

**THE DEVELOPMENT OF GRADING
PRACTICES INVENTORY AND THE
DIFFERENCES OF GRADING PRACTICES
AMONG ACADEMIC AND RELIGIOUS
TEACHERS IN THAILAND**

MAYUTI DUERAMA

UNIVERSITI SAINS MALAYSIA

2018

**THE DEVELOPMENT OF GRADING
PRACTICES INVENTORY AND THE
DIFFERENCES OF GRADING PRACTICES
AMONG ACADEMIC AND RELIGIOUS
TEACHERS IN THAILAND**

by

MAYUTI DUERAMA

**Thesis submitted in fulfilment of the requirements
for the degree of
Doctor of Philosophy**

August 2018

DEDICATION

**THIS THESIS IS DEDICATED TO MY PARENTS
IBRAHIM DUERAMA AND AMINAH DUERAMA, MY
SISTERS AND BROTHERS WHO GAVE GREAT
SUPPORT AND ENCOURAGEMENT**

ACKNOWLEDGEMENT

All praise is to Allah. The completion of a doctorate degree is an enormous undertaking. A lot of work have been carry out during the preparation of this dissertation besides getting consultation, advice and support from the specialists and experts in related fields. I would like to express my deepest gratitude and thanks to the following individuals: My former main supervisor, Assoc. Prof. Dr. Ong Saw Lan (retired), and Assoc. Prof. Dr. Shuki Osman as Main Supervisor and Assoc. Prof. Dr. Rohizani Yaakub as co-supervisor, of School of Educational Studies, USM for their support, encouragement, contribution, valued advice and commitment throughout the process and completion of this dissertation. Dean, lecturers and administrative staffs in School of Educational Studies, USM for their great and kind support. Dean, Faculty of Education, University Fatoni, Pattani, Thailand for his permission and support to conduct this study. All academic and non-academic staff and students Faculty of Education, University Fatoni for their support and cooperation during the early stage of this study. Foremost, I would like to thank my parents Ibrahim and Aminah, my brother Mohd Husni, my sister Huda for their unconditional love an support throughout my life.

TABLE OF CONTENTS

ACKNOWLEDGEMENT	ii
TABLE OF CONTENTS	iii
LIST OF FIGURES	vi
LIST OF TABLES	vii
ABSTRAK.....	viii
ABSTRACT.....	x
CHAPTER 1 - INTRODUCTION	1
1.1 Introduction.....	1
1.2 Background.....	4
1.3 Statement Of Problem	8
1.4 Purpose and Objective of Study.....	12
1.5 Research Questions	13
1.6 Significance of the study	14
1.7 Limitation of the study	14
1.8 Definitions of Terms	15
CHAPTER 2 – LITERATURE REVIEW	17
2.1 Introduction.....	17
2.2 Educational Assessment	17
2.3 Grading Practice.....	24
2.4 Grading Approach.....	26
2.5 Guiding Principles of Grading Practice.....	31
2.6 Measuring Teachers’ Grading Practices.....	37
2.7 Psychometric Properties of The Scale.....	41
2.7.1 Validity	42
2.7.2 Reliability	45
2.8 Theories Related To The Study	48
2.8.1 Teacher’s Assessment Decision Making.....	48
2.8.2 Teacher Professional Judgment	55
2.8.3 Teacher Assessment Competency.....	61
2.9 Theoretical framework	68
2.9.1 The Use Of Assessment Method/Type To Obtain The Data To Grade	71
2.9.2 Grading Factors: Factors Included in Assigning Grade	74

2.9.3 Grading Process	78
2.9.4 Grading Behavior.....	81
2.10 Conceptual framework	83
2.11 Conclusion.....	87
CHAPTER 3 - METHODOLOGY	88
3.1 Introduction.....	88
3.2 Research Design.....	88
3.3 Sample	89
3.4 Instrumentation	91
3.4.1 Examination of Content Validity	92
3.4.2 Pilot Study 1.....	94
3.4.3 Pilot study 2	95
3.5 Data Collection	105
3.6 Data Analysis	105
3.6.1 Checking Rating Scale Structure	106
3.6.2 Checking for Dimensionality.....	107
3.6.3 Checking for Differential Item Functioning and Person Bias	108
3.6.4 Item Calibration	109
3.6.5 Item Fit and Person Fit	109
3.6.6 Item Reliability and Person Reliability	110
3.6.7 Item Map.....	111
3.6.8 Checking For Factorial Invariance.....	112
3.7 Testing Group Differences in Grading Practice.....	112
3.8 Conclusion	113
CHAPTER 4 – FINDINGS OF THE STUDY	114
4.1 Introduction.....	114
4.2 Rating scale category structure of GPI.....	115
4.3 Item and person reliability	122
4.4 Differential Item Functioning (DIF) analysis	126
4.5 Confirmatory Factor analysis.....	130
4.6 Factor structure of the GPI	135
4.8 Discrimination Validity	139
4.9 Measurement Invariance Analysis	140
4.10 Grading Practice Inventory – Final Version.....	144
4.11 Profile of Islamic Private Teachers in the Grading Practice.....	147

CHAPTER 5 – DISCUSSION AND CONCLUSION.....	159
5.1 Introduction.....	159
5.2 Research Findings and Discussion.....	159
5.3 Psychometric Properties of GPI.....	159
5.4 Teachers’ profile in the grading practice.....	165
5.5 Implication of the study.....	169
5.6 Recommendation For Future Research	170
BIBLIOGRAPHY	172
APPENDIX	

LIST OF FIGURES

	Page
Figure 2.1 Teachers' assessment decision making theory (McMillan,2003)	55
Figure 2.2 Conceptual Framework of Grading Practice.....	84
Figure 3.1 Instrument Development Process (adapted from Du Preez, Visser, & Janse Van Noordwyk, 2008; Moore & Benbasat, 1991)	90
Figure 3.2 Category Threshold	98
Figure 3.3 Logit Map of Person and Item From GPI.....	103
Figure 4.1 Rating scale structure of GPI	115
Figure 4.2 Rating scale structure of Assessment method subscale.....	117
Figure 4.3 Rating scale structure of Grading factor subscale.....	118
Figure 4.4 Rating scale structure of the grading process subscale.....	120
Figure 4.5 Differential Item Functioning (DIF) plot between Islamic and Academic teachers.....	126
Figure 4.6 Differential Item Functioning (DIF) plot of Assessment method subscales.....	127
Figure 4.7 Differential Item Functioning (DIF) of Grading factor subscale	128
Figure 4.8 Differential Item Functioning (DIF) of the Grading process subscales.....	129
Figure 4.9 Differential Item Functioning (DIF) of Grading behavior subscale ...	130
Figure 4.10 Measurement model for Assessment method construct	132
Figure 4.11 Measurement model for grading factors construct.....	133
Figure 4.12 Measurement model for the grading process construct.....	134
Figure 4.13 Measurement model for grading behavior construct.....	135
Figure 4.14 Final Measurement Model Of GPI With Parceled Items.....	137
Figure 4.15 CFA model of GPI across Academic teachers.....	140
Figure 4.16 CFA model of GPI across Islamic teachers	141

LIST OF TABLES

		Page
Table 2.1	The theoretical constructs of the grading practice.....	86
Table 3.1	Reliability, Number of items and Coefficient alpha.....	95
Table 3.2	Item-total correlation of GPI.....	96
Table 3.3	Summary of measure steps	97
Table 3.4	Measure, Infit and Outfit Mean Square, Infit and Outfit.....	100
Table 3.5	Item and person reliability, Item and person separation index	104
Table 4.1	Summary of measure steps of GPI.....	116
Table 4.2	Summary of rating scale category structure for Assessment method subscale.....	117
Table 4.3	Summary of rating scale category structure for the Grading factor subscale.....	119
Table 4.4	Summary of rating scale category structure for the Grading process subscale.....	120
Table 4.5	Rating scale structure for the grading behavior subscale	121
Table 4.6	Summary of rating scale category structure for Grading behavior subscale.....	122
Table 4.7	Item and person reliability and separation indices of GPI.....	122
Table 4.8	Item fit statistic and item reliability for each subscale	123
Table 4.9	Item fit statistic and person reliability for each subscale.....	125
Table 4.10	Summary of fit indices from CFA model of GPI for full sample (N=1224).....	137
Table 4.11	Convergent validity and reliability of constructs.....	138
Table 4.12	Discriminant validity of constructs	139
Table 4.13	Summary of fit indices from CFA for Islamic teacher samples data .	142
Table 4.14	Summary fit indices of five nested model for multi-group CFA.....	143
Table 4.15	Grading Practices Inventory – Dimensions, Constructs and Items.....	144
Table 4.16	Descriptive statistics for Grading Practice Inventory (GPI).....	147
Table 4.17	Descriptive statistics of each two group of teachers	152
Table 4.18	Independent T-test for GPI between Isalmic and Academic teachers.	156
Table 4.19	Independent T-test for subscale between Islamic and Academic teachers	156
Table 4.20	Independent T-test for each component between Islamic and Academic teachers.....	157

**PEMBINAAN INVENTORI AMALAN PENGGREDAN DAN AMALAN
PENGGREDAN DIKALANGAN GURU AKADEMIK DAN AGAMA
THAILAND**

ABSTRAK

Penilaian dan penggredan adalah tanggungjawab profesional bagi guru-guru dalam menjalankan pengajaran dan membuat kesimpulan tentang hasil pembelajaran pelajar. Proses penilaian dan penggredan boleh berubah dikalangan guru-guru dalam keadaan yang berbeza dan konteks. Oleh itu, untuk memahami amalan penilaian dan penggredan dikalangan guru adalah penting bagi meningkatkan keupayaan guru tentang penilaian dan penggredan. Tujuan utama kajian ini adalah untuk membangunkan dan menilai ciri-ciri psikometrik “Inventori Grading Practice (GPI)”, dan untuk memeriksa profil “invariance” amalan penggredan guru merentasi disiplin yang berbeza. Seramai 1394 orang guru yang mengajar mata pelajaran agama dan akademik di sekolah swasta Islam di Selatan Thailand terlibat dalam kajian ini, dan Rasch Rating Scale Model (RRSM) dan pengesahan analisis faktor (CFA) telah digunakan untuk menganalisis data. Hasilnya menunjukkan bahawa GPI adalah instrumen yang sah dan boleh dipercayai untuk mengukur amalan penggredan guru. Analisis ciri psikometrik menunjukkan bahawa kesahan kandungannya memadai, kesahan konstruk yang memuaskan, dan kebolehpercayaan instrumen adalah tinggi. Keputusan analisis Rasch mencadangkan bahawa GPI adalah unidimensional, dan semua ciri berfungsi secara sama bagi kedua-dua kumpulan sampel guru. Hasil daripada pelbagai kumpulan analisis CFA menunjukkan bahawa model GPI patut diterima oleh data dari dua kumpulan, membawa kepada sokongan bahawa model itu tak berubah. Hasil analisis

menunjukkan bahawa guru-guru dalam mata pelajaran akademik dan Islam adalah berbeza dalam menggunakan kaedah penilaian, mempertimbangkan faktor penggredan, dan menentukan proses penggredan. Kajian ini memberikan pemahaman yang bermanfaat tentang amalan penggredan guru bagi seluruh mata pelajaran akademik dan Islam.

**THE DEVELOPMENT OF GRADING PRACTICES INVENTORY AND
THE DIFFERENCES OF GRADING PRACTICES AMONG ACADEMIC
AND RELIGIOUS TEACHERS IN THAILAND**

ABSTRACT

Assessment and grading is a professional responsibility for teachers to carry out teaching and make conclusions about students' learning outcome. The process of assessment and grading can be varied among teachers in different situations and contexts. Thus, to understand the practice of assessment and grading among teachers is essential for improving teacher's capability in assessment and grading. The primary purpose of this study was to develop and evaluate psychometric properties of the Grading Practice Inventory (GPI), and to examine the invariance profiles of teachers' grading practices across different disciplines. In this study, 1394 teachers teaching religious subjects and academic subjects in Islamic private schools in southern Thailand were involved, and the Rasch Rating Scale Model (RRSM) and Confirmatory Factor Analysis (CFA) were applied to analyse the data. The results revealed that the GPI is a valid and reliable instrument to measure teachers' grading practices. The analysis of psychometric properties indicated that content validity was adequate, the construct validity was satisfactory, and the reliability was high. The results of Rasch analysis suggest that the GPI was unidimensional, and all items functioned similarly across two group of sample teachers. The results of multi group CFA analysis showed that the GPI model is an acceptable fit to the data across two groups lending to support that the model was invariant. The results of the analysis indicate that teachers in academic and Islamic subjects differed in using assessment methods, considering grading factors, and determining the grading

process. This study provided a beneficial understanding of teachers' grading practices across academic and Islamic subject areas.

CHAPTER 1

INTRODUCTION

1.1 Introduction

Teachers need to understand classroom assessment in their teaching. Their understanding is essential for developing and implementing instructions that help to achieve the desired learning outcomes (Brookhart & Durkin, 2003; Stiggins & Conklin, 1992). Essentially, classroom assessment is a continuous process that allows teachers to gather information and provide feedbacks and expectations to students that motivate their learning (Wiggins, 1998). A well-designed assessment by teachers would enhance instruction, influence students' motivation, and provides feedback to their learning (McMillan, 2000)

Therefore, one major purpose of classroom assessment is to gather information to make conclusions about the level of students' learning outcome, that is the "grade" level of the outcome of learning by the students (Olsen, 2004, Lekholm, 2008). Through classroom assessment, assigning grades to students by teachers is one of their most obligatory responsibilities. The practice of assigning grades by teachers is the process of making judgement on the quality of students' performance, which is termed as "grading practice" (Airasian, 2001).

The grades that teachers assign to students should encompass a meaningful basis for making an evaluative statement related to their achievements (Musial et al., 2009.) Classroom assessment practices lead to significant effects on students' success in the long term, and their ability to retain and apply what they have learned in different contexts (Crooks,1988). Therefore, there is a need for

understanding teachers' assessment (and grading) practice in order to find out if a teacher uses quality assessment methods in their teaching that fulfil the learning needs of students (McMillan, 2001).

According to Adams and Torgerson (1964), in teaching and learning, the practice of grading serves at least four different functions: administration, information, guidance, and motivation. For the administrative function, grading helps schools in sorting and grouping students, such as into different ability groups. It also assists in making decisions about selecting individual students for scholarships, honors, graduation, and employment. In terms of providing information, grading can inform students progress toward certain educational goals. The guidance function of grading refers to the use of grades to identify areas of strength and weakness so that students can plan their study agendas and their educational and vocational future. In the aspect of motivation, grades become incentive that motivate students to work harder to improve or to sustain their performance (Elikai & Schuhmann, 2010). Furthermore, when grading also includes students assignments, it will encourage students to complete them, thereby increases their understanding of subject matter and improving their class performance (Norman,1981).

Typically, the grading practice involves the process of making decisions based on the assessment of performance and evaluative symbols to represent what students know and can do or are able to do (Musial et al.,2009). It reflects the conclusion of decision-making processes which indicate how well a student progressess (Donaldson & Gray, 2012). According to MacMillan (2001), grading practices require professional judgment. The judgment of grading depends upon

two points: first, it depends on the information about the student being judged, and the basis of comparison to translate the particular information into judgment that indicates the quality level (Airasian et al., 2001). In practice, before teachers make a decision on a particular student, they assess the students' performance and then they make judgments in the light of their knowledge and experience. Therefore, teachers' grading practices can be different from one teacher to another. For instance, one teacher might look at the test and conclude that students had mastered a skill, while another teacher might conclude the opposite. These differences of the grading process among teachers might affect the students in their learning motivation and also may reduce the reliable meaning of students' grades if it assigned by different criteria.

In Thailand, students' grades are used as the indicators of learning outcomes for their school subjects. Thus, grades are released to students and reported to parents in report cards to indicate the level of learners' performance in all the subjects taught. When grades are converted to a point system, students' grades point average (GPA) would be used for the sorting and selection of students for various certification purposes (Ministry of Education, 2008). These grading practices are being employed throughout the education system in all government or public schools in the country.

In the context of Islamic schools that combine the religious education together with general education, the differences of grading practices among teachers that came from different educational philosophies are really an interesting issue for a researcher to explore. Thus, this study seeks to explore the grading

practices among both religious and academic teachers in Islamic private schools in Thailand.

1.2 Background

Since the study is investigating teachers' grading practice in the context of Islamic private schools in Thailand, it is essential to understand that the private school was formed, at the beginning, to provide only Islamic religious education.

Islamic education refer to the efforts of the Muslim community to educate its own, to pass along the heritage of Islamic knowledge, first and foremost through its primary sources, the Qur'an (the Holy Book) and the Sunnah (the action of Prophet). This education of Muslims commonly takes place in mosques, schools or universities (Douglass & Shaikh, 2004). Fundamentally, the aim of Islamic education is to educate a Muslim to know Allah who is the creator of mankind and all the universe. All Muslims are obligated to seek knowledge of Allah, thereby encouraged to praise the Greatness of Allah and thank Him for His Mercifulness (Mohammad Hasan, 2007). In this regard, the Islamic education system is really needed for the Muslim community either in a Muslim majority country or minority one including Thailand.

The growth of Islamic education in Southeast Asia including Thailand, was closely related to the spread of Islam in the region with a very distinctive one compared to other areas of the Muslim world. It started in the mosques and/or local educational institution such as "Pondok" which is a place for Muslim to study Islam with a classical system. Then it was modernized by some adopted practices; firstly, by fully integrating Islamic educational institutions into national education which are run and financed by the government, and making a standardization of

Islamic education in accordance with national standards while the ownership and administration remain mostly in the hand of Muslims (Azra, 2010).

Consistent to other countries in Southeast Asia, Islamic education in Thailand has been provided with traditional system in the traditional Islamic boarding schools or “Pondok” for a long period. Islamic knowledge teaching was undertaken by the principle of the Pondok called “Tok Guru”, in various subject areas including the fundamental belief in Allah (Aqidah) , the Qur’an (Holy Book), and the Hadith (the actions of Prophet). Academically, the Islamic education that took place in Pondok was seen as unsystematic. It was because the content of the courses and the graduation depended solely upon the judgment of Tok Guru (Narongraksakhet, 2003).

Subsequently, the new system of Islamic education was implemented in Pondok since it was reformed by a government policy in 1966 to become Islamic Private Schools (Narongraksakhet, 2003). The transformation of Pondok into Islamic Private schools was under-taken by Thailand’s government. The Pondok were viewed as religious institutions rather than educational institutions. This means that they play a lesser significant role for social order, and are obstacles to national education, economy, and political development in the country (Surin,1982).

As a result of reformation, the Islamic Private Schools were transformed from a traditional Islamic education system to a more formal Islamic educational system (Othman & Wanlabeh, 2012). Namely, they have formal curricula, and provide religious studies together with general basic education (Liow, 2009). The school curriculum was modified from pure Islamic religious studies into integrated

education that combine religious and general subjects. In terms of religious studies, the Islamic Private Schools have implemented the curriculum of Islamic studies that covers eighth subject areas for both primary and secondary levels; namely Ibtida'e level, Mutawasit level, and Thanawi (Ministry of Education, 2003).

In fact, the curriculum of Islamic studies for Islamic private Schools has been continuously developed and adjusted by the Educational Development Center Region II (Regional Education office No.12 as currently) to be in line with the national education curriculum. Thus, the reformation of curriculum has been done from time to time; the curriculum of Islamic studies 1961, the curriculum of Islamic studies 1970, the curriculum of Islamic studies 1974, the curriculum of Islamic studies 1980, Integrated Curriculum 1992, and the Curriculum of Islamic Studies 1996. In 2003, the Curriculum of Islamic studies 2003 which is currently implemented in Islamic private schools. It was developed to be in line with the Curriculum of Basic education 2008 which was updated from The Curriculum of basic education 2001 (Manyunu, 2008 ;Narongraksakhet, 2003).

Despite the fact that Islamic Private Schools have been supported financially by the government, the educational quality in these schools still faces many problems and needs to improve urgently (Intarak, 2010). According to Aree (2011) there are various problems that Islamic Private Schools are facing, such as students not prepared to compete in the job market or gain admittance to university, test scores are below the national average, and also the low competency of teachers.

It appears that the academic achievement of students from Islamic private schools is low, especially the schools situated in the south. They are among the lowest ranks according to the national standardized testing during the academic year

2010 (Ramkaew, 2010). Also, it has been found that a number of students from Islamic Private Schools who entered the university have unsatisfactory achievement and dropped out of university due to their weakness in core subjects (Boonphikam, 2008). According to Rajapat Yala University (2006) the unsatisfactory achievement of students in Islamic private schools is due to poor management of the school curriculum and instruction.

Although researchers found that school administration, teachers' instructional management, and students' characteristic were keys factor of the Islamic private schools effectiveness (Uma et al.,2009), in fact another study indicated that teachers' competency was positively related to students' achievement in these schools (Leaheem, 2005) . Besides, interaction between teachers and students in the classroom was best predictor of teaching competency in the Islamic private schools teachers (Wea-u-seng, 2008). These indicate that teachers' action in classroom is a major aspect in improving students' achievement in Islamic private schools.

Generally, teachers' assessment practice in the classroom is an important component for addressing students' learning needs, and improving the education system. Teachers' assessment practice possesses strong effects on student because they can inform the students about what decisions should be taken and what to study and how to learn (Brookhart, 1994). Additionally, the classroom assessment practice affects more than just the students' achievement; their motivation and emotional response. Thus, understanding teachers' assessment practice helps researchers to determine whether the teachers utilized quality assessment method to meet the needs of students (McMillan, 2001).

In line with that, in order to choose an appropriate format of assessment instrument, teachers have to be aware of the strengths and weaknesses of various assessment methods, especially related to the learning objectives of the students (Stiggins, 1992). Furthermore, assessment activities that are used by the teachers appear to have significant effects on students in the long term, including their ability to retain and apply the learned materials in varied contexts (Crooks, 1988).

Accordingly, teachers should bear in mind that formative and summative assessment should consistently evaluate students' achievements for the purpose of grading (Ohlsen, 2004). Grades can be considered as a feedback about how well student meet the expectations, thus parent can use grades to understand how their children are doing at school, while school use grades as a part of the formula to determine which student passes to the next level of study and which student are required to repeat (Musial et al., 2009). Grades can also offer quick and concise data points for counselors to help students in planning their future education.

As a process of judging the quality of students' performance, grading provides an important source for making evaluative conclusions and decisions related to their achievements (Airasian & Russell, 2001, Donaldson & Gray, 2012; Musial et al., 2009). Therefore, the usefulness of grading lies on its accuracy in measuring what it meant to measure.

1.3 Statement Of Problem

Educators agreed that the purpose of grading is to communicate students' academic achievement (McMillan, 2001; Musial et al., 2009; and Stiggins, 2008). Therefore, teachers should base their students final grades primarily on academic performance (McMillan, 2001). In other words, student grades indicate how much

they have learnt from the instruction. Hence the formal assessments of students' achievement of the course objectives should be the major component of subject matter grades, then they should be accorded major weight in assigning grades (Airasian & Russell, 2001). Additionally; the use of achievement as the basis for grading appear to be reliable than with other methods of grading (Banks, 2005).

However, many studies indicated that non-academic factors have been used in assigning student grades; for instance, McMillan (2002) found that non-tests and behavior such as effort, participation, and extra credit work are very important for many teachers. Similarly, Kushniruk's (1994) study revealed that teachers included non-achievement factors such as, effort, aptitude, improvement, work habits, attitude, class participation, and cooperativeness in grading. Teachers also reported that the other non-achievement factors such as classroom behavior, student attendance, and student character played less significant roles in deciding students' grades (Rich, 2001).

The differences in grading practice might have an impact on students' achievement. According to Betts and Grogger (2003), students respond favorably to the incentives provided by a higher grading standard, thus the student test scores rise more for schools with a higher standard, and higher grading standards also rise students' post schooling earning. Similarly, Iacus & Porro's study (2011) revealed that harder grading standards are associated with higher achievement levels. Bonesrønning (2004) stated that the grading effects depend on how the grading is designed. Namely, students who are exposed to hard grading perform better than other students. In addition, the study also revealed that teachers used grading as an instrument to potentially manipulate student effort. These indicate that the

differences of teachers' grading practice do have an impact on students' achievement.

According to Dyb's (2012), grading practice is affected by some categories of factors, such as experiences, pre-service training, district policies and school guideline, professional development, peer and school culture, philosophies of teaching and learning, teaching experience and motivation, instructional leadership, and state or national standard.

Consequently, teachers who believe that the primary purpose of grading is to communicate a summative evaluation of student achievement and performance always base their grade on a test score, report or project, overall assessment, and other culminating demonstrations of learning (Musial et al., 2009). While some other teachers often use them as a motivator in reward and punishments to manage students' behavior, because they believe that those who do more learn more, thereby by grading on effort, for instance, could drive students toward greater achievement (Stiggins, 2008).

Furthermore, it is difficult for teachers to handle sources of information that represents achievement in nature versus additional variables such as effort, behavior, and ability (Allen, 2005). Therefore, the teachers' grading practice is varied. Thus, for some teachers, academic achievement is a major factor affecting their grading practice, while some others tried to be fair to the students by including factors other than academic achievements, such as student efforts and abilities, although it might not reflect their academic achievement (Liu, 2008). Thereby, teachers might not be sure what factors should be included in their student grades,

or how to compile the grades that communicate the actual quality of student performance.

In practice, grading practice involves the process that translate assessment information into marks or letters to indicate the quality of each student's learning and performance. This process depends on two points: first, it depends on the information about the student being judged, and second, the basis of comparison which can be used to translate the particular information into judgment that indicates the quality level (Airasian & Russell, 2001). In fact, before the teachers make a decision on a particular student, they assess a student's performance and then make judgments in the light of their knowledge and experience. Thus, the difference of grading practices among teachers may be due to the difference of assessment proficiency among individual teachers.

However, even though scholars have suggested the appropriate procedures to carry out, there has been differences among teachers about the grading practice. The differences in teachers' grading practices might occur due to teachers' different expectancy towards different students (Randall & Engelhard, 2009). It might also be due to the teachers' attempt to be fair to the students by including some non-academic factors (Brookhart, 1994). In addition, the grading practice among teachers might be influenced by teachers' personal philosophy of learning and teaching, the local official grading policy, as well as perceived and actual consequences in grading (Randall & Engelhard, 2009; Tomlinson, 2001). Thus, an initial investigation into the differences of teachers' grading practice will contribute to the understanding of teachers' development in relation to assessment and evaluation of students' performance.

Grading is not done only at the end of a course, but it should be incorporated into classroom through out a course (Speck, 1998). In the practice of grading, teachers are required to be proficient and competent in developing a valid grading procedure such as choosing an appropriate method of assessment to assess students in order to obtain accurate information for making a decision about an individual student (AFT,NCME & NEA, 1990). Hence, teachers who are competent in the grading practice would implement an adequate procedure in evaluating students' achievement. In contrast, incompetent teachers who have low proficiency and competency may use an inappropriate procedure in their grading practice, and then it would affect the accuracy the evaluative information. Thus, the level of proficiency and competence in grading among individual teachers should be looked into.

In an effort to understand the teacher's grading practice, it is critical to develop a valid and reliable instrument for measuring the grading practice of teachers. Such instrument will be used to derive the information to serve in determining the actual grading practice among teachers. Therefore, the development of the Grading Practice Inventory (GPI) to investigate the teachers' grading practice is the main endeavor of the present study.

1.4 Purpose and Objective of Study

The primary purpose of this study is to develop and validate an instrument known as the Grading Practice Inventory (GPI) for measuring teachers' grading practice. The instrument will be evaluated for its psychometric properties to ascertain the validity and reliability. The instrument will be used to evaluate the teachers' grading practice to determine the differences of teachers' grading practice

across different groups of teachers. Therefore, the objectives of the study can be simplified as follows:

- 1) To develop the Grading Practice Inventory (GPI),
- 2) To evaluate the psychometric properties of the Grading Practice Inventory,
- 3) To examine the invariance profiles of teachers' grading practice for different disciplinary areas.

1.5 Research Questions

In developing the GPI, the psychometric properties concerning validity and reliability of the instrument will be examined, and an evaluation of the profile of teachers' grading practice, the present study will employ appropriate statistical analysis that allows valid interpretation of the results. Therefore, the present study seeks to answer the following questions:

- 1) To what extent does the Grading Practice Inventory (GPI) demonstrate its validity?
- 2) To what extent does the Grading Practice Inventory (GPI) demonstrate its reliability ?
- 3) Does the Grading Practice Inventory (GPI) demonstrate measurement invariance among teachers in different disciplines ?
- 4) Are there any significant differences in grading practices between religious teachers and academic teachers teaching different areas of disciplines ?
- 5) How do religious teachers and academic teachers teaching different areas of disciplines, differ in their grading practice?

1.6 Significance of the study

This study had planned to develop a sound research instrument to measure grading practices called the Grading Practices Inventory, tested with the data collected. Two elements of a good instrument to be assessed are the its validity and reliability, by the following: the Rasch Analylis Model. Second, this research also expected that the findings will provide a comprehensive picture of the actual practice of grading among teachers of Islamic Private Schools in Thailand, in relation to the appropriate grading practices recommended by experts as found in the literature. The implications of the findings of this study might determine the intensity and limitations of teachers in terms of grading practices, which would suggest for policy revisions on teaching quality, involving grading practice by teachers in schools in Thailand. The results of this study are also significant for decision making on training of futurer teachers about testing and evaluation, classroom assessment and grading practice. Also it may serve the universities with information in improving thier teacher education courses, particularly the specific course related to educational assessment.

1.7 Limitation of the study

This study contains some limitations as follows.

First, the findings of this study were based on self-report data collected using the instrument developed for the study, and the information obtained from the respondents about their grading practices are considered as teachers' actual practice of grading.

Second, the subjects of the study were limited to in-service teachers who teach in the Islamic private schools for both academic and religious subjects.

Third, only content and construct validity of the Grading Practice Inventory(GPI) were examined in the process of developing the instrument.

Finally, the study investigates teachers' practice of grading from different areas of disciplines, although the differences in grading may exist between teachers of different levels of training or qualifications.

1.8 Definitions of Terms

In this subsection, several operational definitions of terms are introduced. The terms are defined exclusively for this particular study.

Grading practice. Grade refers to students' grades assigned by teachers to represent students' summative performance representing their learning ability and knowledge in a particular course or subject. Grading practice refers to the process that teachers use to determine students' grades at the end of a course for each subject. It involves their considerations on the assessment methods employed, the factors taken into account when determining students final grades, and the processes involved in carrying the grading procedures, and the behaviors involved in grading students' work. In this study, these processes were measured by the score that teachers rate on self reported items in the Grading Practice Inventory (GPI) employed in the study.

Validity refers to the ability of the items used in the Grading Practice Inventory (GPI) to accurately measure and reflect the content or concept, the construct being measured, examined by a panel of judges and statistical analysis.

In this current study, the content validity was based on evaluation by a panel of measurement experts, while construct validity was evaluated statistically following the Rasch Measurement Model analysis.

Reliability refers to the stability or consistency of scores obtained from the Grading Practice Inventory (GPI) . The stability is based on the likelihood that “same person hierarchy” would result if the same group were given a different item measuring the same construct, and the likelihood that the “same item hierarchy” would result if the same items were given to different groups of persons. (Bond & Fox, 2001)

Religious Teachers refers to teachers teaching subjects in religious stream, which include Al-Quran, Hadith, Tauhid, Feqah, Seerah, Tafsir, Arabic Language, Malay Language in Islamic Private Schools in Thailand.

Academic Teachers refers to teachers teaching academic subjects, which include Mathematics, Science, Thai Language, English Language, Technology, Living Skills, Social Studies, and Physical Education in Islamic Private Schools in Thailand.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

The following sections provide a review of literature focusing on educational assessment and the common concepts related to classroom assessment, particularly the concept of grading practice, including the grading approach, guiding principles of grading, impact of grading practice on students' achievement, and measurement of teachers grading practices. The literature also discusses the psychometric properties in developing a scale, the theories related to the main focus of the study, and the proposed theoretical and conceptual framework for this study.

2.2 Educational Assessment

The term assessment has so far been used broadly; it covers other terms such as measuring, testing, examining, and evaluating, i.e. the terms involving the gathering of information about students' learning. All these terms are often used interchangeably. Linn and Miller (2005), however, stressed that assessment is more comprehensive and inclusive than measurement or testing, because the term measurement is limited to quantitative information about students' learning, whereas the term assessment includes both the quantitative and qualitative description of students' performance.

Assessment is considered as a process of gathering information of students' learning to help teachers in their decision making in teaching (Airasian and Russell, 2001; Banks, 2005; Linn and Miller, 2005). There are various techniques of measuring students' achievement in assessing students' learning. Teachers can use

formal assessment techniques to obtain information from students' involving conventional testing methods such as multiple choice, true-false, short answers, matching or essay items, or by using informal assessment techniques such as asking questions during teaching, or observing students while they are working in class, or assigning students with certain tasks outside class that will demonstrate their abilities. All those methods of assessment will help teachers in making better decisions about helping students to learn (Cunningham, 1998; Linn and Miller, 2005).

In general, educational assessments can serve a variety of purposes. When a teacher decides to operate any assessment of students' learning, he or she has to clarify the purpose of gathering the information as the first step (McMillan, 2001), because clearly identifying the purpose of a particular assessment type helps teachers to determine the appropriate technique of assessment to be used (Hogan, 2005).

Primarily, the assessment has been run to certify students' learning. For example, the end of a unit and final exams are intended mainly to certify students' learning. Consequently, it often results in the assignment of a grade when it is being pursued (Airasian, 2012 ;Hogan, 2005). The assessment for placement purpose have taken place when teachers decide to make decision about the placements of their students in a particular group or ranking students' performance for a particular use.

In designing instruction activities, a teacher may need to examine the students' status of knowledge and achievement in order to plan for their action. The action might be setting an appropriate level of instruction or varying the approach to

instruction. For example, a teacher might administer a reading test not for assigning grades but to identify the students' reading skills that need to be improved. This assessment is conducted for the purpose of planning instruction (Airasian, 2001; Hogan, 2005). Furthermore, providing feedback to students is an important purpose for classroom assessment, because its information not only can help in guiding their learning but also inform parents and schools administrators about the students' progress. In order to provide such feedback, teachers should assess students' learning and behavior constantly (Airasian, 2001; Hogan, 2005). In addition, providing feedback to students is a way for teachers to communicate the assessment information that they can understand and will help them to think about what should be done in the next step. For this reason, providing feedback to students is claimed to be the primary purpose of the assessment (Musial et al., 2009).

In some cases, teachers need the data from students' assessment in order to identify students with learning difficulties or social problems in the classroom to carry out the remedial activities needs. This diagnostic assessment allow teachers to develop better understanding of the Frequently Misunderstood Concepts (FMCs) that usually occur in a particular subject area to consequently determine the way to help the student understand the concept correctly (Linn & Miller, 2005; Musial et al., 2009).

In many circumstances, assessment can also be utilized to monitor and predict students' success in the future. Results of the assessment can be used to monitor the students' progress related to the learning target overtime (Newton, 2007). Consequently, this information can be used to predict future success of students.

The prediction purpose of assessment is often closely related to the planning function, because some teachers may want to predict which student are likely to be successful in advance in particular courses or next level of education for better planning and guidance to students. Furthermore, prediction plays a significant role for college or university admission (Hogan,2007).

Assessment can also motivate students' learning. According to Brookhart, Walsh and Zientarski (2006) findings, student's perception of difficulties, importance, and interesting facts will keep them making efforts to accomplish the assessment task. Brookhart and Durkin's (2003) discovered that the strongest motivation for students is to accomplish the assessment goal that are assigned by the teachers with good grade, to learn for their own sake, and to show what they had learned. Additionally, when the content of an assessment were interesting they wanted to learn more; for instance, Rodriguez (2004) found that the assignment of the homework has strong effects on students' performance on the TIMSS. Thus, it is evident that assessment is a significant predictor for students' motivation to learn.

In short, the educational assessment results have been used for various purposes depending on the user. For students, teachers and parents, results of assessments are usually used for guidance, monitoring students' progression, and placing or ranking the students' performance, and information transfer for a new class or school. On the other hand, administrators, researchers and policy makers used assessment information for system monitoring, program evaluation, qualification and selection, as well as organizational intervention and national counting (Newton, 2007).

In practice, educational assessment can be formative and summative. formative assessment is utilized to examine learning progress in the time of teaching in order to improve learning and instruction. Thus, providing feedback about learning success and failure to both students and teachers is the key point of its purpose (Linn and Miller, 2005). Garrison and Ehringhaus (2007) affirmed that formative assessment is part of instruction. For instance, some instructional strategies can be used formatively such as, establishing and defining quality of work together with students which can help them to understand and know the learning target and criteria to achieve it, and asking better questions can help students deeply think and provide teachers deeper understanding. Hence, by observing students working during the class, teachers can gather evidence of students' learning to inform instructional adjustment, (Garrison & Ehringhaus, 2007).

Miller et al., (2012) stated that not only teachers-made tests and assessments can be used for formative function; standardized testing and others instructional material customized by publishers might as well serve these functions. Sometimes, formative assessment involves collecting information concerning students' need in specific concept and skills, a short-selected response or a set of questions that focus on a specific concept or skill work well for this case. Teachers can also use assessment in the form of competitions to challenge the students with the increasingly more difficult task; for example, teachers conduct a spelling competition to indicate the best spellers in the class (Banks, 2005)

According to Brookhart (2001) formative assessments occur only when the assessment information is used to improve students' performance. Due to this, the

formative process must be done by both teachers and students simultaneously, thereby, students must be able to use the assessment information to improve their learning goal by comparing the actual scores with the desired performance and work harder to close the gap. At the same time, teachers should provide helpful feedback to students. Thus, the role of the students and teachers in improving the performance is the central role for formative assessment (Brookhart, 2001). Garrison and Ehringhaus (2007) noted that assessments will not be formative if the students are not involved in the assessment process; without students' involvement, the formative assessment is not implemented its full effectiveness. Thus, it can be characterized that formative assessment has to be part of instruction and the process of improving the progress and effort of learning; it provides a valid and useful diagnostic information for teachers and students, encourages the students to be active and plays a role in their own learning (Harlen & James,1997).

On the other hand, summative assessment intend to obtain the information on what the students have learned during the period of instruction of a particular course or subject in relation to the curriculum or learning targets (Musial et al.,2009); therefore, summative assessments usually take place at the end of instruction, chapter or unit. It is used to verify how much or how far the instructional goals have been reached. It also can be used for assigning a course grade or the certification of students' achievements, such as, judging students who can pass in a particular course or who have skills in a particular area. The final exam is a familiar example for this assessment type (Banks, 2005; Linn & Miller, 2005).

Brookhart (2001) stated that the summative assessment shows the over view of learning outcomes by collecting evidence over time, at the end of the instruction, or at any flexible period. Assessments for this purpose can be done with any technique including a teacher-made test, standardized test, performance test, portfolio, laboratory, oral report, paper or research report, because all these assessment task can be used to show the students' ability and achievement (Miller et al., 2012).

Generally, summative assessment is more formal and systematic than formative assessment in terms of the procedure in gathering the evidence of learning outcomes, because this type of assessment is usually administered at the end of instruction such as final examinations, or the end of a unit or chapter of teaching, such as the end-unit achievement test; thus, evidence of learning achievement that was gathered from a test or other assessment task is interpreted as the acquirement of skills or knowledge, understanding, and attitudes. This interpretation of achievements is based on the same criteria for every student in order to obtain the comparability across students . It can be noted that summative assessment occurs when the teaching process was ended and its result is interpreted based on criterion-reference because it is relevant to learning criteria, and its technique should be reliable and systematic procedure (Airasian, 2001; Harlen & James,1997).

In summary, educational assessments are the process of gathering information of students' learning through various methods including testing, observing, asking, interviewing, and so on both informal and formal to provide both formative and summative proposes.

2.3 Grading Practice

In general, the summary of classroom assessment has been assigned in a summative statement format such as a letter grade (e.g. A,B,C,D) to indicate the student's attainment. Thus, it seems that the assessment and grading are closely connected. Lekholm (2008) concluded that grades are a summative measure gained from several assessment events and functions explicitly as an indicator of students' achievement, a selection instrument of education system, and an instrument of evaluation. Additionally, grade can also motivate students' learning (Stiggins, 2008).

Educators defined grading as the assignment of symbolic numbers or letters at the end of a specified period of time that will serve as a summary statement of evaluations of the students (Marzano, 2000). It is the process that translate assessment information into marks or letters that indicate the quality of each students' learning and performance (Airasian, 2001).

Basically, the primary goal of grading is to provide high quality feedback to parents and students in order to make them understand and appropriately use the information to support the learning process and encourage student success (Airsian, 2001.) Mostly school grading serves a variety of functions, based on how the grades are used, such as to improve students' learning or report students' accomplishment, to rank students' achievement, to indicate progress toward clearly defined targets, to inform parents , to evaluate teachers, for guidance and administrative uses (McMillan, 2001).

Educators agreed that the primary reason for grading is to provide students, parents, and others with information about the achievement of learning objectives,