.471		
Total	409.478	68			

TABLE 31 ONE -WAY ANALYSIS OF VARIANCE (ANOVA) RESULTS OF SIGNIFICANT DIFFERENCE IN ASSESSMENT LITERACY BY THE LEVEL OF TEACHER EDUCATION

Comparison	**Mean Difference**	**S.E**	***P*-level at < 0.05**
B Ed vs Masters	-2.77	.89	0.017

TABLE 32 POST-HOC TESTS (BONFERRONI) RESULTS OF SIGNIFICANT DIFFERENCE IN ASSESSMENT LITERACY BY TEACHER EDUCATION

4.11. QUESTION 5

How is the dependent plus endogenous variable of teachers' assessment practices affected by the independent variables of gender, years of teaching experience, level of teacher education, grade taught, subject taught, and assessment education/training; and by the exogenous variables of assessment beliefs and assessment literacy?

Although independent sample t-test and series of one way ANOVA tests confirmed that there were no significant differences in assessment beliefs, assessment practices and assessment literacy measured by demographic/teacher characteristics (independent variables) such as gender, length of teaching experience, highest teacher qualification, grade taught, subject taught and types of assessment education/training received, except for few sub-constructs, this study investigated how much these demographic characteristics plus assessment beliefs and assessment literacy affected assessment practices of Bhutanese elementary teachers. It was hypothesised that the demographic/teacher characteristics and the two constructs of

assessment; assessment beliefs and assessment literacy would have positive effects on assessment practices as shown in Figure 2 (Chapter 3).

4.11.1Structural Equation Model (SEM)

To investigate the effects of demographic variables, assessment beliefs, and assessment literacy on assessment practices, Structural Equation Models (SEM) were conducted. The preceding CFA models (Figures 9 & 12) were further extended to construct this structural model to demonstrate the effects of the chosen variables on assessment practices.

The structural model interpreted the effects of all the independent variables on the dependent variable of assessment literacy, effects among the latent variables, and also the effect of exogenous variables on endogenous variables. During the initial run of the path model, majority path coefficients were found to be non-significant at (p<0.05). All such non-significant direct paths were subsequently deleted for achieving a more parsimonious model. Eventually, the structural model analysed using AMOS showed the final models using unstandardized coefficients (b) Figure 15 and standardised coefficients (β) Figure 16 to determine the influence of all identified variables in this study on assessment practices.

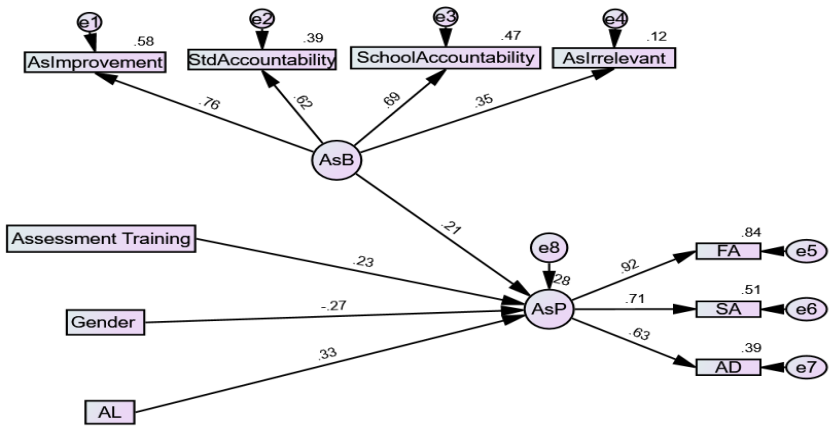

FIGURE 15. SEM PATH MODEL (STANDARDISED)

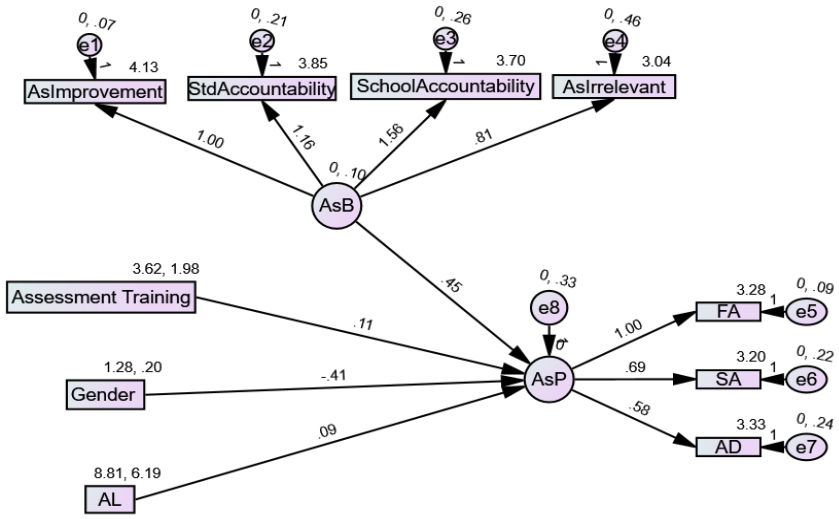

FIGURE 16. SEM PATH MODEL (UNSTANDARDISED)

4.11.2. Direct, Indirect, and Total Effects on Assessment Practises

Table 33 displayed the direct, indirect, and total effects of independent variables of gender, assessment literacy, assessment beliefs, and assessment education/training on the dependent variable, assessment practices. It is evident from Figures 15 and 16 and Table 33 that two independent variables (gender and assessment education/training) and two exogenous variables (assessment beliefs and literacy) had direct effects on assessment practices of Bhutanese elementary teachers. Assessment literacy had a greater direct effect on assessment practices, (β=.326, b=.089). It may be interpreted that the higher the assessment literacy, effective the assessment practices would be. The second greater direct effect on assessment practices was from assessment training (β=.229, b=.110). It may be, thus, understood that when teachers receive adequate and frequent assessment education/training in their teacher education and through national/district/school level PD programs they tend to implement effective assessment practices such as effective assessment design, formative assessment, and summative assessment. The next impacting

variable on assessment practices was assessment beliefs (β=.210, b=.454). It suggested that when teachers have positive attitudes toward assessment such as assessment is for improving teaching and learning, they tend to value formative assessment. The effect was also statistically significant in terms of gender (β=-.272; b=-.411). However, the path coefficient is negative. It may imply that the path coefficient of gender to assessment practices indicate that female teachers perform effective assessment practices than male teachers. Table 33 also demonstrated the indirect and total effects of the variables that were accounted for direct effects on assessment practices. Surprisingly, there were no indirect effects on assessment practices from the shown variables in the model. Therefore, the total effects from all the independent variables (assessment beliefs, assessment education/training, gender, and assessment literacy) on the dependent variable of assessment practices are the same as the direct effects. The absence of indirect effects further implied that assessment beliefs and assessment literacy did not have a mediating influence between independent variables (demographic) and dependent variables (assessment practices).

		Direct Effects					Indirect Effects		Total Effects	
Dependent Variable	Independent Variable	Unstandardised (b)				Standardised (β)	Unstandardised	Standardised	Unstandardised	Standardised
		Estimates	S.E	C.R	p	Estimates				
AP (Assessment Practices)	AB (Assessment Beliefs	.454	.250	1.81	.058	.210	.000	.000	.454	.210
	Assessment Training	.110	.047	2.35	.019	.229	.000	.000	.110	.229
	Gender	-.411	.147	-2.79	.005	-.272	.000	.000	-.411	-.272
	CALI (Assessment Literacy)	.089	.030	2.96	.018	.326	.000	.000	.089	.326

TABLE 33 DIRECT, INDIRECT AND TOTAL EFFECTS OF INDEPENDENT AND EXOGENOUS VARIABLES ON THE DEPENDENT, ENDOGENOUS VARIABLE (ASSESSMENT PRACTICES)

4.11.3.The goodness of model fit indices

It is a prerequisite to evaluate the goodness of fit of the SEM for explaining the relationship between the latent and measured variables (Li & Qiu, 2018). The Chi-Square, Chi-square/df, Tucker-Lewis index (TLI), Comparative fit index (CFI) and Root mean square error of approximation (RMSEA) is considered as the tests of the goodness-of-fit. The fit indices in Table 34 highlight that the chi-square (x^2= 54.353) is a reasonable value. The ratio of chi-square to degrees of freedom (x^2/df= 1.599) which may be considered acceptable as the lower the better, and there is no universally accepted value (Li & Qiu, 2018). The Comparative Fit Index (CFI) was 0.838 which was acceptable. The Tucker-Lewis Index (TLI) was found .738 which seemed quite poor. The indices suggested that the model fitted the data poorly yet explained the impact of observed variables on latent variables. Generally, RMSEA is acceptable with a 0.08 value, but less than 0.05 is better, and less than 0.01 is a perfect model (Kline, 2015). The (RMSEA) was found 0.073 which was acceptable but not better. Overall, these fit indices demonstrated a theoretically accepted but not a perfect model that explained the data well.

Indices	Scores	Remarks
Chi-Square (x^2)	54.353	Acceptable
Chi-Square/df (x^2/df)	1.599	Acceptable
Tucker-Lewis Index (TLI)	.738	Quite poor
Comparative Fit Index (CFI)	.838	Acceptable
Root Mean Square Error of Approximation (RMSEA)	.073	Poor

TABLE 34 SUMMARY OF GOODNESS-OF-FIT INDEXES FOR SEM MODEL.

CHAPTER FIVE: DISCUSSION, RECOMMENDATIONS, AND CONCLUSION

5.1 INTRODUCTION

The purpose of this study is to investigate Bhutanese elementary teachers' assessment beliefs, assessment practices, and assessment literacy level to gain better understandings of teachers' beliefs and practices related to formative assessment. This chapter presents a discussion of findings in response to the research questions identified in Chapter 1. These questions are discussed in comparison with the findings of the previous studies reviewed in Chapter 2. According to the findings of this study, recommendations are provided for future assessment policies, frameworks, and education/training, along with possible directions for further research on the area. Finally, this chapter also presents the scope and limitations of the study.

5.2. ASSESSMENT BELIEFS

To measure the assessment beliefs of Bhutanese elementary teachers, Brown's Conceptions of Assessment III abridged version (COA-III A) Inventory (2006), with slight modifications by Clevair (2010) and Rosas (2014), was employed. Brown (2004, 2006, 2008, 2011) identified four purposes/intentions of assessment: assessment for improvement, assessment for student accountability, assessment for school accountability, and assessment as irrelevant). The assessment beliefs of Bhutanese elementary teachers were measured within these four constructs of assessment, specifically, to what extent Bhutanese teachers agreed or disagreed with these conceptions.

The results of the descriptive analysis showed that teachers scored the highest mean value concerning the belief that assessment is for improvement (M=4.13) and the lowest mean concerning assessment as irrelevant (M=3.03). These results suggest that teachers agree (but not strongly agree) that the purpose of assessment is for improving teaching and learning.

This study's findings, therefore, mirrored the findings of Brown (2008) with New Zealand and Queensland teachers, and Clevair (2010) with teachers in central Virginia in the United States. Primary school teachers in New Zealand and Queensland, as well as Grade 3 through Grade 5 teachers in central Virginia, agreed more with the belief that the purpose of assessment is for improvement (Clevair, 2010). Unlike those studies, though, the current study found that teachers held strong neutral views on assessment for student and school accountability (M= 3.86 and 3.70) and held weak neutral views for assessment is irrelevant (M=3.03). By contrast, the New Zealand and Queensland teachers weakly agreed with student accountability but disagreed with school accountability and assessment is irrelevant. In Clevair's (2010) study, the teachers held a slightly neutral view about school accountability and that assessment is irrelevant. Further, the mean scores of assessment beliefs in this study revealed that, though teachers viewed assessment is to improve teaching and learning, they usually practised assessments to make students accountable, and that through assessment teachers and schools are also made accountable. However, these beliefs related to accountability were not made explicit, as there was a strong neutral response to them by teachers. A possible explanation for this is that it is culturally more likely in Bhutan, negative beliefs are not clearly expressed, and therefore teachers exhibited negative beliefs by not agreeing or disagreeing. Further, teachers' weak neutral score for assessment is irrelevant suggests that they believe assessment interferes in their teachings.

The results of the relationships between the four beliefs showed that they were positively correlated. There was a statistically significant positive association of assessment for improvement and student accountability (r=.72; p=0.01), implying that promoting student accountability by examination can improve their learning. This may also mean that teachers can improve students' learning by teaching for exams. A similar and strong positive correlation was observed with Hong Kong teachers (r=.91) (Brown et al ., 2009), and Chinese teachers (r=.80) (Brown et al., 2011); as well as with New Zealand and Queensland teachers, though with only a weak positive correlation (Brown, 2004; 2008). This comparison shows that, Bhutanese elementary teachers believe in making students accountable by teaching them for examinations, and that examining them can improve students' learning, this belief is almost as strong as Chinese and Hong Kong teachers' beliefs. Bhutanese students receive their entry in tertiary education based on their performance in Grade 12 and Grade 10 which

might have shaped teachers' beliefs of assessment that showed a strong positive correlation between improvement and student accountability.

The two accountability beliefs related to schools and students also had a positive moderate correlation (r=.66), similar to the observation made with New Zealand and Queensland teachers (Brown, 2004; 2008). This implies that, as teachers and schools in Bhutan are evaluated based on students' overall examination results, making students accountable may improve the performance of teachers/schools, or that by making teachers/schools accountable, student accountability can be raised. In other words, this result implies that if an assessment can be used to improve student accountability, this also provides a basis for schools to be evaluated. A moderate positive correlation between improvement and school accountability (r=.61) suggests that Bhutanese teachers believe that assessments for school accountability functions to improve teaching and learning. In other words, evaluating schools/teachers based on examination results is a way of improving teaching and learning. A similar finding was observed in Brown's (2004; 2008) study of New Zealand and Queensland teachers, and Clevair's (2010) study of teachers in central Virginia.

Similarly, the belief assessment is irrelevant was positively correlated to student accountability (r=.37). This finding matched the findings from New Zealand's primary teachers (Brown, 2004; 2008) that assessing students by examining and grading is irrelevant. On the other hand, Bhutanese teachers expressed a belief that making students accountable through gradings and examinations may improve their learning. Surprisingly, unlike any of the past studies, this study found that assessment for improvement was positively related to assessment is irrelevant (r=.15). This finding suggests that Bhutanese elementary teachers do not view these two beliefs as opposites. This association illustrates that the more Bhutanese elementary teachers believe that assessment is for improvement of teaching and learning, they tend to consider that, assessment interferes with their teaching. This finding conflicts with Bhutanese elementary teachers' highest mean score for assessment is for improvement. Finally, the belief that assessment is irrelevant had almost no relation with school accountability (r=.05). This finding too mirrored the finding with New Zealand teachers, which rejected the international notion that assessment for school accountability is not healthy for quality schooling (Brown, 2008). The finding illustrated that Bhutanese teachers believe that school accountability through assessment brings quality schooling. One of the significant

concerns that emerged in this study is the positive relation of irrelevance with three remaining conceptions (improvement, student, and school accountability), which means teachers believe that assessment is irrelevant and interferes with teaching and learning.

Overall, the structure of Bhutanese elementary teachers' assessment beliefs appears conflicting and complex. A high moderate correlation between improvement and accountability (student and school), and a low but positive correlation between improvement and irrelevance suggest that Bhutanese elementary teachers want to practise assessment for improvement, but as teachers and schools are evaluated based on students' end of year academic results, the teachers tend to understand that assessment is irrelevant if an assessment has to be used for improvement rather than for accountability.

5.3. ASSESSMENT PRACTICES

Examining the results of the descriptive analysis it was found that Bhutanese elementary teachers valued prompt constructive feedback the most (M=4.58). 94.8% of teachers either rated prompt constructive feedback as very or extremely important. Effective feedback is constructive which is central to formative assessment (Wood, 2019). Similarly, high mean scores for self-assessment, peer assessment, and questioning indicate that the teachers valued formative assessment practices as “very important”. More than 80% of the respondents rated these elements either very or extremely important. The findings of self-reported assessment practices in this study, which also included the elements of formative assessment and the findings of real-time practices of formative assessment observed by Utha (2014) in the Bhutanese context, are not coexisting. Utha (2014), in her qualitative study performed through observation, interviews, focus groups, and action research, found that the features of formative assessment such as questioning (dialogue), feedback, peer and self-assessments were poorly established and almost absent in her participants' teaching. The most conflicting feature between teachers' self-reported ratings in this study and Utha's (2014) observations was related to feedback. The high mean score of the importance of constructive feedback found by this study contrasts with the generic and minimal feedback observed by Utha (2014) with her participants, which raises concerns about what teachers report and what they practice. Without these core elements, in particular feedback, an assessment cannot be formative. According to Andrade, et al. (2015), the effects of formative

assessment stem from the strength of feedback given to the students. Feedback does not necessarily come just from teachers, but also students themselves. Therefore, self and peer-assessments are essential in formative assessment. Meanwhile, asking questions builds a series of dialogues in a lesson.

In terms of types of assessment practices, the Bhutanese elementary teachers valued authentic assessment the most, with a mean score of 4.13. A similar finding was noted by Cleviar (2010) (M=4.32)., while McMillian et al. (2002) found that authentic assessment was among the least-observed practices. This difference seems to stem from McMillian's approach, which sought to record the frequency of usage, whereas Cleviar (2010) and this study aimed to record the value teachers associated with each of the types of assessment practices. The difference indicates that teachers may value an assessment type as essentially crucial, but they may not practise it as much as they value it. Oral presentation, too, received a higher mean score than other types of assessment (M=4.04), indicating that Bhutanese teachers feel that oral presentation is “very important”. This result indicates that Bhutanese teachers value both alternative assessments and traditional assessments.

On the other hand, traditional assessment types such as objective assessments (MCQ and matching), major exams, essay type questions still received higher mean scores. However, only 4.2 % of participants rated major examinations as “not at all important”. With the removal of examinations from Pre-primary to Grade 3 in 2020, it was expected that a higher percentage of teachers would have indicated this. Simultaneously, alternative assessment types such as projects by individual students / a team of students, performance assessments, and quizzes obtained similar scores with traditional assessment types. This balanced value assigned to both alternative and traditional assessment types suggests that Bhutanese teachers attach a greater value to a wider spectrum of assessment type. Meanwhile, all types of assessment (alternative or traditional) were rated either “very” or “extremely important”, which represented a major percent in comparison to other ratings.

A balanced value was also noted in the assessment practices of self-designed assessments and published assessments in manuals and guidebooks (M=3.78; 3.72). This rating was similar to the findings made by McMillian et al (2002) (M=3.78; 3.38), but differed from Clevair’s (2010) findings, especially in terms of published assessments, which was valued much lower

than self-designed assessment (M=3.84; 2.69). The similar mean scores in this study imply that Bhutanese teachers make use of both types of assessment (designed by self and published) equally. Additionally, when the percentage result in Table 18 (Chapter 4) was examined, 68.4% rated self-designed assessment as very important to extremely important, and 60.4% rated the same for published assessment. Valuing self-designed assessment slightly more (both in mean scores and percentage) than published assessment indicates the need for better training/education and assessment literacy for teachers so that they can prepare reliable and valid assessment tools.

The results of the three constructs of assessment practices, (formative assessment, summative assessment, and assessment design) highlighted that teachers value them equally. This finding replicated the findings of the item-wise analysis of assessment practices. The composite mean scores of the three constructs were very close to each other (M=3.94, 3.71, and 3.75), indicating that Bhutanese teachers working under a policy emphasising formative assessment were aware of its implications, and yet valued summative assessment to meet the expectations of high-stakes accountability for teachers and schools. There thus appears to exist tensions in what teachers valued and what they have to practise. With an over-emphasis on formative assessment, tensions between formative and summative assessment practices are revealed especially where high-stakes accountability exists (OECD, 2006; Song & Koh, n.d). This is a paradox when evidence is available suggesting that formative assessment practices have positive impacts on students' standardised scores (Wiliam et al., 2004; Kingstong & Nash, 2011; 2015). However, it is difficult to conclude from this study whether formative assessment and summative assessment are equally practised in the real classroom as much they are valued because this study did not evaluate the usage of assessment types, but the rated value of different types of assessment practices.

Further, a significant positive association between the three constructs of assessment practices (formative and summative assessment, and assessment design) reveals that the better the assessment design, the more effective both the formative and summative assessments are. This finding also indicates that the integration of formative assessment and summative assessment is essential, rather than practising them in isolation, for which knowledge of effective assessment design is crucial for teachers. Researchers argue that formative assessment and summative assessment can be used as complementary to each

other (Black et al 2010; 2011; Looney, 2011). Besides supporting learning, formative assessment can be utilised to summatively assess the understanding of learners (Bennett, 2011; Kingston & Nash, 2011). According to Birenbaum et al (2006), a learner's understanding may be evaluated at the terminals (end of a chapter, unit, term, year, or a course) by using the evidence collected during the instruction through formative assessment. This evaluation can then be used to adapt subsequent instruction (Black & Wiliam, 2009; Perie, Marion & Gong, 2009). Therefore, better assessment training/education is vital for teachers to develop their capacities and competencies in applying formative and summative assessments in integrated and complementary ways. This way tensions between formative and summative assessment may be addressed. Brown and Harris (2016) posit that balancing formative and summative assessments and navigating their contending purpose is what assessment knowledgeable teachers do rather than emphasising one. However, balancing the two and navigating their competing purpose may be a daunting challenge for teachers, which necessitates support and interventions in building the capacity of the teachers.

5.4. ASSESSMENT LITERACY

Measured using CALI, the overall mean score of 8.8 (42%), with a maximum score of 13 (61.9%), revealed low level of assessment literacy in Bhutanese elementary teachers. Similar findings were noticed in Thailand with primary teachers (Yamtim & Wongwanich, 2014) and in the Philippines with primary and secondary teachers (Hailaya, 2014). This study confirmed the claim made by Popham (2009), who asserts that most teachers know little about educational assessment, and that of Stiggins (2014), who observed low levels of assessment literacy among in-service teachers and leaders. This study also supported the past findings of low levels of assessment literacy among in-service teachers (e.g. Plake et al 1993; Mertler, 2005) and provided evidence in the Asian context in addition to the other few research projects (e.g. Yamtim & Wongwanich, 2014; Hailaya, 2014). Therefore, teachers in general have a low level of assessment literacy.

Meanwhile, examining standard wise performance, Bhutanese teachers obtained the highest mean for Standard 6, (communicate assessment results) and the lowest mean for Standard 3, (ability to administer score) (M=1.62; .83, respectively). This result was the opposite of one

of the earliest studies conducted by Plake et al (1993) and also by Mertler (2005) on assessment literacy. The participants in the studies by Plake et al's (1993) and Mertler (2005) were more literate in Standard 3 than they were in Standard 6. The result suggested that while Bhutanese teachers, to a certain extent, are more literate in communicating assessment results, they nonetheless have low literacy in administering and interpreting assessment scores.

In terms of Standard 1 (choosing assessment methods for instructional decision) and Standard 2, (developing assessment methods appropriate for instructional decision), the mean scores (M= 1.37 and .95, respectively) revealed that Bhutanese teachers can choose appropriate assessment methods but cannot develop assessment methods by themselves. This indicates that teachers may lack clear methods of establishing the reliability and validity of any measuring tools. For Bhutanese teachers, it is argued that, since there are no separate assessment standards documented in the policy, they are not aware of general assessment standards. However, since all public-school teachers undergo certain teacher education courses, they receive assessment education in the course. For that matter, it is expected that teachers are better literate in the assessment. Hailaya (2014), as a tutor in the Curriculum and Assessment of Learning, observed that pre-service teachers lack adequate knowledge and understanding of key assessment ideas, but they are expected to conduct effective assessment tasks when they join the teaching profession. Similar circumstances must be occurring in the Bhutanese context which needs further investigation. Surprisingly, none of the studies referred to by this study showed a level of assessment literacy above the minimum benchmark of 70%, both in terms of overall scores and standard wise scores. The past trends and the findings of this study suggested there are concerns with the assessment literacy of teachers, for which rigorous interventions are necessary.

5.5 SIGNIFICANT DIFFERENCES OF INDEPENDENT VARIABLES ON DEPENDENT VARIABLES

The characteristics of gender did not show statistically significant differences in assessment beliefs and assessment literacy. However, there was a statistically significant difference identified between gender and sub assessment practice (formative assessment), suggesting that gender can influence values that teachers associate with formative assessment. Higher mean score by female teachers on formative assessment indicates that they strongly agree

with improvement conception about assessment, unlike the male teachers. Since no past study noted such an effect determined by gender on assessment practice, this finding needs further validation. Further, since only one of the three assessment practices showed a statistically significant difference, this also suggests that gender was not powerful in shaping teachers' assessment choices.

Years of teaching experience tended to influence assessment beliefs, but not assessment practices and assessment literacy. A statistically significant difference was noted for assessment beliefs (school accountability) caused by years of teaching experience. The difference appeared between teachers with an experience of 11-15 years and more than 15 years. Given the higher mean as scored by teachers with 11-15 years of experience (M= 4.00), than teachers with more than 15 years of experience indicates that new teachers believe the assessment is a means to make a school a better performing school. However, since only one conception (school accountability) had a statistically significant difference determined by years of experience, it was not powerful in shaping teachers' beliefs about assessment.

The level of teacher education showed a significant difference in assessment belief (student accountability), assessment practices (assessment design), and assessment literacy. For assessment belief of student accountability, differences were noted between master's degree and PTC qualification, and master's degree and B Ed qualification. Given the higher means scored by PTC education level, there is a higher degree of agreement that assessment is for student accountability. In other words, teachers with higher qualifications tend to believe less that assessment is for student accountability. Clevair (2010) had noted a similar trend, where teachers who did not attain degrees beyond the bachelor's level believed that assessment is for student accountability. Since the effect of the level of teacher education was not present in all sub-groups of assessment beliefs and assessment practices, it is however negligible in shaping the constructs of assessment beliefs and practices. The higher mean scored by master teachers on assessment literacy is clear to conclude that obtaining higher degrees in teacher education improves the assessment literacy level. This finding mirrors the finding of Rosas (2014), where teacher education level was one of the strongest indicators to determine teachers' assessment literacy level. The significant differences in assessment beliefs, practices, and literacy suggest that teacher education is pivotal in changing assessment beliefs

and practices and increase the level of literacy for which teacher educators and institutions have the bearing towards this progress.

The subjects taught had a significant difference in assessment beliefs (Irrelevance) but not on practices and literacy levels. Dzongkhag teachers agreed more than other subject teachers that assessment is irrelevant. It may be interpreted that Dzongkha is comparatively a new developing language/subject it does not yet have enough assessment tools and techniques. It is also likely that Dzongkha teachers are not given equivalent teacher education compared to other teachers due to a lack of resources written in Dzongkha.

The grade taught and assessment education/training received did not have a significant difference in assessment beliefs, practices, and literacy. The result is not consistent with prior studies. For example, a study by Clevair (2010), found that assessment education/training had a significant impact on the assessment belief of student accountability as well as on assessment practices. On the other hand, Rosas (2014) found no significant difference in assessment beliefs and practices, determined by assessment education/training on assessment, similar to the current study. Further, the lack of a significant influence on assessment literacy by assessment training/education found in this study mirrors the finding of Rosas (2014) who found that assessment education/training in isolation did not influence assessment literacy. However, Rosas (2014) disclosed that, when assessment education/training interacted with other independent variables (years of teaching experience and grade taught), it showed a significant influence on assessment literacy but not on assessment beliefs and assessment practices.

It was expected in this study that types or intensity of assessment education/training received by teachers would have a significant influence, at least on their assessment literacy. The reason for this expectation is that nearly all public-school teachers undergo teacher education training for two years (PTC-which does not exist now), for four years (B. Ed) and for 1 year (PGDE), where they are taught the principles of assessment and evaluation. Furthermore, need-based workshops are conducted with teachers whenever new reforms are introduced. Nonetheless, only those teachers whose roles match with the purpose of workshops are required to attend need-based workshops. Teachers who are teaching pre-primary to Grade 3 received formative assessment training this year (2020) beginning. Therefore, it was

anticipated that assessment education/training would have a significant impact on all three dependent variables. It is considered that this sample did not reflect the Bhutanese teachers' levels of education and training in assessment received during their teacher education, and national and school professional development programs. The non-significant difference of assessment education/training received on all three major dependent variables of assessment beliefs, assessment practices, and assessment literacy may indicate that such education and training have failed to address, or influence teachers' beliefs, practices, and literacy levels related to assessment.

Overall, the results of demographic characteristics and assessment beliefs and assessment practices confirmed/support the findings of Brown (2008), that none of the teacher characteristics had a significant influence on the conceptions of assessment held by New Zealand teachers. Therefore, the demographic characteristics of teachers seemed to be irrelevant in determining how strongly a teacher agreed with each conception of assessment beliefs, or how strongly a teacher valued each of the assessment practices, or how much literate a teacher was in the assessment.

5.6. EFFECTS ON ASSESSMENT PRACTICES

The structural model in Figures 15 and 16 detected that assessment literacy had the greatest effects on assessment practices (β=.326, b=.089=, and (β=.229, b=.110, respectively). 32% of assessment practices are influenced by the level of assessment literacy. This implies that when teachers are sufficiently literate about assessment, they tend to practise effective assessment methods/tools/techniques. However, this finding contradicted the findings of Siegal and Wissehr (2011), who noted the misalignment of assessment knowledge and assessment tools designed among preservice teachers. Similarly, in the study conducted by Ogan-Bekiroglu and Suzuk (2014), gaps were noted between assessment literacy and its implementation into practice with preservice teachers. From the literature search conducted for this study, it appears that no studies to date have been carried out to test the direct effects of assessment literacy on assessment practices among in-service teachers. The next impacting variable on assessment practices was the assessment education/training that teachers receive (β=.229, b=.110), with assessment practices being influenced by assessment education/training by almost 23%. On contrary, ANOVA analysis determined no significant

differences by the types or intensity of assessment education/training. This finding requires further validation as there appears to be no research yet conducted in the literature that has examined the mechanisms of how assessment literacy affects assessment practices. The effect in the result suggests that when teachers receive adequate and frequent assessment education/training in their teacher education and through national/district/school level PD programs, they tend to implement effective assessment practices.

Assessment beliefs influenced assessment practices by 21% (β=.210, b=.454). The existing research into teacher beliefs or conceptions shows that teachers' beliefs about assessment influence their assessment practices (Brown, 2008; Brown &Gao, 2009; Barnes, Fives & Dacey, 2015; Fives & Buhel, 2012). In particular, Brown and Gao (2009) found among Hong Kong teachers that assessment beliefs predicted assessment practices. Assessment practices related to diagnosing student learning needs were aligned with the conceptions of improvement of assessment. Similarly, preparing students for examination was aligned with the conception of making students accountable. Using examinations to evaluate the quality of school was aligned with the conception of school accountability, and irrelevance practice was aligned with the belief that assessment is irrelevant (Brown et al. 2009). Thus, if teachers hold positive beliefs about assessment, such as improvement of teaching and learning, they may practise more formative assessment and design their assessment activities concerning learning goals. Similarly, if they have negative beliefs about assessment, such as accountability and irrelevance, they may practise more summative assessment and use published assessment items, and sometimes may not even frequently assess their teaching and students' learning. The effect was also statistically significant in terms of gender (β=-.260; b=--.390). However, the path coefficient was negative. The path coefficient of gender to assessment practices indicated that female teachers perform more effective assessment practices than male teachers and thus suggest that male teachers require further assessment education or perhaps develop positive assessment beliefs in them.

Assessment literacy did not affect assessment beliefs, either directly or indirectly. The finding corresponds with the findings of Deneen & Brown, (2016) who conducted their study with 32 in-service and pre-service teachers. They noted that teacher education programs increased participants' assessment literacy but did not alter their conceptions.

Brown (2004) argued that if teachers believe that assessment is irrelevant, helping teachers to gain a higher level of assessment literacy and change their beliefs is of little value, especially in the context where assessment is more visible for accountability purposes. Since the teachers in this study did not disagree with the irrelevance conception about the assessment and rather, they associated it positively with the conceptions of improvement, it appears that assessment literacy did not affect their assessment beliefs. The implication of disconnection between assessment literacy and assessment beliefs in this study suggests that assessment programs in teacher education and professional development programs on assessment are not aligned with teachers' conceptions of assessment. Therefore, investigating teachers' conceptions of assessment and accounting for them in assessment education and training programs is imperative in the Bhutanese context to establish a clear connection between assessment beliefs, practices, and literacy.

5.7. RECOMMENDATIONS

As Bhutanese elementary teachers in this study agreed that the purpose of assessment is for improving learning/teaching, they require knowledge of a range of assessment tools to effectively implement formative assessments in the classroom. Similar to preceding studies by Black and Wiliam (1998) and Brown (2008), the prominence attached to an array of assessment practices emphasised that teachers need to use a variety of assessment tools and methods, both informal and formal, aimed at improving learning/teaching. Therefore, teachers need to be provided with high-quality assessment methods and tools that facilitate improvement.

The high positive association found between assessment for improvement and student accountability suggests that removing major examinations is unlikely to be successful. Similarly, the higher medium positive association between assessment for improvement and school accountability indicates that assessment policy needs to draw strong commitment from teachers to improve their instruction as well as improve their students' learning. Brown (2004, p.315), suggests that assessment may be planned in such a way that it becomes a means of giving feedback to the practitioners of their practices. Further, the positive

association between assessment for improvement and assessment as irrelevant is a concern for policymakers.

Bhutanese elementary teachers highly valued formative assessment components such as giving constructive feedback, self and peer assessment, and questioning. These findings suggest that when professional development plans related to assessment are crafted at any level (school, district, and national), the planners may wish to further probe into teachers' understandings of formative assessment, and more importantly into feedback aspects. Teachers' capacities in giving constructive feedback should be developed because the information given in feedback must be scaffolded to ensure that it is constructive. Safii and Wong (2017) found that primary school students still look for teachers' affirmation as feedback although it may not help in their learning. Therefore, it is expected that elementary students receive praise along with constructive feedback, as praise gives students a sense of achievement and success.

Within the given range of assessment practices, authentic assessment and oral presentations were rated with higher values than other assessment practices. The rest of the practices were rated as moderately important. Therefore, it can be recommended that teachers' understandings of authentic assessments and oral presentations, as well as other varieties of assessment tools, may need to be further explored and developed. Meanwhile, the lack of distinct values rated between self-designed and published assessments indicate that teachers need to develop their capacities for designing valid and reliable assessment tools. At the same time, textbooks and manuals may also present a wide variety of assessment options.

Similar values rated against formative assessment practices and summative assessment practices types within the array of given assessment practices suggest that Bhutanese elementary teachers are willing to adopt both types of assessments with proper integration and as complementary, rather than as contenders to each other. Therefore, it is crucial to support teachers through well- defined professional learning programs in building their competencies to implement formative and summative assessment as complementary modes and to remove tensions that teachers experience while implementing formative assessment.

The data related to assessment literacy revealed that Bhutanese elementary teachers have low levels of assessment literacy, as measured by the Classroom Assessment Literacy

Inventory (CALI). It is thus clear that primary education teachers (elementary teachers) in Bhutan need to acquire expertise in assessment to build their classroom assessment competencies. This requires teacher education colleges to design relevant assessment courses directed toward improving assessment literacy and assessment practices. To strengthen teachers' skills, knowledge, and experiences gained during their education, regular professional development related to assessment could be offered to in-service teachers. This can be achieved through collaborative learning; which teachers may adopt with the support of the district and school management. Further, as suggested by Stiggins (2014) and Englsen and Smith (2014), to develop effective assessments, it is not only teachers who should become assessment literate, but also related administrators such as school principals and policymakers. Therefore, training and professional developments could also be conducted for these administration staff. Developing assessment literacy standards in the Bhutanese context can be another strategy that guides assessment courses in teacher education colleges and school-level professional development programs related to assessment.

The data in this study showed that the assessment practices in realising educational goals are influenced by teachers' assessment beliefs, assessment literacy levels, the amount of assessment training/education they receive, and their gender. This finding implies that making visible teachers' beliefs about assessment is critical (Brown, 2008), as well as ensuring that these beliefs develop along with the implementation of new policies and guidelines. Brown (2008), further proposes that, through their belief systems, teachers tend to comprehend, experience, and implement assessment policies and frameworks, including the knowledge and skills received during their teacher education and professional development. Improving assessment literacy and increasing assessment training/education is therefore another step towards improving assessment practices. Meanwhile, the data indicate that male teachers require additional training and education to improve their assessment practices.

Given the positive beliefs about assessment held by Bhutanese elementary teachers, their considerable value they place on alternative assessment modes such as authentic assessments and oral presentations, and their desire to implement both formative and summative assessment approaches equally (despite low levels of assessment literacy), pre-

service teachers must be trained effectively, and in-service teachers are given relevant capacity-building opportunities. Utha (2014, p. 136), however, warns that professional developments targeted towards selected teachers and of a single cycle are not successful in building teachers' expertise. Consequently, teachers must be provided with enough time, resources, and professional support by school administration to help them realise their positive beliefs and practices of assessment. As such, Luyten and Dolker (2010, p. 434), advocate for investment in teacher education and assessment resources.

5.8. SCOPES AND LIMITATION AND FUTURE RESEARCH

This study offers several theoretical and practical implications for the field of assessment. The assessment beliefs, practices, and knowledge/literacy of elementary teachers are made visible to the teachers themselves, school leaders, and policymakers. Through this process, teachers are made aware of what they believe about assessments and which assessment practices they value most, while also being made aware of their assessment knowledge. It is anticipated that the visibility and awareness of assessment beliefs, practices, and knowledge of elementary teachers can facilitate them to develop more effective assessment practices. Likewise, with this context, school leaders and policymakers can make more informed decisions about assessments and support teachers to implement their positive attitudes about assessment.

This study adds an extra dimension to the theory and practice of teachers' assessment beliefs, practices, and knowledge. In particular, studies on assessment literacy to date have been limited – not only in Asia but across the world, with an exception of America, as indicated in the literature review. Similarly, studies on assessment beliefs and practices are quite common in the Western context but are rare in Asia. Therefore, this research project represents an addition to the scarcity of empirical studies on teachers' assessment beliefs, practices, and literacy. To the knowledge of the researcher, there has been no other study conducted to explore the statistical relationships between these three constructs.

Despite being significant, this study is not without limitations. The first limitation is related to the data collection instrument. The Classroom Assessment Literacy Inventory (CALI) that measured teachers' assessment literacy does not have credible reliability and validity. In the

attempt to adapting the instrument to the context of Bhutan, 14 items were deleted from its original version, leaving 21 items remaining. As a result, the number of items to measure each of the Seven Standards was 3, which was found to be too few to adequately describe the standards. The CALI measured the cognitive dimension of assessment, and not the aptitude, which was found challenging by the teachers. Consequently, 32 teachers out of 112 participants did not attempt this section of the survey.

The second limitation derives from the sample size. The sample size (n=112) was small in proportion to the population (2,296) of teachers teaching in primary education, and proportion to the number of items (69 items) included in the survey. The bigger the sample size, the better the measurement of included variables in a study, and the more that the results can be generalised. Therefore, generalisation and credibility are left at the discretion of the readers. Additionally, the ratio of female and male respondents was a limitation. 72% of the respondents were female and 28% were male, resulting in a difference of 44%. This must have presented biased results in terms of assessment practices affected by gender.

The third limitation is related to the nature of responses that sought to measure assessment beliefs and assessment practices. Since the responses are self-reported and at the discretion of participants alone, the accuracy of these self-reported responses is questionable.

Owing to the scope and limitations of this study, future researchers initially may wish to study teachers' assessment beliefs, assessment practices, and assessment literacy levels separately. This way, the number of items or questions to measure variables will be reduced, which will encourage respondents to participate in the study. Besides, measuring limited variables would simplify the process of operationalising variables in the data analysis. Further, authentic and true responses on assessment beliefs and practices may be achieved through a mixed-method approach of research, rather than through a quantitative method or qualitative method alone. More empirical studies are also required to further investigate the relationship between assessment beliefs, assessment practices and assessment literacy level.

5.9. CONCLUSION

The purpose of this study was to investigate Bhutanese elementary teachers' assessment beliefs, practices, and literacy level, to gain better understandings of teachers' beliefs and practices related to formative assessment. Overall, the teachers showed positive beliefs about assessment and agreed that assessment is for improving learning and teaching. However, teachers did not disagree that assessment is irrelevant rather they were neutral about this conception. The highest positive correlation was shown between assessment for improvement and assessment for student accountability. At the same time, there was positive association between improvement and irrelevance. Likewise, the value that teachers associated with numerous assessment practices are promising. Teachers expressed that they value most the formative assessment methods such as giving prompt feedback, peer and self-assessment, and questioning/dialogue. Relatively high values were found for alternative assessment methods such as authentic assessments and oral presentations, indicating that teachers give importance to this type of assessment. Teachers also valued both self-designed and published assessments. The level of teachers' assessment literacy, as measured by the Classroom Assessment Literacy Inventory (Mertler, 2003), was found to be significantly low. The Bhutanese elementary teachers are more literate in Standard 6 (communicate assessment results) and less literate in Standard 3 (the ability to administer scores). The assessment practices as the overall outcome variable was significantly influenced by assessment literacy, assessment beliefs, assessment education/training received and gender.

REFERENCES

Abbot, M.L. (2011). SPSS BASICS (Chapter 4). In Martin. L.A. *Understanding Educational Statistics Using Microsoft Excel and SPSS,* PP. 23-39 John Wiley & Sons, Incorporated

American Federation of Teachers (AFT), National Council on Measurement in Education (NCME), and National Education Association (NEA) (1990). *Standards for teacher competence in educational assessment of students*. Retrieved from http://www.unl.edu.buros/article3.html

Andrade, H., Lui A., Palma, M., & Hefferen, J. (2015) Formative assessment in dance education. *Journal of Dance Education, 15*(2), 47–59. http://dx.doi.org/10.1080/15290824.2015.1004408

Arbuckle, J. L. (2007). *AMOS (Version 16.0.1) [CFA and SEM analysis program]*. Spring House, PA: Amos Development Corporation.

Azis, A. (2015). Conceptions and practices of assessment: a case of teachers representing improvement conception. *TEFLIN Journal, 26 (2), 129-154.*

Barnes, N., Fives, H., & Dacey, C. M. (2015). Teachers' beliefs about assessment. In H. Fives, & M. G. Gill (Eds.), *The handbook of research on teachers' beliefs* (pp. 229-233). New York, NY: Routledge.

Barnidge, M. & De Zúñiga. (2017). Amos (Software). In J, Matthes., (General Editor), C.S. Davis and R. F. Potter (Associate Editors). *The international encyclopedia of*

communication research methods. John Wiley & Sons, Inc. DOI: 10.1002/9781118901731.iecrm0003

BCSEA. (2015). *BCSEA Journal of Educational Assessment*. Assessment and Monitoring Division. Thimphu.

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Principles, Policy & Practice, 18,* 5–25. doi: 10.1080/0969594x.2010.513678

Bentler, P. M. (2007). On tests and indices for evaluating structural models. *Personality and Individual Differences, 42*, 825-829.

Berry, R. & Kennedy, K.J. (2008). Basic Concepts of Assessment. In R.Berry & K.J. Kennedy. *Assessment for learning* (pp.5-22), Hong Kong University press. 5-22

Birenbaum, M., Breuer, K., Cascallar, E., Dochy, F., Dori, Y., Ridgway, J., . . . Nickmans, G. (2006). A learning integrated assessment system. *Educational Research Review, 1,* 61–67. doi:10.1016/j.edurev.2006.01 .001

Black, P. & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in education*, *25*(3). https://doi.org/10.1080/0969594X.2018.1441807

Black, P. & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy, & Practice, 5*(1), 7-74. doi:10.1080/0969595980050102

Black, P. & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 80(*2), 139-148.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Principles, Policy & Practice, 5,* 7–74. doi:10.1080/096959598005010

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21,* 5–31. doi:10.1007/s11092-008-9068-5

Black, P., Harrison, C., Hodgen, J., Marshall, M. & Serret, N. (2010) Validity in teachers' summative assessments. *Assessment in Education,* 17(2), 215-232.

Black, P., Harrison, C., Hodgen, J., Marshall, M., & Serret, N. (2011). Can teachers' summative assessments produce dependable results and also enhance classroom learning? *Assessment in Education: Principles, Policy & Practice, 18*(4), 451–469.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: Putting it into practice*. Buckingham: Open University Press.

Briggs,D.C.,Ruiz-Primo,M.A.,Furtak,E.,Shepard,L.,&Yin,Y. (2012). Meta-analytic methodology and inferences about the efficacy of formative assessment. *Educational Measurement: Issues and Practice, 31*(4), 13–17. http://dx.doi.org/10.1111/j.1745-3992.2012.00251.x

Brown, G. T. L. (2006). Teachers' conceptions of assessment: Validation of an abridged instrument. *Psychological Reports, 99*(1), 166-170.

Brown, G. T. L., (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education, 11*(3), 301-318.

Brown, G. T. L., Chaudhry, H., & Dhamija, R. (2015). The impact of an assessment policy upon teachers' self-reported assessment beliefs and practices: A quasi-experimental study of Indian teachers in private schools. *International Journal of Educational Research*, *71*, 50–64. https://doi.org/10.1016/j.ijer.2015.03.001.

Brown, G. T. L., & Harris, L. R. (Eds.). (2016). *Handbook of social and human conditions in assessment*. New York, NY: Routledge.

Brown, G. T., Hui, S. K. F., & Yu, F. W. M. (2010, July). *Teachers' conceptions of assessment: Developing a model for teachers in Hong Kong*. Paper presented at the Biannual conference of the International Test Commission, Hong Kong, China.

Brown, G., Lake, R., & Matters, G. (2011). Queensland teachers' conceptions of assessment: The impact of policy priorities on teacher attitudes. *Teaching and Teacher Education*, *27*(1), 210–220. https://doi.org/10.1016/j.tate.2010.08.003

Brown, G.T.L & Gao, L. (2015) Chinese teachers' conceptions of assessment for and of learning: Six competing and complementary purposes, *Cogent Education*, *2*(1), 1- 19. DOI: 10.1080/2331186X.2014.993836

Brown, G.T.L. (2008). *Conceptions of Assessment: Understanding what assessment means to teachers and students.* Nova Science Publishers Inc.

Brown, G.T.L. (2011). Teachers' conceptions of assessment: Comparing primary and secondary teachers in New Zealand. *Assessment Matters, 3*(2011). 45-70.

Brown, G.T.L., Kennedy, K.J., Fok, P. K., Chan, J.K.S. & Yu, W.M. (2009) Assessment for student improvement: understanding Hong Kong teachers' conceptions and

practices of assessment, *Assessment in Education: Principles, Policy & Practice, 16*(3), 347-363, DOI: 10.1080/09695940903319737

Calveric, S. B. (2010). *Elementary teachers' assessment beliefs and practices* (Doctoral Thesis Virginia Commonwealth University) Retrieved from http://proxy.library.adelaide.edu.au/login?url=https://search-proquest-com.proxy.library.adelaide.edu.au/docview/855823398?accountid=8203

CAPSD. (2008). *Syllabus for classes IV to VI*. Curriculum and Professional Support Division. Paro

CAPSD. (1999). *Continuous Assessment*. Curriculum and Professional Support Division. Thimphu.

Cauley, K. M., & McMillan, J. H. (2010). Formative assessment techniques to support student motivation and achievement. *Clearing House, 83*(1), 1-6.

Centre for Educational Research and Innovation. (2008). *Assessment for learning formative assessment.* Paris, France: OECD.

Churchill, R., Ferguson, P., Godinho, S., Johnson, N. F., Keddie, A., Letts, W., Mackay, J., McGill, M., Moss, J., Nagel, M. C., Nicholson, P., and Vick, M. (2011). *Teaching: Making a difference.* Milton Qld: John Wiley & Sons Australia, Ltd.

Colbert, P. & Cumming, J. J. (2014). Enabling All Students to Learn Through Assessment : A Case Study of Equitable Outcomes Achieved Through the Use of Criteria and Standards. In C. Wyatt-Smith. *Designing Assessment for Quality learning* (pp.211-232). Springer Science + Business Media Dordrecht.

Creswell, J.W. (2014). *Research Design; Quantitative, Qualitative and Mixed Methods Approaches.* SAGE Publications, Inc.

Cumming, J. J. & Kleij, F.M.V. (2016). Effective Enactment of Assessment for Learning and Student Diversity in Australia. In D. Laveault and L. Allal (eds.), *Assessment for Learning: Meeting the Challenge of Implementation, The Enabling Power of Assessment* (pp. 55-73). Springer International Publishing Switzerland. DOI 10.1007/978-3-319-39211-0_4

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), 1–15. https://doi.org/10.1080/2331186X.2016.1225380

Dunn, K. E. & Mulvenon, S. W. (2009). A critical review of research on formative assessment: The limited scientific evidence of the impact of formative assessment in education. *Practical Assessment, Research & Evaluation*, *14*(7), 1-11.

Engelsen, K.S. & Smith, K. (2014). Assessment literacy**.** In C. Wyatt-Smith. *Designing Assessment for Quality learning* (pp, 91-107). Springer Science + Business Media Dordrecht.

Fives, H., & Buehl, M. (2012). Spring Cleaning for the "messy" construct of teachers' beliefs: What are they? Which have been examined? What can they tell us? In K. R. Harris, S. Graham, & T. Urdan (Eds.), *APA educational psychology handbook: Volume 2 individual differences and cultural and contextual factors* (pp. 471-499). Washington: American Psychological Association.

Fowler, F.J. (2009). *Survey Research Methods* (4th edition). Thousand Oaks, CA: Sage.

Gebril, A. & Brown. G.T.L. (2014). The effect of high-stakes examination systems on teacher beliefs: Egyptian teachers' conceptions of assessment. *Assessment in Education: Principles, Policy & Practice, 21* (1), 1-19.

Griffiths, T., Gore, J., & Ladwig, J. (2006, November). *Teachers' fundamental beliefs, commitment to reform, and the quality of pedagogy*. Paper presented at the Australian Association for Research in Education (AARE).Adelaide, Australia.

Hailaya, W.M. (2014). Teacher assessment literacy and student outcomes in the province of Tawi-Tawi, Philippines. (Doctoral Thesis, University of Adelaide, Australia). Retrieved from http://digital.library.adelaide.edu.au/dspace/handle/2440/56314

Hattie,J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to student achievement.* New York: Routledge.

IBM Corp. (2019). *IBM SPSS Statistics for Windows, Version 26.0*. Armonk, NY: IBM Corp

Keller, H. D., & Utha, K. (2017). Pedagogical discourses in Bhutanese upper secondary schools. *Kowledge Culture*. n. pag. Print.

Kennedy, J. K., Chan, J. K. S., Fok, P. K. & Yu, W. M. (2008). Forms of assessment and their potential for enhancing learning: Conceptual and cultural Issues. *Educational Research Policy Practice*, *7*, 197- 207.

Kingston, N. & Nash, B. (2011). Formative Assessment: A Meta-Analysis and a Call for Research. *Educational Measurement: Issues and Practice, 30(4), 28–37.*

Kippers, W.B., Wolterinck, C.H.D., Schildkamp, K., & Poortman, C.L. (2018). Teachers' views on the use of assessment for learning and data-based decision making in classroom practice. *Teaching and Teacher Education, 75*, 199-213.

Kline, P. (1994). *As easy guide to factor analysis*. NY: Routledge.

Kline, R. B. (2016). *Principles and practice of structural equation modeling*. New York: The Guilford Press.

Koh, K., Burke, L.E.C.A., Luke, A., Gong, W. & Tan, C. (2018). Developing the assessment literacy of teachers in Chinese language classrooms: A focus on assessment task design. *Language Teaching Research*, (22)3, 264-288. DOI:10.1177/1362168816684366

Leavy, P. (2017). *Quantitaive, Qualitative, Mixed Methods, Arts-Based, and Community-Based Participatory Research Approches*. The Guilford Press. New York.

Leighton, J. P., Gokiert, R. J., Cor, M. K., & Heffernan, C. (2010). Teacher beliefs about the cognitive diagnostic information of classroom- versus large-scale tests: Implications for assessment literacy. *Assessment in Education: Principles, Policy, & Practice*, 17(1), 7-21. doi: 10.1080/09695940903565362

Li, Z. & Qiu, Z. (2018). How does family background affect children's educational achievement? Evidence from Contemporary China. *The Journal of Chinese Sociology, 5*(13), 1-21. https://doi.org/10.1186/s40711-018-0083-8

Looney, J. W. (2011). *Integrating formative and summative assessment: Progress toward a seamless system?* OECD Education Working Papers, No. 58, OECD Publishing. http://dx.doi. org/10.1787/5kghx3kbl734-en

Luyten, H. & Dolkar, D. (2010) School-based assessments in high- stakes examinations in Bhutan: a question of trust? Exploring inconsistencies between external exam scores, school-based assessments, detailed teacher ratings, and student self-ratings. *Educational Research and Evaluation, 16*(5), 421-435, DOI: 10.1080/13803611.2010.530437

Masters, G.N. (2013). *Reforming Educational Assessment: Imperatives, principles and challenges*. Australian Council for Educational Research (ACER), Melbourne.

Matsunaga, M. (2010). How to Factor-Analyze your data right: Do's, don'ts, and how- to's. *International Journal of Psychological Research, 3*(1), 98-111.

Matzke, D. (2014). Model Comparison and the principle of parsimony. In J. Busemeyer, J. Townsend, Z. J. Wang, & A. Eidels (Eds.), Oxford handbook of computational and mathematical psychology (pp, 125-144). Oxford, UK: Oxford University Press.

Maxwell, T. W., Rinchen, P., & Cooksey, R. (2010). Evolutionary trajectories in school assessment systems: The case of Bhutan. *Asia Pacific Journal of Education*, *30*(3), 273–288. https://doi.org/10.1080/02188791.2010.495840

McManus, S. (Ed.) (2008*). Attributes of effective formative assessment.* Washington, DC: Council of Chief State School Officers.

McMillan, J. H., & Schumacher, S. (2006). *Research in education: Evidence-based inquiry* (6th ed.). Boston, MA: Allyn and Bacon.

McMillan, J., Myran, S., & Workman, D. (2002). Elementary teachers' classroom assessment and grading practices. *The Journal of Educational Research, (95)*4, 203-213

Mellati, M., & Khademi, M. (2018). Exploring teachers' assessment literacy: Impact on learners' writing achievements and implications for teacher development. *Australian Journal of Teacher Education*, *43*(6), 1–18. https://doi.org/10.14221/ajte.2018v43n6.1

Mertler, C. A. (2005). Secondary teachers' assessment literacy: Does classroom experience make a difference? *American Secondary Education, 33*(2), 76-92.

Mertler, C. A. (2009). Teachers' assessment knowledge and their perceptions of the impact of classroom assessment professional development. *Improving Schools, 12*(2), 101-113.

Mertler, C. A., (2003). *Pre-service versus in-service teachers' assessment literacy: Does classroom experience make a difference?* Columbus: Mid-Western Educational Research Association.

Mertler, C.A. (2019). Quantitative Research Methods. In C.A. Mertler (2nd Eds), *Introduction to Educational Research* (pp 107-143). Sage Publication, Inc.

MOE. (2014). *Bhutan Education Blueprint 2014-2024.* Ministry of Education, Thimphu.

MOE. (2019). *Annual Education Statistics (2019*). Policy and Planning Division, Thimphu.

Mugizi, W. (2019). Constructing a Conceptual Framework for Quantitative Data Analysis in Social Science Research. *Interdiscilinary Journal of Education, 2*(1), 74-88.

Ning, N. T., Kam, H. W., Sundara, R., Chin, Y. K., Hoe, L. N., & Thiruman, M. (2010). *Teacher Quality Enhancement Project Policy Guidelines and Strategic Plan Report prepared by*. *Royal Education Council*. Retrieved from www.educare.swww.rec.org.bt

OECD (n.d). *Assessment for learning; Formative assessment*. Retrieved from https://www.oecd.org

OECD. (2006). *Education Policy Analysis: Focus Policy Analisys*. *Higher Education*. https://doi.org/10.1787/epa-2006-en

Ogan-Bekiroglu, F. & Suzuk, E. (2014) Pre-service teachers' assessment literacy and its implementation into practice. *Curriculum Journal, 25*(3), 344-371, DOI: 10.1080/09585176.2014.899916

Palden, T. (2018). Govt. to do away exams for primary schools in phases. *Kuensel*.

Panadero, E., and A. Jonsson. (2013). The Use of Scoring Rubrics for Formative Assessment Purposes Revisited: A Review. *Educational Research Review 9*, 129–144.10.1016/j.edurev.2013.01.002

Parveen, H. & Showkat, N. (2017). *Quantitative Methods: Survey*. Retrieved from https://www.researchgate.net/publication/318959206

Perie, M., Marion, S., & Gong, B. (2009). Moving toward a comprehensive assessment system: A framework for considering interim assessments. *Educational Measurement: Issues and Practice, 28,* 5–13. doi:10.1111/ j.1745-3992.2009.00149.x

Perry, L.M. (2013). Teacher and Principal Assessment Literacy. *Graduate Student Theses, Dissertations, & Professional Papers*. 1391. https://scholarworks.umt.edu/etd/1391

Plake, B. S., Impara, J. C. & Fager, J. J. (1993). Assessment competencies of teachers: A national survey. *Educational Measurement: Issues & Practice, 12*(4), 10-12. doi: 10.1111/j.1745-3992.1993.tb00548.x

Popham, W. J. (2009) Assessment Literacy for Teachers: Faddish or Fundamental? *Theory into Practice, 48(*1), 4-11, DOI: 10.1080/00405840802577536

Popham, W. J. (2010). *Classroom assessment: What teachers need to know (6th ed.*). Boston, MA: Allyn and Bacon.

Postareff, L., Virtanen, V., Katajavuori, N., & Lindblom-Ylanne, S. (2012). Academics' conceptions of assessment and their assessment practices. *Studies in Educational Evaluation, 38*(4), 84-92.

Rahman, S. (2016). The Advantages and Disadvantages of Using Qualitative and Quantitative Approaches and Methods in Language "Testing and Assessment" Research: A Literature Review. *Journal of Education and Learning, 6(*1), 102-112.

Remesal, A. (2007). Educational reform and primary and secondary teachers' conceptions of assessment: The Spanish instance, building upon Black and Wiliam (2005). *The Curriculum Journal, 18*(1), 27-38.

Remesal, A. (2011). Primary and Secondary teachers' conceptions of assessment: A qualitative study. *Teaching and Teacher Education, 27*(2011), 472-482.

RGOB. (2019). *Draft National Education Assessment Framework 2019*. Royal Government of Bhutan. Thimphu.

Rosas, R. (2014). *Elementary school teachers' and principals' formative assessment beliefs, practices and assessment literacy*. (Doctoral Thesis, California State University) Retrieved from http://proxy.library.adelaide.edu.au/login?url=https://search-proquest-com.proxy.library.adelaide.edu.au/docview/855823398?accountid=8203

Rowntree, D. (1987). *Assessing students: How shall we know them*? London: Kogan

Royal Education Council (REC). (2019). *Continuous Formative Assessment Guidebook for Primary Education (Classes PP to VI)*. Paro. Bhutan.

Ryan, K.A. (2018). *An Investigation Of Pre-Service Teacher Assessment Literacy And Assessment Confidence: Measure Development And EDTPA Performance*. (Doctoral thesis, Kent University College of Education, Health and Human Services). Retrieved fromhttps://etd.ohiolink.edu/!etd.send_file?accession=kent1522746692960315&dis position=inline

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*, 119-144. https://doi.org/10.1007/BF00117714.

Sadler, D. R. (1998). Formative assessment: Revisiting the territory. *Assessment in Education: Principles, Policy & Practice, 5,* 77–85.

Safii, L., & Wong, H. M. (2017, June). *Seeing self-assessment and teacher feedback through students' lenses.* Paper presented at the Seventh Redesigning Pedagogy International Conference, Singapore

Schools will implement the Continuous Formative Assessment. (2020, January 23). *The Kuensel*, p.1

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting Structural Equation Modeling and Confirmatory Factor Analysis results: A review. *The Journal of Educational Research, 99*(6), 323-338.

Shavelson, R. J., Young, D. B., Ayala, C. C., Brandon, P. R., Furtak, E. M., Ruiz-Primo, M. A., . . . Yin, Y. (2008). On the impact of curriculum- embedded formative assessment on learning: A collaboration between curriculum and assessment developers. *Applied Measurement in Education, 21,* 295–314. doi:10.1080/08957340802347647

Shepard, L. A., Hammerness, K., Darling-Hammond, L., & Rust, F. (with Baratz-Snowden, J., Gordon, E., Gutierrez, C., & Pacheco, A.). (2005). Assessment. In L. Darling-Hammond & J. Bransford (Eds.), *Preparing teachers for a changing world: What teachers should learn and be able to do* (pp. 275–326). San Francisco, CA: Jossey-Bass.

Siegel, M. A., & Wissehr, C. (2011). Preparing for the plunge: Pre-service teachers' assessment literacy. *Journal of Science Teacher Education, 22*, 371–391.

Song, E. & Koh, K. (n.d). *Assessment for learning: understanding teachers' beliefs and practices.* Retrieved from https://www.academia.edu

Stigging, R. & DuFour, R. (2009). Maximising the Power of Formative Assessment. *Phi Delta Kappan.* May (2009). 640-644.

Stiggins, R. (2014). Improve assessment literacy outside of schools too. *Phi Delta Kappan*, *96*(2), 67–72. https://doi.org/10.1177/0031721714553413

Stiggins, R. (2014). Improve Assessment Literacy outside of Schools too. *Kappan 96* (2), 67-72.

Stiggins, R. J. (1995). Assessment literacy for the 21st century. Phi Delta Kappan, 77(3), 238–245.

Stiggins, R. J. (1999). Assessment , Studen Confidence , and Scho. *The Phi Delta Kappan*, *81*(3), 191–198.

Stiggins, R. J. (1999b). Are You Assessment Literate? *High School Journal, 6*(5), 20-23

Stiggins, R. J.(2018). Better assessments require better assessment literacy. *Educational Leadership*, *75*(5), 18–19.

Swaffield, S. (2011). Getting to the heart of authentic Assessment for Learning, *Assessment in Education: Principles, Policy and Practice*, 18(4) 433–449.

Tabachnick, B.G., & Fidell, L.S. (2001). Using Multivariate Statistics, 4. Needham Heights, MA: Allyn & Bacon.

Thinley, P. (2016). Overview and 'Heart Essence' of the Bhutanese Education System. In M.J.Schuelka & T.w.Maxwell, *Education in Bhutan* (pp. 139–152). Springer Science+Business Media Singapore https://doi.org/10.1007/978-981-10-1649-3

Utha, K. (2014). *Formative Assessment practices in Bhutanese Secondary Schools and its impact on Quality of Education*. https://doi.org/10.5278/vbn.phd.socsci.00011

Van der Kleij, F. M., Vermeulen, J. A., Schildkamp, K., & Eggen, T. J. H. M. (2015). Integrating data-based decision making, Assessment for Learning, and diagnostic testing in formative assessment. *Assessment in Education: Principles. Policy & Practice, 22*(3), 324-343.

Van Staden, S., & Motsamai, P. (2017). Differences in the quality of School-based assessment: Evidence in Grade 9 mathematics achievement. *Pythagoras, 38*(1), 1-10.

Webb, N. L. (2002). *Assessment literacy in a standard-based urban education setting*. Paper presented to the American Educational Research Association Annual Meeting in New Orleans, Louisiana.

Wiliam, D. (2011). *Embedded formative assessment*. United State: Solution Tree

Wiliam, D., & Thompson, M. (2007). Integrating assessment with instruction: What will it take to make it work? In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 53–82). Mahwah, NJ: Erlbaum.

Wiliam, D., Lee, C., Harrison, C., & Black, P. J. (2004). Teachers developing assessment for learning: impact on student achievement. *Assessment in Education: Principles Policy and Practice, 11*(1), 49-65.

Williams, C. (2011). Research methods. *Journal of Business & Economics Research (JBER)*, *5*(3).

Wood, D.F. (2019). Formative assessment: assessment for learning. In Swanwick, T., Forrest, K. & O'Brien, B.C. *Understanding Medical Education: Evidence, Theory, and Practice*, (3rd Eds).(PP. 361-373). The Association for the Study of Medical Education (ASME). Published 2019 by John Wiley & Sons Ltd.

Yamtim, V. & Wongwanich, S. (2014). A study of classroom assessment literacy of primary school teachers. *Procedia - Social and Behavioural Sciences, 116* (2014), 2998 – 3004

APPENDICES

APENDIX 1 ETHICS APPROVAL

Our reference 34511

RESEARCH SERVICES

OFFICE OF RESEARCH ETHICS, COMPLIANCE AND INTEGRITY
THE UNIVERSITY OF ADELAIDE

LEVEL 4, RUNDLE MALL PLAZA 50 RUNDLE MALL
ADELAIDE SA 5000 AUSTRALIA

13 July 2020

Dr Lynda Maree MacLeod School of Education

Dear Dr MacLeod

ETHICS APPROVAL No: PROJECT TITLE:

TELEPHONE +61 8 8313 5137 FACSIMILE +61 8 8313 3700 EMAIL hrec@adelaide.edu.au

CRICOS Provider Number 00123M

H-2020-121
Bhutanese elementary teachers' beliefs, practices, knowledge and literacy of formative assessment

The ethics application for the above project has been reviewed by the Low Risk Human Research Ethics Review Group (Faculty of Arts and Faculty of the Professions) and is deemed to meet the requirements of the *National Statement on Ethical Conduct in Human Research 2007 (Updated 2018)* involving no more than low risk for research participants.

You are authorised to commence your research on: 13/07/2020 The ethics expiry date for this project is: 31/07/2023

NAMED INVESTIGATORS:

Chief Investigator: Dr Lynda Maree MacLeod

Student - Postgraduate Masters Mrs Hemlata Karki by Research:

CONDITIONS OF APPROVAL: Thank you for addressing the feedback. The revised ethics application provided on the 13th of July 2020 has been approved.

Ethics approval is granted for three years and is subject to satisfactory annual reporting. The form titled Annual Report on Project Status is to be used when reporting annual progress and project completion and can be downloaded at http://www.adelaide.edu.au/research-services/oreci/human/reporting/. Prior to expiry, ethics approval may be extended for a further period.

Participants in the study are to be given a copy of the information sheet and the signed consent form to retain. It is also a condition of approval that you immediately report anything which might warrant review of ethical approval including:

serious or unexpected adverse effects on participants,
previously unforeseen events which might affect continued ethical acceptability of the project,
proposed changes to the protocol or project investigators; and
the project is discontinued before the expected date of completion.

▪▪▪▪

Yours sincerely,

Dr Anna Olijnyk Convenor

Dr Jungho Suh Convenor

The University of Adelaide

APPENDIX 2 APPROVAL LETTER FROM THE MINISTRY OF EDUCATION

དཔལ་ལྡན་འབྲུག་གཞུང་། ཤེས་རིག་ལྷན་ཁག།

Ministry of Education
Department of School Education
School Planning and Coordination Division

DSE/SPCD/SLCU(2.2)/2020/471 March 10, 2020

Principal
Participating School(s)
All Dzongkhags and Thromdes

Subject: Approval to conduct and collect data for research project

Dear Sir/Madam,

Ms. Hemlata Karki, a teacher of Dagapela Middle Secondary School is currently pursuing a Master of Education at the University of Adelaide in South Australia. To fulfil the requirement of the graduate program, she is working on the research project entitled: ***"Bhutanese Elementary Teachers' Beliefs and Practice of Formative Assessment and the associated Challenges".***

This proposed study attempts to explore: how beliefs shape the practice of formative assessment in Bhutanese teachers, the level of assessment literacy among the Bhutanese teachers, gaps in assessment policy and practice. Therefore, she will be collecting data (using online survey questionnaires and online interviews) from the teachers of your school.

In this regard, you are requested to facilitate and grant permission to Ms. Hemlata Karki to collect data under the following conditions:

- ***Seeking prior permission from the school management before collection of data.***
- ***Ensuring minimal disruption to instructional time of the school.***
- ***Providing research participants with sufficient information to make an informed decision as to whether to take part in research (informed consent).***
- ***Ensuring that participants are not subject to coercion to take part or penalty for not taking part.***
- ***Ensuring that participants are, and are aware that they are, free to withdraw from the research at any time without giving a reason and without a prejudice.***
- ***Protecting and respecting personal data provided by participants through rigorous and appropriate procedures for confidentiality and anonymisation.***

Thanking you.

Sincerely yours,

(Karma Tshering)
DIRECTOR GENERAL

CC: *1. Chief DEO/TEO, for information. 2. Ms. Hemlata Karki, DMSS, for follow-up.*

Post Box No. 112, Kawajangsa, Thimphu, Bhutan, Tel: PA: +975 2 325325, www.education.gov.bt

APPENDIX 3 PARTICIPATION INFORMATION SHEET

PROJECT TITLE: Bhutanese Elementary teachers' beliefs, practices and assessment literacy of formative assessment.

HUMAN RESEARCH ETHICS COMMITTEE APPROVAL NUMBER: H-2020-121
PRINCIPAL INVESTIGATOR: Lynda Maree MacLeod Dr.
STUDENT RESEARCHER: Hemlata Karki, Ms
STUDENT'S DEGREE: Master of Education

Dear Participants

You are invited to participate in the research project as described below.

What is the project about?

The project intends to investigate Bhutanese elementary teachers' beliefs, practices, knowledge and assessment literacy of formative assessment.

In order to conduct this research, about 200 Bhutanese elementary teachers teaching from Pre-Primary to Grade 6 from about 25 schools are invited.

Who is undertaking the project?

The research is undertaken by Hemlata Karki, studying a Master of Education degree at the University of Adelaide, South Australia under the supervision of Dr Lynda Maree MacLeod.

Why am I being invited to participate?

You are invited as you are a Bhutanese elementary teacher teaching any subject or grade from Pre-Primary to Grade 6. Your participation will be valuable source of information for this study as it intends to investigate Bhutanese elementary teachers' beliefs, practices, knowledge and assessment literacy of formative assessment. This is significant as written examinations from PP to grade 6 in Bhutan are being removed by next year and formative assessment is the basis for students' promotion to next grade. Further, data-driven decisions and high-stakes assessments continue to grow. I am hopeful that with your help, we can identify areas in need of best practices for teachers in Bhutan and enhance students' learning.

What am I being invited to do?
You are being invited to take *an online survey questionnaire which is sent to you with this introductory message via the survey link.* The online survey consists of four sections under Demographic information, Teacher Conception of Assessment (beliefs), Self-reported assessment practices (practices), and Classroom assessment literacy inventory (assessment literacy).

Completion and submission of the questionnaires to the researcher will be taken as an indication of giving consent to participate, giving permission to use the data and confirmation that you have read the Participation Information Sheet prior to filling out the survey.

How much time will the project take?
It will take approximately 45 to 60 minutes but you have the option to pause and continue by saving the work as you complete.

Are there any risks associated with participating in this project?
There are no foreseeable risks in this research. However, in the process of filing out the survey if you experience that you need to learn more about assessment/formative assessment, please feel free to contact me. I am ready to share the resources with you for your learning and practicing. Or if you wish to share your feelings and reflection, please talk to your colleague whose name is mentioned in the Introductory Message.

What are the benefits of the research project?
Investigating elementary teachers' beliefs, practices, knowledge and assessment literacy, this study may give an insight of what Bhutanese elementary teachers believe about formative assessment and how do these beliefs impact the practice of formative assessment in classrooms. Further, assessment literacy level of teachers will inform the teachers themselves about their competencies, knowledge and practice of formative assessment and reflect over them. The list of recommendations in the research report will potentially inform the policy makers to formulate and amend policies towards enhancing teachers' assessment literacy and competencies. It will also inform the school leaders to promote the practice of formative assessment through sustained professional development programs at school level. Finally, as the data-driven decisions and high-stake assessments are increasingly growing, this research may identify and suggest the areas in need of best practices of formative assessment to enhance students' learning in Bhutan.

Can I withdraw from the project?
Participation in this project is entirely voluntary. If you do not wish to participate, you can simply not complete the survey or withdraw prior to the submission of the survey. However, withdrawing your participation after the submission of the survey is not possible as the data will be completely anonymous and tracing back your data information will not be possible. Please be assured that your participation, non-participation or withdrawal will not impact your ongoing career, promotion or incentives. More so, no negative consequences will follow from non-participation or withdrawal.

What will happen to my information?
All the information provided will be treated in the strictest confidentiality during the research process (i.e. Recruitment, data collection, data analysis) and during the reporting of research results and publications. Your responses in the survey questionnaire will not be shared to anybody including your school authorities and other stakeholders. The survey questionnaire once submitted through Web based survey the identity of the respondents will automatically disappear. Therefore, do not worry about the data and information that you will be sharing in the form of responses to the survey questions/items. They are absolutely anonymous and confidential.
The data will be stored as Non-identifiable. Access to the data will be restricted only to the researcher and the principal supervisor. The data will be stored with password protection, all the dataset in the UoA R: drive and UofA Box according to Data Management Plan. The data will be retained by the

University of Adelaide, South Australia for a minimum of 5 years from the date of dissertation submission. The data collected may be used for future research in order to compare with a new data set or to examine the change in beliefs, practices and literacy about the formative assessment of teachers. The research outcomes will be made publicly accessible in the form of Master Dissertation upon any future publication and presentations in the conferences.

Who do I contact if I have questions about the project?

Participant with questions or inquiries regarding the project may contact the researcher or the supervisor:

Name, Title	**Telephone Number**	**Email**
Dr Lynda Maree MacLeod	+61 8313 5548	Lynda.macleod@adelaide.edu.au
Ms Hemlata Karki	+61 424682522	hemlatakarki@education.gov.bt

What if I have a complaint or any concerns?

The study has been approved by the Human Research Ethics Committee at the University of Adelaide (approval number…..). This research is conducted within the framework of NHMRC National Statement on Ethical Conduct in Human Research 2007 (Updated 2018). If you have questions or problems associated with the practical aspects of your participation in the project, or wish to raise a concern or complaint about the project, then you should consult the Principal Investigator. If you wish to speak with an independent person with your concerns or a complaint, the University's policy on research involving human participants, or your rights as a participant, you may please contact the Human Research Ethics Committee's Secretariat on

Phone: +61 8 8313 6028

Email: hrec@adelaide.edu.au.

Post: Level 4, Rundle Mall Plaza, 50 Rundle Mall, ADELAIDE SA 5000

Any complaint or concern will be treated in confidence and fully investigated. You will be informed of the outcome.

If I want to participate, what do I do?

If you agree to participate in this research, please complete the online questionnaire by clicking "Next" below.

Please note that submission of your survey responses is considered as your consent to participate and use of your data for the purpose of this research and that you have read and understood the above information.

Yours sincerely,

Dr Lynda Maree MacLeod

Hemlata Karki

APPENDIX 4 ONLINE SURVEY QUESTIONNAIRE

This research aims to investigate the beliefs, practice, knowledge and assessment literacy of Bhutanese elementary teachers about formative assessment. It would also provide valuable information to the policy makers, school leaders and teachers about the formative assessment and explain how each stake holder can contribute in enhancing formative assessment practice to increase students' academic performance. As the data-driven decisions and high-stake assessments are increasingly growing, this research may identify and suggest the areas in need of best practices of formative assessment to enhance students' learning.

Please be noted that submission of your survey response will be taken as an indication of your consent to participate, confirmation that you have read the participant information sheet, and permission to use your data in this research. However, you have the privilege to withdraw your participation from the survey before the submission. For your information, the data collected now may be used in related future research and your privacy, anonymity and confidentiality will be highly maintained throughout the period of research and in future. No information and data will be shared to anybody except the researcher and the primary supervisor.

This document has FOUR Sections as follow:

SECTION ONE: DEMOGRAPHIC INFORMATION

SECTION TWO: TEACHER CONCEPTION OF ASSESSMENT (BELIEFS)

SECTION THREE: PRACTICE OF ASSESSMENT (FORMATIVE ASSESSMENT)

SECTION FOUR: CLASSROOM ASSESSMENT LITERACY INVENTORY (ASSESSMENT LITERCAY)

SECTION ONE: DEMOGRAPHIC INFORMATION

Please Check appropriate box(es) for each item.

1. Gender

☐ Female

☐ Male

☐ Other

2. Years of Teaching Experience

☐ 0-5 years

☐ 6-10 years

☐ 11-15 years

☐ 15^{+} years

3. Highest Qualification Earned

☐ Primary Teacher Certificate (PTC)

☐ Bachelor's degree in education (B ed)

☐ Post Graduate Diploma in Education (PGDE)

☐ Master's degree

☐ Doctor Ph.D.

4. What grade level (s) do you currently teach? Please tick all applicable to you.

☐ Pre-Primary

☐ Grade 1

☐ Grade 2

☐ Grade 3

☐ Grade 4

☐ Grade 5

☐ Grade 6

5. What subject(s) do you teach currently? Please check all applicable to you.

☐ Dzongkha

☐ English

☐ Mathematics

☐ Science

☐ Social Studies

6. What education or training in assessment do you have? Please check all applicable to you.

☐ None

☐ Short workshops provided by school as a part of professional development (PD)

☐ Short workshop provided by Ministry/Royal Advisory Council (REC)

☐ Completed an undergraduate assessment course

☐ Completed a graduate assessment course

☐ Completed a post graduate assessment course

SECTION TWO: TEACHER CONCEPTION OF ASSESSMENT III ABRIDGED SURVEY (BELIEFS)

This section consists of 27 statements related to beliefs and understanding of all types of Assessments. Please choose the rating according to YOUR OWN understanding of assessment. The rating scale from the LEFT to RIGHT is:

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

1. According to YOUR opinion about assessment please give your rating indicating how much you agree or disagree or neutral to each of the 27 statements.
2. A gentle reminder that it is essential to answer each question for this survey to be useful.

Conception of Assessment	Strongly disagree	Disagree	Neutral	Agree	Strongly Agree
1. Assessment provides information on how well schools are doing.					
2. Assessment places students into categories.					
3. Assessment is a way to determine how much students have learned from teaching.					
4. Assessment provides feedback to students about their performance.					
5. Assessment is integrated with teaching practice					

6. Assessment results are trustworthy.					
7. Assessment forces teachers to teach in a way that is contradictory to their beliefs.					
8. Teachers conduct assessments but make little use of the results.					
9. Assessment results should be treated cautiously because of measurement error.					
10. Assessment is an accurate indicator of a school's quality.					
11. Assessment is assigning a grade or level to student work.					
12. Assessment establishes what students have learned.					
13. Assessment informs students of their learning needs.					
14. Assessment information modifies ongoing teaching of students.					
15. Assessment results are consistent.					
16. Assessment is unfair to students.					
17. Assessment results are filed and ignored.					
18. Teachers should take into account the error and imprecision in all assessment.					
19. Assessment is a good way to evaluate a school.					
20. Assessment determines if students meet qualifications.					

21. Assessment measures students' higher order thinking skills.					
22. Assessment helps students improve their learning.					
23. Assessment allows different students to get different instruction.					
24. Assessment results can be depended on.					
25. Assessment interferes with teaching.					
26. Assessment has little impact on teaching.					
27. Assessment is an imprecise process.					

SECTION THREE: ASSESSMENT PRACTICES IN ELEMANTRY GRADES

Please give a rating for each of the following 16 statements based on YOUR opinion about assessment practices. Use the following rating scale and choose the response that comes closest to describing each assessment's level of importance to you.

- Not at all Important
- Low Importance
- Moderately Important
- Very Important
- Extremely Important

You may notice that the ratings are ordered from "Not at all Important" on the LEFT to "Extremely Important" on the RIGHT.

Check one box for each item.

PART A

Assessment Practices	Not at all important	Low important	Moderately important	Very important	Extremely important
1. Assessments designed primarily by yourself					
2. Performance quizzes					
3. Objective assessments (e.g MCQ, short answer, matching)					
4. Essay type questions					
5.performance assessments (e.g structured teacher observations or ratings of performance such as speech or paper)					
6. projects completed by individual students					
7. Major exams					
8. Authentic assessments (e.g real world performance tasks)					
9. Projects completed by teams of students					
10. Assessments provided to teachers such as in manuals or instructional guides					
11. Oral presentations					
12. Prompt constructive feedbacks					
13.Self -assessment by students					
14.Peer assessments by students					
15.Questioning					

SECTION FOUR: CLASSROOM ASSESSMENT LITERACY INVENTORY (ASSESSMENT LITERACY)

This section consists of 21 MCQs that test your assessment literacy. Choose the best option according to your knowledge and understanding of assessment.

1. What is the most important consideration in choosing a method for assessing student achievement?
☐ The ease of scoring the assessment.
☐ The ease of preparing the assessment.
☐ The accuracy of assessing whether or not instructional objectives were attained.
☐ The acceptance by the school administration.

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

☐ Student scores from the test can be used for a large number of educational decisions.
☐ If a student retook the same test, he or she would get a similar score on each retake.
☐ The test score is a more valid measure than teacher judgments.
☐ The test score accurately reflects the content of what was taught.

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

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

4. Ms. Deki, the principal, was evaluating the teaching performance of Mr. Penjor, the fourth-grade teacher. One of the things Ms. Deki 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. Deki to make this decision?

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

5. 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?

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

6. Which of the following would most likely increase the reliability of Mrs.Dema's multiple-choice end-of-unit examination in physical science?

☐ Use a blueprint to develop the test questions.
☐ Change the test format to true-false questions.
☐ Add more items like those already on the test.
☐ Add an essay component

7. Ms. Shova 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?
☐ compare, contrast, criticize
☐ identify, specify, list
☐ order, match, select
☐ define, recall, restate

8. Several students in Ms. Vijoy's class received low scores on her end-of-unit test covering 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?

☐ Use the test provided in the "teacher's guide."
☐ Have the students take a test that has separate items for each step of the process.
☐ Look at the student's records and standardized test scores to see which topics the students had not performed well on previously.
☐ Give students story problems to complete and have them show their work.

9. 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?
☐ The student answered 90% of the items on this test correctly.
☐ The student knows 90% of the instructional content of the unit covered by this test.
☐ The student scored higher than 90% of all the students who took the test.
☐ The student scored 90% higher than the average student in the class.

10. Students in Mr. Jeewan'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?

☐ When the models are turned in, Mr. Jeewan identifies the most attractive models and gives them the highest grades, the next most attractive get a lower grade and so on.
☐ Mr. Jeewan asks other teachers in the building to rate each project on a 5-point scale based on their quality.
☐ Before the projects are turned in, Mr. Jeewan constructs a scoring key based on the critical features of the projects as identified by the highest performing students in the class.
☐ Before the projects are turned in, Mr. Jeewan 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.

11.At the close of the first month of school, Mrs. Karma gives her fifth grade students a test she developed in social studies. Her test is modelled 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 much lower than other students. Which of the following types of additional information which would be most helpful in interpreting the results of this test?

☐ The gender of the students.
☐ The age of the students.
☐ Reliability data for the standardized social studies test she used as the model.

☐ Reading comprehension scores for the students.

12. Ms. Chandra 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?

☐ The principal needs to report the results of this assessment to the state testing director.
☐ Ms. Chandra wants to give the students practice in taking tests early in the semester.
☐ Ms. Chandra wants to check for prerequisite knowledge in her students before she begins the unit on factoring.
☐ Ms. Chandra wants to measure growth in student achievement of these concepts, and scores on this test will serve as the students' knowledge baseline.

13. Which of the following choices typically provides the most reliable student- performance information that a teacher might consider when assigning a unit grade?

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

14. When a parent asks a teacher to explain the basis for his or her child's grade, the teacher should:

☐ explain that the grades are assigned fairly, based on the student's performance and other related factors.
☐ ask the parents what they think should be the basis for the child's grade.
☐ explain exactly how the grade was determined and show the parent samples of the student's work.
☐ indicate that the grading scale is imposed by the school board and the teachers have no control over grades.

15. Which of the following grading practices results in a grade that least reflects students' achievement?
☐ Mr. Leki requires students to turn in homework; however, he only grades the odd numbered items.
☐ Mrs. Tshering uses weekly quizzes and three major examinations to assign final grades in her class.
☐ Ms. Goma permits students to redo their assignments several times if they need more opportunities to meet her standards for grades.
☐ Miss Eden deducts 5 points from a student's test grade for disruptive behaviour.

16. During the most recent grading period, Ms. Karma 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 regarding how she assigned the grades?

☐ The grades probably reflect a bias against minority students that exists in most tests.
☐ Decisions like grade assignment should be based on more than one piece of information.
☐ The test was too narrow in curriculum focus.
☐ There is no significant criticism of this method providing the test covered the unit's content.

17. In a routine conference with Meena's parents, Mrs. Euden observed that Meena's scores on the state assessment program's quantitative reasoning tests indicate Meena is performing better in mathematics concepts than in mathematics computation. This probably means that

☐ Meena's score on the computation test was below average.
☐ Meena is an excellent student in mathematics concepts.
☐ the percentile bands for the mathematics concepts and computation tests do not overlap.
☐ the mathematics concepts test is a more valid measure of Meena's quantitative reasoning ability.

18. Mr. Kelly bases his students' grades mostly on graded homework and tests. Mr. Kencho 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 formal and informal assessment.
☐ performance and applied assessment.
☐ customized and tailored assessment.
☐ formative and summative assessment.

19. Mrs. Bida 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. Bida's action?

☐ The students can see the other students' graded tests, making it a violation of the students' right of privacy.
☐ The students have to wait until after school, so the action is unfair to students who have to leave immediately after school.
☐ Mrs. Bida 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.

☐ The students who were absent will have an unfair advantage, because her action allows the possibility for these students to cheat.

20. 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?

☐ Mr. Jackson darkened circles on the answer sheet at random. He assumed Deki, who was not a good student, would just guess at the answers, so this would be a fair way to obtain Deki's score on the test.

☐ Mr. Hasta filled in the answer sheet the way he thought Jerry, who was not feeling well, would have answered based on Jerry's typical in-class performance.

☐ Mr. Sonam turned in the answer sheet as it was, even though he thought Geeta, an average student, might have gotten a higher score had she finished the test.

☐ Mr. Leki read each question and darkened in the bubbles on the answer sheet that represented what he believed Zangmo, a slightly below average student, would select as the correct answers.

21. Mrs. Tshering was concerned that her students would not do well in annual examination in grade 6 (question paper comes from national board of examination) to be administered in respective school. 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?

☐ Instructed students in strategies on taking multiple choice tests, including how to use answer sheets.
☐ Gave students the items from an alternate form of the test.
☐ Planned instruction to focus on the concepts covered in the test.
☐ None of these actions are unethical.

APENNDIX 5: APPLICATION TO THE PRINCIPALS

[Date]

To
The Principal
[insert school name]
[insert district name]
Bhutan

Sub: Seeking permission to collect research data.

Dear Principal,

I, Hemlata Karki am a teacher at Dagapela MSS, currently studying the Master of Educational degree in the University of Adelaide in South Australia. As a Master student I am required to write a dissertation for which I have chosen Bhutan as a place of data collection. The research title is "**Bhutanese Elementary Teachers' Beliefs, Practice, Knowledge and Literacy of Formative Assessment".**

According to the letter number DSE/SPCD/SLCU(2.2)/20202/471 dated 10th March 2020, awarded by the School Planning and Coordination Division, Department of Education, MOE, I am to seek prior permission from the school management before the collection of data and follow ethical aspects in strictest sense. Therefore, I have prepared a separate Participation Information Sheet (PIS) for the participants (elementary teachers teaching any subject from PP to grade 6 in the current year) to inform about the research project. The Information Sheet ensures voluntary participation, right to withdraw any time, no coercion and penalty for not participating or withdrawing, maintaining of highest confidentiality, anonymity and privacy of the participants' information and data at all times and details of contact for any questions and queries. In addition, I ensure that there will be minimum disturbance to the teachers as the survey will take 45 to 60 minutes to complete and it is an online survey questionnaire which can be completed any time in their own personal space. A week time will be given to complete and submit the survey to the researcher. Thus, it will not hamper teachers' daily job in the school.

Therefore, I would like to request you to kindly provide me with the permission letter. I would further like to request you to kindly email my Introductory Message attached here, to all the primary teachers in your school. For your kind information, the Introductory Message includes the introduction of the student researcher, topic of the research project, aims, a brief description of the research project, link to the survey and the steps to follow to get access to the online survey questionnaire. The survey link will take the teacher participants to the Participation Information Sheet and the Survey Questionnaires.

Your kind cooperation, support and permission are highly appreciated. Please, find attached herewith the letter of approval from Department of Education and the Introductory Message to the teacher participants.

Thanking you in anticipation.

Yours sincerely,

Hemlata Karki

EID 2008267

APPENDIX 6: INTRODUCTORY MESSAGE AND CONSENT TO TEACHERS

Dear Teacher,

My name is Hemlata Karki, a Master student in the University of Adelaide, studying a degree of Master of Education. For my research, I have chosen to explore **on Bhutanese elementary teachers' beliefs, practice, knowledge and literacy of formative assessmen**t. This is significant as written examinations from PP to grade 6 in Bhutan are being removed by next year and formative assessment is the basis for students' promotion to next grade. Further, data-driven decisions and high-stakes assessments continue to grow. I am hopeful that with your help, we can begin to identify areas in need of best practices for practicing teachers in Bhutan and enhance students' learning.

Therefore, I would much appreciate your participation in this study. Please take 45 to 60 minutes to fill out this online survey. You are being invited because you are an elementary teacher teaching from PP to grade 6 in Bhutan. The detail of the research is explicitly given in the Participation Information Sheet that you can read by clicking the link below. **Before attending the survey, I request you to go through the Participation Information Sheet.** Please feel free to contact if you have any queries and clarification to make. Contact details of the University of Adelaide are given in the Participation Information Sheet, while in your school you may contact

Name: _____

Mobile Number: _____

Email address: _____

To participate in the survey, please follow the steps given below.
Step 1 - Click on the link to the survey: https://www.surveymonkey.com/
Step 2- Read the Participation Information Sheet
Step 3 - Follow the instructions, clicking "next" at the bottom of every screen
Step 4 - If you need to take a break please click "save" and then come back to finish up the survey.
Step 5 - Remember to click "done" at the end of the survey when you have completed the entire survey.
My sincere thanks to you for your time and consideration. This study could not be completed without your help.

Sincerely yours,

Hemlata Karki

APPENDIX 7 ACKNOWLEDGEMENT AND THANKING EMAIL TO THE PRINCIPALS

Dear Principal

I am thankful to you for granting me the permission letter and for inviting the potential participants from your school in taking part in this research.

As mentioned in the previous email, this research is essential to provide an insight to Bhutanese teachers' beliefs, practices, knowledge and literacy of formative assessment in identifying the areas in need of best practices of assessment for teachers and enhance student learning.

Therefore, once again I would like to request you to kindly remind your teacher participants to attend the survey if they have not yet done.

I extend my appreciation and thanks for supporting this research project.

Yours sincerely,

Hemlata Karki

The University of Adelaide

APPENDIX 8 SAMPLE SCHOOLS

SL No.	Name of the school	Email address	Local Contact Person
1	Tang CS	bt.tangcs@education.gov.bt	Gyem Tshering/Lhamu
2	Chumigthang MSS	ch.chumigthangmass@education.gov.bt	Tandin Wangdi
3	Phuentsholing LSS	plinglss@education.gov.bt	Lhapchu Tshering
4	Dagapela MSS	dg.dagapaleamss@education.gov.bt	Sherub Jigme
5	Daleythang LSS	dg.daleythanglss@education.gov.bt	Phurba Sing Tamang
6	Tashiding LSS	dg.tashidinglss@education.gov.bt	Bimal Sunar
7	Drukgyal CS (lower)	drukgyellss@education.gov.bt	Tshering Yangden
8	Norbuling CS	sp.norbulingcs@education.gov.bt	Dawa
9	Gelephu MSS	gelephumss@education.gov.bt	Karma Rinzin
10	Gelephu LSS	gelephulss@education.gov.bt	Radhika Chhetri
11	Kuzuchen MSS	tp.kuzuchenmss@education.gov.bt	Deki
12	Lungtenphu LSS	lungtenphulss@education.gov.bt	Kabita Nepal
13	Jigme Namgyal LSS	jigmenamgyallss@education.gov.bt	Pem Dem
14	Dorokha CS	dorokhacs@education.gov.bt	Parmila Sharma
15	Samtse LSS	st.samtselss@education.gov.bt	Phul Maya Zimba
16	Damphu LSS	tr.damphulss@education.gov.bt	Kinzang Dorji
17	Rangthangling PS	tr.rangthanglingps@education.gov.bt	Cheku
18	Khandothang PS	kandothangps@education.gov.bt	Chador Wangmo
19	Gesarling CS	dg.gesarlingcs@education.gov.bt	Tika Chawan
20	Daga LSS	dg.dagalss@education.gov.bt	Tshering Dema
21	Kidheykhar CS	kidheykharcs@education.gov.bt	Pema Rinzin
22	Taju PS	tajupryschool@education.gov.bt	Quenley Dema
23	Samdrup Jongkhar PS	sjongkharps@education.gov.bt	Yeshey Wangmo
24	Tashi Yangtse LSS	ty.trashiyangtselss@education.gov.bt	Tashi Phuntsho
25	Lobesa LSS	pk.lobesalss@education.gov.bt	Tashi Phuntsho

APPENDIX 9 CFA READING

Confirmatory Factor Analysis (CFA) is a type of Structural Equation Modelling (SEM) that analyses quantitative data. It is theory-driven and aims to verify the hypothesised factor structure of any scale (Dimitrov, 2013; Schreiber, Stage, Barlow & King, 2006). According to Probst (2003), CFA is employed to provide evidence of construct validity. Since it is a theory- driven approach of confirming a hypothesised factor structure, its analysis is directed by theoretical relationships among latent and observed variables. Therefore, CFA allows testing of theoretical models that hypothesise how items define constructs and how the constructs are related to each other (Schumacker & Lomax, 2010).

There are five sequential steps of CFA:

- Model Specification,
- Model Identification,
- Model estimation
- Model Testing
- Model modification

Model specification: This step is concerned with the construction of the theoretical model based on theory and past research (Schumacker & Lomax, 2012). In this step, the researcher is involved in determining the number of factors, the factors related to the observed variables, whether or not a correlation exists, and the equality of factor loading (Dimitrov, 2013). In other words, the researcher specifies the relationships and parameters of the model.

Model identification: In this step, data information is used in determining the possibility of parameter estimation, because in CFA, it is necessary to determine whether the theoretical/hypothesised model is identified (Ryan, 2018, Hailaya 2014).

Model estimation: This third step comes after the model has been identified. During this step, the researcher looks for "fitting function" (Schumacker & Lomax, 2012) in which he / she estimates the parameters in the model employing several methods. Model estimation is crucial to minimize the covariance difference between the population covariance matrix (©) and sample covariance matrix

(*S*). Indeed, when the difference is zero, the x^2 becomes zero, which indicates a perfect model fit to the data (Schumacker & Lomax, 2012).

Model testing: A step that determines how well the data fit the model, or to what extent the data support the proposed model (Schmacker & Lomax, 2012). This is determined using several fit indices. The higher the similarity between the population covariance matrix (©) and sample covariance matrix (*S), the* better the model fit to the data.

Model modification: This final step is undertaken to achieve better fit of the model to the data if the model fit is not adequately achieved. However, Schumacker and Lomax, (2012) suggest conducting the model modification step with caution so that the purpose of CFA is not defeated.

CFA analysis involves two parts, the measurement and structural parts of the model. The measurement model is assessed to ensure significant relationships between observed and unobserved variables. The size of factor loadings provides indications about these relationships. Instruments with multiple choice items or dichotomous data (Yes/No), require 0.3 or less as an accepted factor loading size (Kline, 1994). The measurement model is established to test construct validity and composite reliability. Once the measurement model is found to be acceptable, the structural model – which determines the significant relationships among unobserved/latent variables – can be assessed.

Numerous fit indices are used in the assessment of the proposed model fit to the data. The most commonly used are chi-square (x^2), normed chi-square (x^2 /df), root mean square error of approximation (RMSEA), Comparative Fit Index (CFI) and Trucker-Lewis index (TLI) (Schreiber, Stage, Barlow & King, 2006) . Though there are also other indices – such as standardised root mean square residual (SRMR), goodness-of-fit index (GFI), adjusted goodness-of-fit index (AGFI) and parsimony goodness-of-fit (PGFI), this research utilised the former set of indices, which will now be explained in turn.

The Chi-square index (x^2) is an index of 'exact fit', as it calculates the perfect fit of a model to data (Matsunaga, 2010). However, it is known that this index is sensitive to sample size, and that it almost always indicates a poor model fit. Therefore, Probst (2003) suggests dividing chi-square by

the number of degrees of freedom (df) to further analyse the model. According to Wheaton et al (1977) the normed chi- square (x^2 /df), minimises the impact of sample size on Chi-square. The static values of normed chi-square (x^2 /df) ranges from as high as 5.0 (Wheaton et al 1977) to as low as 2.0 (Tabachnick & Fidell, 2007).

The root mean square error of approximation (RMSEA) suggests how well the model fits the population covariance matrix (Byrne, 1998). This index is known for excluding the influence of sample size as well as performing statistical tests on the values, and is therefore, considered one of the primary indicators for evaluating the goodness of fit of a model. Generally, RMSEA is acceptable with a value of 0.08, but less than 0.05 is better and less than 0.01 is a perfect model (Kline, 2015). There are others who recommend different RMSEA values, such as close to 0.06 (Hu & Bentler, 1999), or a stringent upper limit of 0.07 (Steiger, 2007). An RMSEA value more than 0.10 indicates a poor fit of the model to data (Diamantopoulos & Siguaw, 2000)

The CFI is one of the most popularly reported fit indices (Fan et al, 1999) as it is not impacted negatively by the size of a sample (Tabachnick & Fidell, 2007). The commonly recommended CFI value is $\geq$ 0.95 (Hu and Bentler, 1999) or 0.90 (Matsunaga, 2010). However, CFI values can range from 0.0 to 1.0, with values closer to one considered a good fit.

The Trucker-Lewis index (TLI) is also used to test model fit. Unlike CFI, TLI accounts for the consequences of adding parametersI. However, they are interpreted similarly in defining the model fit.

To summarise, the cut off value for RMSEA is 0.06, 0.95 for TLI; and 0.95 for CFI, according to Hu and Bentler (1999), especially for continuous data. However, Yu (2002) also suggests that these indices and the cut-off values are reasonable even for categorical data. Thus, this research uses the X^2 , X^2 /df, RMSEA, TLI and CFI as goodness-of-fit indicators for categorical data and continuous data.