DEMOGRAPHIC FACTORS AFFECTING SECONDARY MATHEMATICS TEACHERS' CONCEPTIONS OF ASSESSMENT

BIBI MAHDIEH MIRIAN

FACULTY OF EDUCATION
UIVERSITI MALAYA
KUALA LUMPUR

2019

DEMOGRAPHIC FACTORS AFFECTING SECONDARY MATHEMATICS TEACHERS' CONCEPTIONS OF ASSESSMENT

BIBI MAHDIEH MIRIAN

THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF MATHEMATICS EDUCATION

FACULTY OF EDUCATION
UNIVERSITI MALAYA
KUALA LUMPUR

2019

UNIVERSITY OF MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: **Bibi Mahdieh Mirian**

Matric No: **PGM 150004**

Name of Degree: **Master of Mathematics Education**

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"):

Demographic Factor Affecting Secondary Mathematics Teachers' Conception of Assessment

field of study: **Mathematics Education**

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work, I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date:

Subscribed and solemnly declared before,

Witness's Signature Date:

Name:

Designation:

ABSTRACT

Assessment of student learning is a vital step in the mathematics teaching and learning process. Data obtained from assessments enable teachers to reflect on and make improvements to their practice. This single embedded quantitative study explores teachers' conception in the context of a new mathematics curriculum in Iran where major educational reform has been undertaken. Thus, the main purpose of this study was to investigate in-service Iranian mathematics teachers' conception of assessment at three grades of lower secondary school 7^{th}, 8^{th} ,9^{th} in seven educational districts in Mashhad, one of the cities in the North East of Iran. In addition, to determining whether there are differences in teachers' conceptions of assessments according to three demographic characteristics of mathematics teachers such as years of experience, educational qualifications and gender. There were 352 mathematics teachers included in the study. The instrument used to conduct this study was the Teachers' Conceptions of Assessment III Abridged (COA-III, Brown 2006) which had 27 items. The findings of Multivariate Analysis of Variance (MANOVA) indicated that the highest level of agreement among mathematics teachers was student and school accountability purpose of assessment. The only significant differences were noted between educational qualifications and teachers' conception of assessment. The findings of this study contribute to the area of mathematics teachers' conception and assessment in Iran after the reform movement in the mathematics curricula. Besides, the findings suggested that professional development needs to be provided for teachers in order for teachers' conception to be compatible with the ideas in the proposed mathematics curriculum reform.

ABSTRAK

FAKTOR DEMOGRAFI YANG MEMPENGARUHI PEMAHAMAN GURU MATEMATIK SEKOLAH MENENGAH TERHADAP KONSEP PENTAKSIRAN

Pentaksiran pembelajaran pelajar adalah satu langkah yang penting dalam pengajaran dan pembelajaran matematik. Data yang diperoleh daripada pentaksiran membolehkan guru untuk mengimbas dan membuat penambahbaikan dalam pengajaran mereka. Kajian kuantitatif tertanam tunggal ini dapat meneroka konsep pentaksiran yang difahami guru dalam konteks kurikulum matematik baru di Iran, di mana pembaharuan pendidikan yang besar telah dijalankan. Oleh itu, tujuan utama kajian ini adalah untuk mengkaji kefahaman guru matematik dalam perkhidmatan di Iran terhadap konsep pentaksiran untuk tiga gred menengah rendah, iaitu gred 7, 8 dan 9 di tujuh daerah pendidikan di Mashhad, salah satu bandar di utara of Iran. Di samping itu, untuk menentukan sama ada terdapat perbezaan dalam pemahaman konsep taksiran di kalangan guru matematik, guru-guru dibahagikan kepada tiga demografi iaitu tahun pengalaman, kelayakan pendidikan dan jantina. Terdapat 352 guru matematik yang menyertai kajian ini. Instrumen yang digunakan untuk menjalankan kajian ini adalah "*Teachers' Conceptions of Assessment III Abridged*" (COA-III, Brown 2006) yang mempunyai 27 item. Hasil Analisis Varians Multivariat (MANOVA) menunjukkan bahawa persetujuan paling tinggi di kalangan guru matematik adalah konsep kebertanggungjawaban pelajar dan juga sekolah. . Satu-satunya perbezaan yang ketara adalah kelayakan pendidikan dan pemahaman guru terhadap konsep taksiran. Hasil kajian ini menyumbang kepada pemahaman guru matematik di Iran tentang konsep pentaksiran selepas pembaharuan dalam kurikulum matematik. Selain itu, penemuan mencadangkan bahawa pembangunan profesional

perlu disediakan untuk guru-guru supaya konsep tersebut serasi dengan idea-idea dalam pembaharuan kurikulum matematik yang dicadangkan.

ACKNOWLEDGEMENTS

I would never be able to finish my study without support from My family, I must express my very profound gratitude to my dear sister Banafsheh, dear Mom and dear Bahareh, finally my two brothers -in- law; dear Adeel and dear Ali, for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them.

I would like to express my deepest gratitude to my supervisor Dr. Hutkemri. His support and guidance throughout this process helped make this dream come true. His willingness to proofread all of my work and patiently help me work through all of the milestones that need to be accomplished was completed with encouragement in his voice.

I would also like to acknowledge Dr Suzi Head of Mathematics Department and also Dr Norjo as readers of this thesis, and I am gratefully indebted to them for their very valuable comments on this thesis.

Finally, thanks to Faculty of Education, Mathematics Department and University of Malaya for their support that made this journey possible.

TABLE OF CONTENTS

Chapter 3 Research Methodology

Chapter 4 Findings

Chapter 5 Discussion and Conclusion

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

AMS	:	American Mathematical Society
ANOVA	:	Analysis of Variance
ARM	:	Assessment Reform Movement
COA-III	:	Conceptions of Assessment Abridged Version
ERM	:	Education Reform Movement
FRDE	:	Fundamental reform document of education
INC	:	Iranian National Curriculum
MANOVA	:	Multivariate Analysis of Variance
MOE	:	Ministry of Education
MOST	:	Ministry of Science and Technology
NCTM	:	National Council of Teachers of Mathematics
NOFET	:	National Organization for Educational Testing
OECD	:	Organization for Economic Co-operation and Development
PISA	:	The Program for International Student Assessment
TIMSS	:	Trends in International Mathematics and Science Study
TCoA IIIA	:	Teacher Conceptions of Assessment Abridged Scale

LIST OF APPENDICES

Chapter 1

Introduction

1.1 Background of the Study

Teachers' conceptions and how they impact their decisions and actions have been the focus of a large body of research for many years (Fives & Buehl, 2012; Pajares, 1992; Woolfolk Hoy, Davis, & Pape, 2006). These conceptions held by teachers about specific concepts are used for filtering useful information, framing problem spaces as well as guiding their actions (Pajares, 1992).

Various definitions and terms have been used to describe the construct of conceptions in the literature on teachers' conceptions. For instance, Brown (2002, 2004) and Thompson (1992) have used 'conceptions' instead of 'beliefs'. Thompson (1992) stated that conceptions reflect "a more general mental structure, encompassing beliefs, meanings, concepts, propositions, rules, mental images, preferences, and the like" (p. 165). By using the term 'conceptions', it is possible to combine 'knowledge' and 'beliefs' into a single construct, which can resolve some theoretical and operational problems encountered in attempting to differentiate these two concepts (Barnes, Fives & Dacey, 2017).

Assessment of student learning is a vital step in the teaching and learning process. Data obtained from assessment enable teachers to reflect on and make improvements in their practice. Regarding the significance of education and enhancing student learning, teachers have to develop assessment strategies providing the essential information to make instructional enhancements. Therefore, understanding teachers' conceptions of assessment, with regard to their different characteristics such as years of experience or their gender is necessary.

Regarding the concept of assessment, different beliefs and conceptions about the aims of assessment from a teachers' point of view can affect how new information about principles of assessment is processed or how the situations of examining, developing or selecting assessments are framed. Different conceptions about assessment may also be useful as a guideline for teachers in how they affect their decisions on other practices they use in the classroom such as teaching techniques and motivational strategies.

Teachers' conceptions are influenced by the school, school district, and community in which they are situated (Ernest, 1988). It can also be affected by the personal factors such as gender. Furthermore, due to having differing conceptions about mathematics and its teaching and assessment, teachers in the same school are often observed to adopt different classroom practices (Ernest, 1988).

A study by Herrington, Herrington, and Glazer (2002) on mathematics teachers showed that teachers' conceptions mostly reflect into classroom practices and assessment. Related studies also have identified several factors including personal and environmental factors, external pressures, culture, teachers' reflection level and teachers' experiences that act as influences or constraints to teachers' instructional practices in general or to implementing alternative assessment practices in mathematics (Borko & Cumbo, 1993; Cooney & Shealy, 1995; Ernest, 1988; Herrington et al., 2002).

In 1989, the National Council of Teachers of Mathematics (NCTM) outlined a proposal for mathematics that led schools into the 21st century and emphasized procedural skills, real-world uses of mathematics and multi-step problem solving. Later, NCTM (2000) regarded assessment as a powerful tool for learning mathematics that is able to bridge the gap between teaching and learning with the aim to improve the quality of mathematics education.

Therefore, it was essential to change mathematics assessment practices in such ways that support and are in line with other ongoing changes in the education of mathematics. As a matter of fact, if traditional practices continue to dominate, any success in mathematics reform seems unlikely. Considering this issue, many countries started to reform mathematics education and assessment in order to meet the demand of 21st century skills and eventually, in 2011, although it seemed late, the mathematics reform movement was launched in Iran.

This reform movement completely restructured the Iranian school system, affected textbooks, and fundamentally changed the educational goals in the system. It specially had substantial influence on teachers' role. As in other reforms around the world, Iranian mathematics reformers requested teachers to be facilitators during the mathematics teaching and learning process and suggested teachers to give students opportunities to engage in conversations which lead to solutions, give substantive feedback rather than points or grades on assignments. Mathematics assessment must encompass a wide variety of methods, encourage teacher-student interaction, allow for different learning styles, promote problem-solving and self-confidence, and promote active involvement in learning (NCTM, 2014; Reyhani, 2015).

According to Brown (2004) teachers' conceptions influence the view and attitude of teachers with respect to assessment. Therefore, the context in which teachers understand the assessment should be taken into consideration. Therefore, it is important to understand how assessment is conceived by Iranian in-service mathematics teachers and how they practice assessment approaches (Brown, 2002) to improve the quality of learning and teaching and finally improve quality of the educational system.

1.1.1 The Iranian Context

Education in Iran is still very centralized, and it consists of several levels. Education in K-12 and the higher education are controlled by the Ministry of Education (MOE), and generally, higher education is controlled by the Ministry of Science, Research and Technology (MSRT). The structure of school was formerly 5 years of primary school, 3 years of guidance school, 3 years of high school and one year of pre-university course. Out of these different educational levels, the last level that includes four years is not obligatory. In fact, this level is divided into three different fields, namely theoretical, vocational/technical, and manual fields. Each one of these fields has its own specific courses.

The Iranian educational system consists of both public-school system as well as private school system for every one of these levels, starting from primary schools up to universities. The Ministry of Education (MOE) runs the planning system, financing, administration, curriculum, as well as compilation of textbooks. This organization takes care of training teachers, giving marks, holding the exams, and so forth. The basis for the exam questions are the textbooks (Arani, Kakia, & Karimi, 2012).

In 2011, the reform movement introduced a new structure of education; Pre-university year was replaced with an additional year in elementary school. In this way, the structure of school education includes 6 years of primary and 6 years of secondary school and each academic period is divided into two periods of three years. The first nine years is devoted to general education and the final three years of high school at the end of the second period are dedicated to semi-skilled specialized education, whereby students will be awarded a diploma. The second-high school has three branches: theoretical, technical and professional. Each branch of the course consists of a variety of disciplines (FRDE, 2011; Reyhani, 2015).

The new curriculum changed the approaches to mathematics education from behaviouristic view to constructivist view. All mathematics textbooks in all grades have been rewritten. The main feature of the new curriculum is its approach to teaching. Prior to this curriculum, problem solving was considered as an outcome of learning, but the recommended approach to teaching and learning in the new curriculum encourages teachers to engage students in group or in-class as well as problem-solving activities, as well as considering different learning styles when presenting a topic. The following areas are emphasized by the Ministry's guidelines for mathematics: a) content including basic mathematical concepts and facts, b) mathematical processes including real-life situation modelling, problem solving, hypothesis-making and evaluation, reasoning and estimation, and c) general skills including observing, abstracting and generalizing, creative and critical thinking, comparing and ordering, and classifying and sorting (Reyhani, 2015).

Assessment and evaluation in the Iranian education system is not voluntary for each level. The evaluation methods are very fundamental. In the teaching system, the same course books are used all over the country in each and every school. In the grade five of primary school, in grade three of secondary school and in grade four of high school, nation-wide exams are held to evaluate students' levels of acquired knowledge. After reform in the new system the national examinations are held in the sixth (primary school), the ninth (lower secondary) and the eleventh (higher secondary) grades. For the rest, teachers have a responsibility to design and report the result of the exams.

The national examinations' results determine placement of the students into different levels of education; it can be said that Iran is in the high-stakes assessment context category. The source for the content of these exams is the textbook. This leads to the increasing reliance of the teachers on the textbooks for the content of their teaching. However, beside the textbook, teachers have guidebooks for teaching and also workbooks for extra practices (Arani et al., 2012).

In Iran, classroom assessment is considered a tool for making judgments about the quality of schools, teachers and students. In fact, families of the students and also the people in charge of schools share this opinion that better schools help students have better achievement on standardized exams. Perhaps the teachers are mostly concerned with evaluation ideas. This is because they know very well that the marks of the students show the quality of their teaching methods (Saad, Sardareh, & Ambarwati, 2013).

In Iran, the marking system is from 0 to 20. The students are required to get minimum 10 to proceed to the higher classes. The system is approximately the same as that of the US where: A=17-20, B=14-16/9, C=12-13/9, D = 10-11/9, F=below 10. The Iranian people speak Persian as the national language. Therefore, using local and tribal languages in the media, and also teaching their literature are considered to be big achievements in this country (Arani et al., 2012).

1.2 Statement of the Problem

The quality of education depends on students' academic achievements which are mostly based on the routine assessments. In Iran, among different curriculums, mathematics has a special place in academic planning as an important success factor for the students seeking academic and technical education at university level.

According to Pangeni (2014), various factors affect academic achievement as well as students' progress in math, more importantly student characteristics, teacher characteristics, family environment, and school environment.

Teacher's characteristics can make education a process of pleasure and prosperity, or a barren procedure. In addition to having knowledge about the subject, mathematics teachers require a series of skills, attitudes and conceptions that are necessary for effective teaching and academic achievement of students (Fung et al., 2017).

Previous studies showed that teachers' conception of different parts of educational process including assessment plans could affect the success of the entire educational program (Gage, 2001; Delors, 2013; Carneiro, Looney, & Vincent, 2015). Furthermore, over the past few decades, studies on the teachers' conceptions of the assessment process, especially as a result of the critique of the ineffectiveness of teachers' educational process have become an important issue for teacher training for education authorities (Kaiser, Südkamp, & Möller, 2017).

In Iran, considering the significance of the conceptions of assessment among the mathematics teachers, one of the important issues is the lack of data and information on the teachers' conceptions of assessment and examinations. It is especially important in the new assessment process focusing on improving the quality of mathematics education through modifying the learning and teaching process with an emphasis on using new approaches of assessment and practices.

As there is no doubt that teachers play a key role in driving the education reform movement (ERM), their responsibility will be increased considerably in comparison with the past. They are expected to use assessment information 'effectively and professionally' to enhance learning and fine-tune their teaching practice that could easily be affected by their conceptions about this new approach (Leahy, Lyon, Thompson, & Wiliam, 2005).

Besides, as mathematics teachers' conceptions are based on a set of beliefs and body of mathematical knowledge that can influence their teaching approach, including the assessment process, determining the teachers' conceptions could be one of the key factors to better understand the teachers' definition of their practice.

Studies on different educational reforms have shown that in most cases, because of teachers' limited knowledge and poor conceptions about a new approach of practices or the mismatch between teachers' conceptions and new curriculum goals, the expected outcomes are not achieved (Handal & Herrington, 2003; Pfannkuch, 2001; Senk, Beckmann, & Thompson, 1997; Susuwele-Banda, 2005). Furthermore, in new educational approaches such as the improved assessment process, teachers often have faced a conceptual dilemma regarding adopting the new approach that may require more attention (Adams & Hsu, 1998; Brown, 2004; Earl & Timperley, 2014; Suurtamm & Koch, 2014; Webb, 2012; Windschitl, 2002). This condition grounds considerable pressure since it requires teachers' understanding of the logic for 'improvement-oriented' assessment and having the knowledge, skills, and resources for the realization of this policy reform.

Therefore, if policy makers aim to succeed in implementing any new assessment practices or tools, whether in local or national level they should be aware of the complex structure of teachers' conceptions of assessment (Brown, 2004). Furthermore, if teachers' conceptions remain unchanged, it may reduce effectiveness in implementation of the new policy (Brown, 2004). Consequently, as teachers are the last step in a sequence of educational reform, it is critical to take into account teachers' conceptions concerning their present practices in order to realize how they could possibly comprehend, reply to, and accomplish the reform.

It is not always easy to identify and determine teachers' conceptions because of the different factors effecting human conceptions and beliefs, in this case, the different years of experience and academic qualification, gander as well as grades taught that could affect the teachers' conceptions of assessment (Abdollahi, Dadjooye Tavakkoli, & Youseliani, 2014). Different assessment conceptions lead to different practices. For example, teachers who see assessment as a means of collecting data to base their decisions about learning and their teaching methods on, will attempt to integrate assessment into their teaching.

Only a few studies have investigated teachers' conceptions of assessment, specifically on the new approach, so the information in this area is very limited. Therefore, the current study is aimed to determine in-service Iranian mathematics teachers' conceptions of assessment at three grades of lower secondary school (7th, 8th, 9th grades) in seven educational districts. Understanding the mathematics teachers' conceptions of assessment could explain the gap between the goals of the new assessment approach in the educational system in Iran and its inadequate outcomes. To the best of our knowledge, this is the first study that aims to investigate mathematics teachers' conceptions of assessment in Iranian lower secondary schools.

The model of Brown (2002) for teachers' conceptions of assessment guided this study. The model will be discussed in the following section.

1.3 The Conceptions of Assessment Model

Brown's (2002) model for teachers' conceptions of assessment is the underlying model of this study. Brown proposed the four-factor model in order to gather information about how teachers conceived of assessment regarding the nature and purpose of assessment. The model is explained briefly as follows:

First, regarding the terms used in this model, Brown used “conception” instead of “beliefs”. As mentioned before, conceptions are believed to reflect “a more general mental structure, encompassing beliefs, meanings, concepts, propositions, rules, mental images, preferences, and the like” (Thompson, 1992, p. 165). Thus, it is possible to combine ‘knowledge’ and ‘beliefs’ into a single construct by using the term ‘conception’.

Second, teachers’ conceptions about the purposes of assessment is classified by Brown into four major inter-correlated factors, which could be grouped into two further categories: “purpose” including categories 1 to 3 and “anti-purpose” including category 4 only. These categories are as follows: 1) improvement: assessment improves learning and teaching process, 2) school accountability: assessment makes teachers and schools accountable for learning, 3) student accountability: assessment makes students accountable for their learning, and 4) irrelevance: assessment is irrelevant or pernicious to the work of teachers and the life of students.

The first conception refers to this idea that assessment is a useful tool for improving students’ individual learning and the teachers’ quality of teaching. This conception is thoroughly associated with formative assessment. According to Brown (2002) the main purpose of improvement conception is to assess the knowledge, skills, performance, or understanding of the students in order to produce precise information which is helpful to make valid changes in learning and teaching processes and improve students’ achievement (Brown, 2002). In fact, the improvement conception is the fundamental aim of the mathematics educational reform.

The second conception of assessment points out that the purpose of assessment is making the teachers and schools accountable for student achievement; hence, poor outputs have risky results for schools, teachers or the educational programs. The idea of accountability conception about assessment is very much effective on the teachers' performance and also on the management system (Brown, 2002).

The third conception of assessment is aligned with the idea of learners' individual accountability for their learning. The fact that they are categorized into various classes, groups, and that they have to take various exams testify to this claim. The ways this idea is carried out include giving marks to the students' performance, issuing them certificates according to their performances, and decision-making in different situations (Brown, 2002).

The fourth conception implies that assessment is irrelevant to these claims: a) assessment gives equivalent results with testing, therefore, exams are inappropriate for education, or 2) assessment causes teachers, schools and students to become accountable for what students achieve; hence, this is bad for the quality of learning. It means that assessment is unrelated to the lives of teachers and neither is it to those of students. So, systematic evaluation of students has no rational place in the educational objectives. This could be because systematic evaluation is not good and fair and causes anxiety, and this is also because it damages students' self-esteem (Brown, 2002).

There are three basic assumptions in Brown's model for the mathematics teachers' conceptions of assessment in their classroom:

1. It is assumed that the conceptions of teachers function like a filter through which they construe and experience any nationally sponsored or school-wide policy and practice; specialized progress activities are even connected to assessment.

2. It is assumed that making decisions concerning persons, separately or in groups is notified and influenced by assessment.

3. It is assumed that the quality of the educational decisions is influenced by the quality of data obtained through assessment (Brown, 2002).

1.4 Research Purpose

The main goal of this study was to examine in-service Iranian mathematics teachers' conceptions of assessment at three grades of lower secondary school (7th, 8th, 9th grades) in seven educational districts. In addition, an attempt was made to determine whether there were differences in mathematics teachers' conceptions of assessment according to three demographic characteristics of them including years of experience, educational qualification and gender. This study aimed to achieve the following objectives:

1.5 Research Objectives

The research objectives of this study are:

1- To determine mathematics teachers' conceptions of assessment.

2- To determine the differences in teachers' assessment conceptions based on their years of teaching experience

3- To determine the differences in teachers' conceptions of assessment according to their academic qualification

4- To determine the differences in teachers' conceptions of assessment according to their gender

5- To determine the differences in teachers' conceptions of assessment according to grades level taught.

1.6 Research Questions

Based on the research objectives, the following research questions are posed in the present study:

1- What are the Iranian mathematics teachers' conceptions of assessment?

2- Do Iranian mathematics teachers' conceptions of assessment have significant differences based on their years of teaching experiences?

3- Do Iranian mathematics teachers' conceptions of assessment have significant difference based on their academic qualification?

4- Do Iranian mathematics teachers' conceptions of assessment have significant difference based on grade taught?

5- Do Iranian mathematics teachers' conceptions of assessment have significant difference based on gender?

1.7 Research Hypotheses

The following hypotheses were formulated based on the research questions of the study:

$H_{0\ 1.1}$: There is no difference in mathematics teachers' conceptions of assessment in terms of improvement conception based on years of teaching.

$H_{1\ 1.1}$: There is difference in mathematics teachers' conceptions of assessment in terms of improvement conception based on years of teaching.

$H_{0\ 1.2}$: There is no difference in mathematics teachers' conception of assessment in terms of student accountability conception based on years of teaching.

$H_{1\ 1.2}$: There is difference in mathematics teachers' conceptions of assessment in terms of student accountability conception based on years of teaching.

H_0 1.3: There is no difference in mathematics teachers ‘conceptions of assessment in terms of school accountability conception based on years of teaching.

H_1 1.3: There is difference in mathematics teachers’ conception of assessment in terms of school accountability conception based on years of teaching.

H_0 1.4: There is no difference in mathematics teachers’ conceptions of assessment in terms of irrelevant conception based on years of teaching.

H_1 1.4: There is difference in mathematics teachers’ conceptions of assessment in terms of irrelevant conception based on years of teaching.

H_0 2.1: There is no difference in mathematics teachers’ conceptions of assessment in terms of improvement conception based on academic qualification.

H_1 2.1: There is difference in mathematics teachers’ conceptions of assessment in terms of improvement conception based on academic qualification.

H_0 2.2: There is no difference in mathematics teachers’ conceptions of assessment in terms of student accountability conception based on academic qualification.

H_1 2.2: There is difference in mathematics teachers’ conceptions of assessment in terms of student accountability conception based on academic qualification.

H_0 2.3: There is no difference in mathematics teachers’ conceptions of assessment in terms of school accountability conception based on academic qualification.

H_1 2.3: There is difference in mathematics teachers’ conceptions of assessment in terms of school accountability conception based on academic qualification.

H_0 2.4: There is no difference in mathematics teachers’ conception of assessment in terms of irrelevant conception based on academic qualification.

H_1 2.4: There is difference in mathematics teachers’ conception of assessment in terms of irrelevant conception based on academic qualification.

$H_{0\ 3.1}$: There is no difference in mathematics teachers' conceptions of assessment in terms of improvement conception based on grade taught.

$H_{1\ 3.1}$: There is difference in mathematics teachers' conceptions of assessment in terms of improvement conception based on grade taught.

$H_{0\ 3.2}$: There is no difference in mathematics teachers' conceptions of assessment in terms of student accountability conception based on grade taught.

$H_{1\ 3.2}$: There is difference in mathematics teachers' conceptions of assessment in terms of student accountability conception based on grade taught.

$H_{0\ 3.3}$: There is no difference in mathematics teachers' conceptions of assessment in terms of school accountability conception based on grade taught.

$H_{1\ 3.3}$: There is difference in mathematics teachers' conceptions of assessment in terms of school accountability conception based on grade taught.

$H_{0\ 3.4}$: There is no difference in mathematics teachers' conceptions of assessment in terms of irrelevant conception based on grade taught.

$H_{1\ 3.4}$: There is difference in mathematics teachers' conceptions of assessment in terms of irrelevant conception based on grade taught.

$H_{0\ 4.1}$: There is no difference in mathematics teachers' conceptions of assessment in terms of improvement conception based on gender.

$H_{1\ 4.1}$: There is difference in mathematics teachers' conceptions of assessment in terms of improvement conception based on gender.

$H_{0\ 4.2}$: There is no difference in mathematics teachers' conceptions of assessment in terms of student accountability conception based on gender.

$H_{1\ 4.2}$: There is difference in mathematics teachers' conceptions of assessment in terms of student accountability conception based on gender.

$H_{0\ 4.3}$: There is no difference in mathematics teachers' conceptions of assessment in terms of school accountability conception based on gender.

$H_{1\ 4.3}$: There is difference in mathematics teachers' conceptions of assessment in terms of school accountability conception based on gender.

$H_{0\ 4.4}$: There is no difference in mathematics teachers' conceptions of assessment in terms of irrelevant conception based on gender.

$H_{1\ 4.4}$: There is difference in mathematics teachers' conceptions of assessment in terms of irrelevant conception based on gender.

This study used a causal-comparative research design. According to Fraenkel, Wallen, and Hyun, (2012) in the study of cause or consequences of differences that already exist between or among groups of individuals more suitable design is causal-comparative research design. The data collection instrument in this study is the abridged version of Teachers' Conceptions of Assessment Scale (TCoA-IIIA) questionnaire. Mathematics teachers are selected by a simple random sampling method from population of mathematics teachers at lower secondary level in seven educational districts in Mashhad, Razavi Khorasan Province, Iran. The quantitative data are analyzed using SPSS Software, version 24.

1.8 Significance of the Study

This study will hopefully benefit several stakeholders in mathematics education including secondary school mathematics teachers, curriculum planners and developers in the Ministry of Education and university lecturers.

First, this study will benefit teachers, since the main participants of this research are mathematics teachers in lower secondary schools, the findings of this study will help them in examining their conceptions about assessment. In fact, having a better understanding of teachers' view about assessment may assist teachers to implement the assessment to better meet the learning needs of students. Next, the results of this study will provide useful information related to assessment conceptions that teachers need in Iran and other countries around the world. Finally, teachers will use the findings of this study to evaluate the effectiveness of the assessment they use to assess the learning process of students.

Second, curriculum planners and developers of the Ministry of Education (MOE) can utilize the data to develop assessment guidelines in education. In fact, the Ministry of Education is in charge of in-service teachers' professional development such as designing courses or training programmes in Khorasan Razavi province. According to Popham (2009) professional development is key to improving assessment practices. However, designing assessment courses for in-service teachers needs the information about what teachers are doing and what they are not doing well, so, results of this study can help the Ministry to develop courses or training programmes. Then, district authorities are encouraged to apply the findings of this study to create courses for training teachers on the various facets of assessment. Finally, the Iranian educational assessment system may add the findings of the current study to the existing assessment theory and practice that it uses as an outline to develop training courses and professional development programmes for teachers related to the use of assessment in their teaching practice.

Third, this study will benefit university lecturers as it will contribute to the body of literature involving teachers' perceptions and assessment. Many researchers conducted research related to the affective domain and classroom assessment. This study adds to the compilation of research about perceptions, beliefs and conceptions toward assessment in mathematics education. The findings of this study can be used for future studies in affective domain area and assessment, in other geographic locations and with elementary school teachers. By offering different views to problems in a different perspective, the academicians can change the study by increasing the sample or population of the study, changing the methodology of the study and using interviews. Finally, they can focus on different topics in the assessment content area.

1.9 Limitations and Delimitations of the Study

This study contains some limitations and a delimitation. The two limitations are related to the research design, and the data collection method. And the one delimitation of the study is related to the setting of the study.

The first limitation is related to the research design used in the study, namely causal-comparative which means that the research design was non-experimental, and the variables in the research could not be manipulated or randomly assigned (Creswell, 2013).

The second limitation is related to the data collection method used in this study, namely questionnaire. Any method or instrument contains certain weaknesses. Since the questionnaires are standardized, explaining ambiguous points in the questions is not possible. Piloting the questionnaire with a small group of teachers could partially solve this problem. Also, there might be lack of responsibility of some respondents in filling the questionnaire.

The delimitation is related to the setting, namely lower secondary school level in this study. This study focused on grades 7, 8, 9 only. Primary levels and other grades of secondary school 2 (high school) such as grades 10, 11, and 12 are not included in this study. Lower secondary level is highly significant since it links primary and higher secondary school.

1.10 Definition of Terms

This study contains several basic terms including assessment, classroom assessment, conception, new assessment approaches, professional development, mathematics reform and traditional assessment. The definition for each basic term is now described:

Alternative Assessment (new approaches to assessment). According to Wu and An (2007) alternative assessment refers to assessment methods that are 'non-standardized' and 'non-traditional'. For instance, alternative assessment refers to use of non-traditional forms of assessment such as portfolios and projects (Nasab, 2015). Further, Nasab elaborated that these two forms of assessment both emphasize the product and process of learning. Frey and Schmitt (2010) explain that alternative assessment or performance test is such an assessment that requires the performance of a skill or the production of a product. In this study, new approaches to assessment or alternative assessment refers to various forms of non-traditional assessment that concentrate on the process and product of learning in order to support and improve teaching and learning process.

Assessment. Assessment is an integral part of teaching and learning which involves the process of gathering, interpreting and recording information related to student progress in learning and the effectiveness of the teaching strategies (Huba & Freed, 2000). Assessment is seen as an umbrella term which includes diverse practices such as deciding, collecting and making judgements about evidence related to the goals of the learning being assessed (Harlen, 2012). In this study assessment refers to any act of interpreting information or data concerning the performance of students, collected through any of a multitude of means or practices (Brown, 2002).

Classroom Assessment. Classroom assessments refers to the form of tests or the teacher-made assessment which are built, supervised and marked by the teachers and are meant to assess the learning levels of students (Keaikitse, 2012). Classroom assessment includes several types of practices such as building evaluation methods like paper-pencil tests and performance measures, supervision, marking, and interpretation of the achievements in the evaluation. It also includes formation of marking stages, communication of evaluation results, as well as using them in making decisions about education. Classroom evaluation includes all these activities, but it is not limited to them (Z. Zhang & Burry-stock, 2003). In this study, classroom assessment refers to a range of teacher-selected or teacher-made assessments that provide opportunity for students to be active in the learning process. In addition, it includes acquiring information and formative feedback for decision-making within the classroom context with the aim of improving teaching and learning.

Conception. It is the organizing framework "by which an individual understands, responds to, and interacts with a phenomenon" (Brown, 2004, p. 204). Additionally, conceptions are "mental structures, encompassing both beliefs and any aspect of the teachers' knowledge that bears on their experience, such as meanings, concepts, propositions, rules, mental images, and the like" (Thompson, 1992, p. 165). In this study, based on Brown (2002; 2004), the conception of assessment represents the teachers' understanding of the way students' learning is examined, tested, evaluated or assessed and finally the term "conception" is used to refer to the organized structure of ideas and beliefs about the nature and purpose of assessment (Brown, 2002).

Professional Development. It refers to specialized education opportunities about learning brought up for teachers so that they can develop their skills, their knowledge as well as their classroom activities. Indeed, professional development refers to training opportunities in assessment concepts, skills, and instructional strategies that help to improve learning, so teachers should have the opportunity to attend these programmes or courses (Gates, 2008).

Mathematics Reform. It is a movement away from traditional teacher-centered mathematics education including lecturing, modeling, and drills toward a hands-on, problem-solving, and more student-centered approach (Bray, 2011; Ross, McDougall, Hogaboam-Gray, & LeSage, 2003).

Traditional Assessment. Traditional assessment refers to paper-and-pencil tests that consist of multiple-choice, matching, true/false, short answer/fill-in-the--blank or essay questions. They are frequently designed to measure understanding or knowledge (lower level thinking) and scoring of them is often objective (Frey & Schmitt, 2010).

1.11 Operational Definition of Terms

Academic Qualification. This is one of the independent variables of this study that included the highest academic degree of the teachers who participated in the study. This variable is classified into three levels as Bachelor's degree, Master's degree, and PhD degree. This variable is measured by the third question in the first section of the questionnaire: "What is your highest level of education?"

Conceptions of Assessment. Conceptions of assessment in Brown's model (2002, 2004, 2006) represent a teacher's understanding of the nature and purpose of the way students' learning is examined, tested, evaluated or assessed.

Grade Level. This independent variable included the level of study that teachers teach during one academic year. This variable had three categories or levels in this study: grade 7, grade 8 and grade 9. This variable is measured by the forth question in the questionnaire: “Which grade level you are teaching in lower secondary school?”

Gender. This independent variable included the distinction between male and female as reported by the respondents (teachers) in this study. This variable is measured by the first question in the questionnaire: “What is your gender?” and it has recorded in two categories: Male and Female.

Years of Teaching Experience. This independent variable referred to the years of teaching that teachers spent in schools. This variable is measured by the question, “How many years have you been teaching?” and is categorized into five levels as 1-5 years, 6-10 years, 11-15 years, 16-20 years and More than 20 years.

1.12 Summary

This study is presented in five chapters. Chapter One established the theoretical framework of the study. This chapter included several sections; the first section is background of the study that explained the research area and prominent issues related to assessment. The second section is the problem statement where the researcher selected one critical issue for the study with three justifications that included discussion. Next, in the theoretical section, the model used in this study and research questions of the study are explained based on the research objectives. Following that significance of the study is discussed and also limitations and delimitation of the study are justified. Finally, the key terms were clearly defined as they are used in this study. The remaining parts of this thesis are as follows: Chapter two provides the review of related literature, then, Chapter three examines the methodology used, followed by Chapter four which analyzes the data gathered in the study. Finally, Chapter five interprets the findings, provides recommendations, and outlines implications for the future studies.

Chapter 2

Literature Review

2.1 Introduction

This chapter is divided into fifteen main sections, namely introduction, education and assessment system in Iran, mathematics reform movement, significance of assessment and its purposes, definition of alternative assessment (new approaches to assessment), traditional assessment, classroom assessment, assessment, conception, the model of conceptions, conceptual framework, teachers' characteristics and assessment, mathematics teachers' conceptions of assessment and its influence on assessment practices**,** review of research findings using Brown's Model and summary. The first section contains an overview of the chapter. The second section explains the education system in Iran. This section consists of three other subsections, the history of education in Iran, current view on Iranian education and assessment in the Iranian national curriculum.

The third section comprises two subsections, mathematics assessment reform; and assessment reform around the world. Firstly, this section provides an explication of the mathematics reform movement, followed by a discussion about assessment reform in mathematics and presenting a discussion of assessment reform movements around the world. Next, the fourth section offers significance of assessment and its purposes. Next, the fifth section provides several definitions of alternative assessment in the literature and in the current study. The sixth section offers several definitions of traditional assessment in the literature and the current study. Following that, the seventh section includes definitions of classroom assessment and the eighth section contains definitions of assessment. The ninth section provides a definition of conceptions.

The tenth section contains the model for conceptions of assessment that provides an explication of the model employed in this study. This section provides an insight into Brown's model of assessment conceptions and the rationale for using the model. Next, in eleventh section, the researcher explains the conceptual framework of the study. In the twelfth section the literature on teachers' characteristics and assessment is reviewed, followed by the section that provides a review of teachers' conceptions of assessment and their influence on assessment practices in different studies. Then, in the fourteenth section, a review of the relevant findings reported in the literature concerning teachers' conceptions of the purposes of assessment based on Brown's model is presented. Finally, the last section is a summary that provides concluding remarks for the literature review.

2.2 Education and Assessment System in Iran

2.2.1 Historical View on Education in Iran

This subsection explains the history of education in Iran which has been divided into nine periods from 1851 to 2011. The significant changes in the education system in each period consisting of educational structure, textbooks, establishing of new departments and education systems are explained. Lastly, the researcher explains in detail the last educational reform movement from 2011 to the present.

The first period is from *maktabs* to the year 1851. The *maktabs* were religious schools and the teaching of primary education was presented traditionally and its objective was to give the pupils a brief introduction about ethics and alphabets. In this period, the difficulties of textbooks and the inconvenience for the mental abilities of children created problems for them. Most of them had to drop out of *maktabs* in the early years and those who succeeded to complete the educational period in *maktabs* either went after administrative jobs or continued with their studies in the religious schools in different subjects such as mathematics, medicine and astronomy. Education during these years was not limited to educational periods; writing books in different sizes was also popular informally in the form of handwriting (Samadi, 2001).

The second period is from 1851 to 1910. The beginning of this period coincided with the formal establishment of *dar-ul-funun by* Amir Kabir. It became an introduction to paying serious attention to formal education in Iran. This school was established according to the European educational style and scientific centers with the aim of introducing a period of European sciences and technologies to Iranians. By establishing this school, different fields of study were introduced according to the society's needs, and gradually, the need to educate people increased. Therefore, every organisation established a school that was dependent on the organisation itself; the course subjects were selected according to the requirements. After the Russia-Iran wars, the society's need for mathematics became more popular and following that, several people were sent abroad to study. According to the records and historical narratives, mathematics was taught by French teachers in *dar-ul-funun* and some schools were founded by foreign religious commissions; and the textbooks were a selection of the books published in France. In fact, with *dar-ul-funun*, modern education entered a new stage and this school became the precursor of formal education in Iran (Gooya, 2011; Hudaee, 2009).

The founding of *dar-ul-funun* was considered as a movement towards renewal and change, an outstanding point in Iran's history of education and the beginning of the formal educational system in the country. In 1893, the establishment of the Ministry of Culture, was a fundamental step taken for the development of formal and classic education, and from that date on, opening public and private schools started all over the country. During this period, the school had two levels, namely primary and secondary. The secondary level was for 5 years, and in its scientific section, according to the European and American styles, courses such as full accounting, algebra, trigonometry, geometry, geography, history, astronomy, physics, chemistry, botany, zoology, economics, sports and gymnastics, training military operations, and one of the three -French, English or Russian-- languages were taught (Hudaee, 2009).

The third period from 1910 to 1921 marked the establishment of the Ministry of Knowledge in 1911, which paved the way for the formation of formal education and the early forms of schools were divided into four levels called village primary schools, district primary schools, secondary schools and high schools. In the curriculums of these schools, three types of scientific education (teaching subjects in the classes), ethics (speeches in field trips), and physical education (sports and matches) have been proposed. Because the cultural ties between Iran and France were strong during that period, and because the Iranian government and the wealthy families sent their children to France for further education, the Iranian schools, too, were managed in the French style and based on the French curriculum. The textbooks, too, were mostly translated from French and published. In the first decade of the 20th century, some signs of tendency towards a formal school curriculum were observed in Iran (Hudaee, 2009). Gradually, the school levels were separated and after the Act of Knowledge was passed by the National Council of Parliament in 1911, the schools were divided into three levels of primary, secondary and high schools. After the establishment of the Constitution and the formation of Parliament, education received formal attention from the government and the three forms of scientific, ethics and physical education were formally named. In this year, the Minister of Sciences prepared a curriculum for schools for the first time, and this was the beginning of centralisation in the educational system. However, the school principals in different states and districts still personally interfered with the curriculum design. The books were compiled and produced according to the periods and the levels; and upon formation of the Association of Knowledge, the publication of course books started (Samadi, 2001).

During the fourth period from 1921 to 1941, the High Council of Culture was formed and undertook different affairs of the country's education such as approval of educational curriculums, statutes and different regulations. This was a fundamental step for unifying formal education in the country. On the instructions of the 6-year secondary school, according to the approval of the High Council of Knowledge in 1931, after the 6-year primary school, the secondary school was divided into two periods of three years each (first and second). Two fields of study namely science and arts were placed in the second period. In March 1934, according to a directive, the studying place of elementary courses was named *dabestan* (primary school) and its teacher was named *moalem* (teacher), and the studying place of secondary school was called *dabirestan* (high school) and its teacher was named *dabir* (teacher). The research findings show that during the first 15 years of this century, the High Council of Knowledge compiled and revised the educational curriculums and increased as well as decreased them several times, until the last years of this century, it reached an almost coherent curriculum. After much deliberations, it was found that there were three major defects in the educational system, namely too much insistence on the tradition, insistence on religion, and the dominant theoretical nature. From then on, specific attention was paid on vocational education with the aim of limiting theoretical growth of the educational system, and according to the plan of the second half of 1951, it attempted to make education milder and get it aligned with the living requirements (Gooya, 2011).

The fifth period was from 1941 to 1983. During this period, World War II had started and had weakened the government which made the government unable to compile and publish books, and this created an opening for activities of free publishing. In 1955, under the advice of the American counselors, an office called the Headquarters of Studies and Curriculums was formed which took over the responsibility of making preparations for planning curriculums and projects, organising the required coordination for changing the educational system, and evaluation of the curriculums. Secondary schools were divided into two 3-year periods. During the first three years students remained in the general form and in year 4 students were divided into several different fields, namely literature, mathematics, agriculture, technical studies, commerce and housewifery (for girls). This system known as the old system of education in secondary schools was present until 1974-75 which was the first year of the New Educational system. Theoretical education started from three main branches of literature, and mathematics; according to the May 1958 Statute, ratified by the High Council of Culture, the fields of vocational studies included subfields of metal works, car mechanics, electricity, carpentry, building, painting and textile. The new curriculum had a more applied and more moderate content; it had become closer to the life requirements. In addition, except for the subject of housewifery that belonged to girls, there were no differences between the education for girls and boys. However, this curriculum still received harsh criticism; because the main cause of this problem was the fact that it was an imported curriculum which did not conform to the Iranian conditions and characteristics. Another reason was that this education belonged to the special social classes. In March 1962, the approval of the Board of Ministers, the compilation and preparation of course books were exclusively allocated to the Ministry of Education and the Iranian Course Books Organization was thus formed (Samadi, 2001).

The sixth period was from 1963 to 1978. Since 1963, course books were compiled with the cooperation of a group of university professors, teachers from the Education Department and experts from the Ministry. Some activities were carried out in order to reform the educational system and to synchronise it with scientific changes and technologies. These were also carried out in order to fulfill economic and social expectations. Eventually in 1966, the new educational system was established, and in 1966-67, it was started in primary school. Graduates of this period went to secondary schools from 1971-72 after five years; and after spending a three-year sojourn in secondary schools, in 1974-75, they started high school in three different fields of sciences, comprehensive studies and vocational studies.

In the early curriculum, the four-year high school period had been divided into two fields of technical studies and sciences. The purpose of scientific studies in high school was only to prepare adolescents for higher education. This was divided into two periods: one was three years and the other was only for one year. In the first three years, the principles of science, literature and arts were taught generally, and in the fourth year, students were assigned to one of the following fields: physics and math, social studies and economics, literature and arts. Workers who were trained in vocational studies obtained a diploma qualification so that they could start their career in one of the sections of industry, viz services and agriculture. This period included 4 years of vocational studies and 2 years of professional studies. However, this curriculum encountered changes in 1971, and at the time of implementing it, the field of comprehensive studies was added to the science curriculum. Of course, the comprehensive studies lasted only one year, and it was stopped mainly due to lack of the required human forces (Reyhani, 2015).

The seventh period started from 1978 to 1996. During this period, after the victory of the Islamic revolution, the values and attitudes toward education changed. High school education lasted four years. It included two fields in the first year, i.e. sciences and mathematics, and humanities. From the second year, the field of sciences and mathematics was divided into two fields of mathematics-physics and experimental sciences, and the field of humanities was divided into social economy, culture and literature. Since 1981, by execution of the KAD plan which was a mixture of profession and knowledge and a connection between science and action, it undertook to bring high school education out of its mere theoretical form, and the students were obliged to spend one sixth of their educational year learning a profession and building a career in society. Since 1986, those who had passed second grade high school were entitled to enter teachers training colleges for two years and start to work as teachers after getting a diploma. Prior to 1986, rural colleges were active where learners who had completed and passed secondary school studied for four years. In 1998, the general curriculum of the Islamic Republic of Iran's educational system was approved by the Council of Fundamental Changes in the Education Department; and in July 1989, the High Council of Cultural Revolution passed its general points. It was eventually finalised in January 1990 (Reyhani, 2015).

In the eighth period from 1991 to 2011, efforts were made to bring changes to the basis of the new educational attitudes in the curricula. However, unlike the previous educational system which had introduced changes from primary schools, to secondary schools and finally in high schools, by execution of the New System, the Education Department concluded that it was not possible to execute all plans simultaneously and in an integrated way. Therefore, with the intention of selective implementation of plans, it started making fundamental changes in high school education. In the New Educational System, the effort was in conjunction with reinforcement in the field of technical and professional training and upgrading it to the level of early profession. The new field of *kardanesh* (vocation-knowledge), as an innovation and a promising movement, was presented to students who wanted to learn professions in the natural working environments through planning and recognised educational standards. Attempts have also been made to prepare students to participate in educational and productive organisations in the country for purpose of training and employing the literate human forces. In this System, and in order to lessen educational failures, the unit credit system was introduced. The general objective of high school theoretical studies was to elevate the level of culture, general knowledge, having a better knowledge of students' talents and interests, initiating an appropriate foundation to direct them to their favourite field of study, and preparing them for higher education. The general objective of the *kardanesh* course was training skilled and semi-skilled human force as well as skilled workers and supervisors for the sections of industry, agriculture and services. The general objective of the pre-university course was to prepare students to enter higher education. During this period, 32 units were presented in two terms. In addition to participating in the entrance exam of universities, holders of pre-university certificate were entitled to pass the make-up units and take the associate's degree entrance exam and start work after attending the skills training courses. In the first year, when the New Educational System became popular in high schools, the existing structural and operational problems made it go through a lot of changes. It suffered from a lot of criticism which required a lot of discussion. In any case, this curriculum was evaluated many times, of course, every time when changes were made to it, they were generally made without taking into consideration the evaluations. As a result of these reviews, the two-staged entrance test of universities was eliminated, and the term-unit system turned into a year-unit system (Musapour, 2011).

The ninth period from 2011 to the present, after the announcement of the Strategic Act was passed by the High Council of Cultural Revolution in December 2011 to the Ministry of Education (MOE), the 6-3-3 structure was passed for the country. Some of the mentioned reasons for its selection were increasing effectiveness and performance of the Education Department, creating coordination and convergence between its structure and the growth and learning stages, increasing the general education period from 8 years to 9 years, and increasing the compulsory education period from 8 years to 9 years. In explaining the reasons behind the selection of this structure, some research and survey findings reported that the initial compulsory and formal education was from the age of 6. The two-staged structure of primary schools and high schools are more compatible with the implications of some narratives as well as research findings in different fields and aspects of growth and learning psychology. By passing the Act for eliminating *konkur* (university entrance test) and the possibility of entrance of all graduates to universities, the need for pre-university courses have been eliminated; therefore, more emphasis could be given to the general education period. In order to shorten the duration of this system's establishment period, gradual and grade by grade procedure was not taken. Changes started from the middle of the educational period, and simultaneously, a change in the contents, also started from two different grades. This is also closer to the global experience, and in this way, the change which started from the first grade till the fifth grade in 2011-12, continued until 2015-16.

On the other hand, by adding grade 6 in 2012-13, secondary schools do not have grade 1, and this process was transferred to the higher levels reaching high schools in 2015-16. Also, by 2018-19, with the full establishment of the new structure, the pre-university course will be eliminated. On this basis, the educational structure of the Education Department includes 6 grades of primary school and 6 grades of high school in which every educational period is divided into two three-year periods.

In the first 9 years, education is general, and in the second three-year period of high school, it is semi-specific after which the students will receive a diploma. The second period of high school includes three fields of theoretical studies, technical studies and *kardanesh*, and every one of these fields has subfields. The variety of the subfields depends on the requirements of the society, the time factor, as well as the possibility of implementing the subfield. Therefore, altogether, it seems that this period is contemporaneous with the time when bringing changes to the education system which is inevitable. Outdatedness and the inefficiency of the education system is accountable for the various present requirements in different educational, social, economic and cultural fields. This means there is a lack of a holistic perspective, suffering from unilaterality, lack of effective cooperation in the formal educational system of the country with other sectors of the society, and dissatisfaction of people as well as authorities and beneficiaries of education. This, therefore, requires a comprehensive and fundamental change in all factors of education (Reyhani, 2015).

Therefore, designing the comprehensive scientific plan of the country, the plan of cultural engineering, preparing the Education Department's Fundamental Change Act, preparing a national curriculum, doing various studies in philosophy of education and training, altogether, have led to the appearance of changes which is observable in this period (Reyhani, 2015).

2.2.2 Current View on Iranian Education System

In Iran, the education system is regarded as a social and cultural institution with the aim of serving the most prominent organisations. It plays a crucial role in edifying, disseminating and transferring culture in Iran's society, aiding students to initiate appropriate foundations and develop proper standards. As charted in the Statistical Centre of Iran's Constitution, the government of Iran is entrusted with the responsibility to provide free education and nurture the enquiry and exploration spirit in the fields of culture, science, technology, and Islamic studies in secondary schools. Also, mentioned in the Constitution, is that all religious minorities including Zoroastrians, Christians, and Jews are free to teach and practice their own faith.

A number of deputy ministries, organisations and centers that have special administrative responsibilities work under the jurisdiction of the Ministry of Education. Some of the responsibilities within the Ministry that come under the purview of the Ministry include goal and strategy development, educational activity conducting and supervision, curriculum and textbook development, educational material publication and distribution, teacher professional development planning and conducting, as well as conducting examinations and student assessment, teacher education, and defining HR policies.

The Supreme Council of Education, being an autonomous legislative body is not only concerned with the developing the role of education but is also responsible for the gamut of duties including to adopt and communicate executive policies, design mechanisms to achieve the objectives of the policies, improve related structures and processes; update, amend and coordinate programmes and policies, and monitor implementing programmes related to the strategic educational reform. According to the terms of the Council, a progress report must be submitted to the Cultural

Revolution's Supreme Council annually. This progress report will highlight the execution of the fundamental reform plan and the Ministry's performance.

At the tertiary level, three other ministries oversee other educational matters. For example, universities affiliated with the Ministry of Science, Research and Technology offer non-medical degrees, medical schools and para-medical degrees are handled by the Ministry of Health and Medical Education, and non-formal vocational education is overseen by the Ministry of Labour and Social Affairs. Ultimately, the goals of Iran's system of education are categorised into scientific and instructional goals, cultural and ethical goals, economic goals and social, environmental, and life goals.

At the upper secondary level, the structure of education is highly centralised. The Ministry of Education manages and funds elementary and secondary schools (Grade 1 to 12). Formal education in Iran consists of elementary education (6 years), lower secondary education (3 years) starting at age 12 and upper secondary education (3 years) making a total of 12 years of formal education. Students who intend to enter university must attain a high school diploma.

Children commence their primary education formally during their pre-primary years. Although Farsi is not the primary language spoken, it is compulsory for them to undertake a course in Farsi language lasting one month, especially in areas where Farsi is not the main spoken language. In such communities, Farsi is additionally taught along with the normal pre-primary activities. The supervision and educational preparation of pre-school centers come under the purview of the Organisation for Educational Research and Planning (OERP) and the Welfare Organisation. Pre-schools that are open to both boys and girls can either be public or privately owned. The main objectives of pre-primary education as stipulated by OERP are as follows:

- Contribution to the physical, mental, emotional, and social growth of children

- Promoting children's socio-emotional development, self-confidence, sense of environmental conditions, and sense of aesthetics,
- Providing children with opportunities to enjoy and be interested in group activities
- Strengthening religious and ethical values and national identity
- Promoting desired social and individual behaviour in children
- Promoting oral language development and communication skills

Primary education which lasts for six years (Grades 1 to 6) is the first stage of formal education for students between the ages of 6 to 11. The main aims of primary education are as follows:

- Creating a moral development atmosphere
- Developing numeracy skills and literacy
- Teaching personal hygiene to students
- Developing students' social skills
- Developing students' talents, abilities, and physical strength

The subjects approved by the Ministry to be taught at primary schools include Holy Quran, Farsi (reading, writing and dictation), mathematics, science, religious education, social studies (social studies, history, and geography), art and physical education. Other subjects that are also approved are, thinking and inquiry, technology, and occupations which are also taught at the sixth grade. Very often, only one teacher is assigned to teach all the subjects except for art, religion and physical education in Grades 1 to 3. However, in most schools, specialist teachers teach science and mathematics to students from Grades 4 to 6.

Students whose age are between 12 to 14 are admitted to the lower secondary level of education (Grades 7 to 9) which lasts for three years. At this level, students familiarise themselves with numerous subjects in social and physical sciences, art and humanities. The primary objectives of the lower secondary level are as follows:

- Developing students' intellectual and moral abilities
- Increasing students' general knowledge
- Strengthening scientific imagination and academic discipline among students
- Identifying individual talents and preferences with the aim of directing students towards appropriate study programmes.

On top of this, students in lower secondary school also have the option to learn a second language either English, French or German of their choice or vocational education and defence education which are open to the male students only.

The upper secondary level (Grades 10 to12) lasts for three years and is available for students between the ages of 15 to 17. During this period, students can choose one of the theoretical, technical and vocational, or Kar-Danesh (skill knowledge) programmes. The objectives of these programmes are varied to cater to students with different interests and abilities. Students who enrol in university choose the academic programmes on natural sciences, social sciences and mathematics according to their education level and career interests. On completion of the technical and vocational and Kar-Danesh programmes, students will either receive a post-diploma degree or a skill certificate which will help them to enter the labour market. Alternatively, students who have completed either of these programmes can either choose to further their education at a vocational school choosing a programme based on their capabilities and interests or apply to sit for the university entrance examination after completing vocational training.

The Statistical Centre of Iran (2013) reported that among the public and private schools ranging from elementary to university level, with elementary schools comprising about 7%, lower secondary schools 10%, upper secondary schools 18% and technical and vocational institutions 18% which are privately operated. These schools are regulated by the Ministry of Education although they are mostly funded by tuition fees payable by students. However, free education is open to all citizens to study at public schools in Iran.

Before 2002, a large number of Iranian teachers and educational staff received their qualifications from Higher In-Service Education Centers of the Ministry of Education. However, since 2002, the in-service long-term degree programmes offered by universities and higher education institutions have made it possible for science and mathematics teachers to study one of these teacher education programmes at Associate, Bachelor's or Master's degree level. Of course, other teacher education centers that have affiliation with the Bureau of Scientific Promotion of Human Resources in the MOE and are recognised as post-secondary institutions, offer a two-year associate degree programme. High school graduates who have been successful in their nationwide university entrance examination can choose these programmes. A wide range of subjects are offered under this programme including mathematics, science, physical education, social studies, primary education, Persian language and literature, internal affairs, Islamic and Arabic languages, art and special education. Those who graduate from this programme are qualified to teach in primary or lower secondary schools.

A four-year teacher education programme is also offered by Iranian universities and other higher education institutions which leads to a bachelor's degree. Those seeking entrance to universities or higher education institutions must pass the nationwide university entrance examination. Upon acceptance, students need to study pedagogy and educational psychology, beside taking specialised courses in their subject areas.

2.2.3 Assessment in the Iranian National Curriculum (INC)

In this subsection, the importance of assessing in the INC, definition of assessment, division of assessment based on NCTM, the purposes of assessment, a brief explanation about assessment in mathematics, the role of teachers in assessment and finally, the current assessment structure in Iranian schools are presented.

Regarding the importance of assessment, the Iranian National Curriculum (INC) emphasised that assessment is one of the important factors in the curriculum and it is a necessity in each educational system as it will help to protect mobility, health and life of the educational system. Gathering information, analysis and interpretation systematically during the assessment process would determine to what extent the goals of the curriculum and educational system have been achieved and how much is to be accomplished. The result of this process will show whether or not the components of the educational system such as textbooks, programme content, teachers' teaching activities, and students' learning activities have played their role properly (Reyhani, 2015).

In the INC, assessment is defined as a systematic process to collect, analyse and interpret the information in order to determine whether the desired objectives have been achieved or are being achieved and how much. The assessment of educational progress can be defined as assessing the learner's performance and comparing the results with the aim of training to make a decision in this regard where and to what extent the activities of teacher training and the efforts of student learning or student outcomes have led. These findings are essential in evaluating the academic achievement in terms of two basic things: determining the purpose and measuring the performance of learners (Saif, 2011).

Moreover, in terms of definition of assessment, Gholamāzad (2011) noted that collecting the data or information is not only meant to take the tests, because if the assessment is only restricted to the tests, then the important aspect of assessment which can facilitate the growth of the students and provide information for learning process will be worthless. Assessment in mathematics in the INC is based on the NCTM principles. The NCTM mentioned four stages of assessment, namely: (1) Setting clear goals, and designing assessment, (2) gathering evidence or using a different method, (3) Using the results to make decisions, and (4) interpretation.

The first and the most important one, is determining the purpose of assessment. It is significant that teachers must first have a clear knowledge about the purpose of assessment. In other words, without knowing about what exactly teachers want to assess they will be unable to design or select the assessment. Therefore, they will be unable to have an accurate decision and interpretation of the result.

Gholamāzad (2011) classified the purposes of the mathematics assessment into four categories. She mentioned that although there are other ways to classify these purposes, the following categories indicate the first few things that fixing them requires an assessment. This classification of assessment purposes consists of the following categories: The first purpose of assessment is enhancing the growth of students, meaning that to achieve the learning objectives, teachers should provide constant feedback from the assessment of students. Assessment not only should provide the information of the growth of ability in problem solving for both students and teachers during teaching process but should also focus on the procedural skills. Therefore, in this way assessment can be effective in promoting the students' growth (Gholamāzad, 2011).

The second purpose of assessment is the improvement of education. according to Gholamāzad (2011) the daily assignments that the teacher has designed to develop students' understanding, should include daily information on how students think and what kind of idea they use and develop. Indeed, she highlighted the daily problems for solving and discussions in comparison with the examination at the end of each chapter which provides more useful, valuable and effective data. In this way, assessment can contribute to improving education.

The third purpose of assessment is identifying priority and perfection. In this category, assessment is seen as a process of determining the worth or value of something based on the reviews and also enabling accurate judgments by teachers (NCTM, 1995a). Beside the teacher's judgment, assessment consists not only of tests' data but should also take into account the sources and types of information gathered while teaching. Most importantly, assessment should be carried out on the basis of what students know and understand. Therefore, assessment can be a tool to recognise students' achievement and their success.

The fourth purpose of assessment is the modification of programmes. The results of the assessment should be used as a component in response to this question: "Has the programme worked well to achieve the objectives?" Thus, the results should lead to an adjustment to the programme. In this way, the programme is not limited to the textbook; it also includes the teachers' educational programme that teachers design for themselves in the classroom and other school resources (Gholamāzad, 2011).

The INC emphasised that today the conception of assessment goes beyond the traditional understanding and teachers should see assessment as a dynamic process during teaching and learning. Considering this, it is expected that teachers use a variety of assessment tools and strategies to assess students' achievements such as journal writing, alternative assessment, open-ended tasks, self-assessment and so on. Indeed, teachers are responsible for creating educational opportunities for students, adjustment, codification, implementation and assessment of the curricula inside the classroom. Therefore, they should use any activities inside the classroom in order to assess students' achievements and use the results of the assessment to improve learning and adjust teaching (Reyhani, 2015).

Regarding the current assessment structure in schools, it can be said that prior to the start of primary school, a school readiness assessment is administered to screen and diagnose students with special needs, and to determine whether students are ready to start first grade or need to be referred to a programme for students with special needs. The evaluation of the students before grade 6 includes a consistent formative assessment. It includes an observation of their in-class activities as well as their oral or written examinations in the class and their homework.

Schools give the students report cards in January and June which are descriptive of each student's performance. Students who do not qualify for promotion to the following class with a minimum of three subjects, have to enroll for make-up sessions to fulfil further achievement activities before the following September. When grade 6 comes to an end, students sit for a regional examination prepared by the Office of Assessment in the Ministry of Education located in every province. Students who pass this exam are issued a certificate confirming that they have passed elementary school. Those who fail this examination get the chance to retake it in September. Students who cannot pass this examination even in September, still get the chance to re-sit the following year. At this level, the grading system uses marks gained in continual assessment as well as in written and oral exams which are based on specific criteria. Every student has to get 10 points out of 20 to qualify for promotion.

The promotion system in the lower secondary level is similar to the one in the elementary school but it is not as descriptive. Students in the lower secondary schools are graded on a point system (out of 20) in each subject of study and are issued report cards. At the end of grade nine, students who pass the national final examination receive a lower secondary school certificate and in higher secondary school in grade eleven students who pass a national final examination receive a diploma. High school graduates who want to go on to further their education take a university entrance examination, administered by the National Organization for Educational Testing (Reyhani, 2015).

2.3 The Mathematics Reform Movement

This section first presents a discussion on mathematics reform movement and then discusses about mathematics assessment reform and explores assessment reform around world.

In the mathematics education, the roots of the reform movement can be found in the mid-1980s when it served as a response to the failure of traditional ways of teaching. Also, technology has an influence on the curriculum as well as the appearance of new approaches to the scientific methods of studying how mathematics is learned. As a main point in the reform movement, there was a standards-based approach as to “what and how” mathematics teaching was (Reyhani, 2015).

In modern mathematics, though, the main focus is on problem solving, mathematical reasoning, how to justify opinions, how to make sense of complex situations and, of course, it is also about learning new ideas. Students must be given the chance to solve complex problems, to formulate, and also to test mathematical ideas. They should also be entitled to draw conclusions. They should be able to read, write and discuss mathematics, apply demonstrations as well as drawings and real-world objects; and also, be able to take part in formal mathematical as well as logical debates (Reyhani, 2015).

The leading source for the standards-based approach to mathematics instruction are the standards formed by the National Council of Teachers of Mathematics (NCTM). The handbook, *The Principles and Standards for School Mathematics* published by NCTM in 2000, outlined the foundations and standards required in developing a comprehensive programme about mathematics in the schools. This publication described six main principles in terms of equity, curriculum, teaching, learning, assessment, and technology. It also outlined the five content and process standards indicating what stages and contents the students need to know and should be able to use. The standards related to content are organised in the section related to content standards about numbers and operations, algebra, geometry, measurement and data analysis as well as probability. The standards of the process are systematised on the areas of problem solving, reasoning and proof, communication, connections and representations (NCTM, 2000).

A collection of basic assumptions on the teaching and schooling activities is indirectly shown in this reform movement. The first point is that all students should have the opportunity to learn new mathematics. The second point is that all students have the capability to learn mathematics more than we traditionally thought they can. The third point is that new applications and changes that have appeared in technology have brought changes to the instructional importance of some mathematical ideas. The fourth point is that by using technological tools, new instructional environments can be created. The fifth point is that learning mathematics in a meaningful way can be achieved through purposeful engagement and interaction which can be met through prior experience (Romberg, 2000).

The American Mathematical Society published a concept paper which was helpful in identifying some of the common agreement areas in the teaching and learning of mathematics. These areas are founded on three basic premises; 1) having fundamental skills with numbers is still important and students need to have proficiency with procedures of computation; 2) mathematics requires precise reasoning about subtly defined objects and concepts; 3) students should be capable of formulating and solving problems. The areas where these premises overlap in agreement are the following: fluency in mathematics requires the students to recall automatically certain procedures and algorithms; making use of calculators in teaching is permissible, but it should not hinder the students' fluency development with basic facts and computational procedures. It is crucial to use and understand the basic algorithms of whole number arithmetic and to build and understand numerical meaning fractions; teachers are required to maintain a focus on mathematical concepts in order to use "real-world' contexts in teaching mathematics. There has to be multiple strategies for teaching mathematics, however, the teacher is responsible for selecting the strategies which are proper for specific concepts. Teachers of mathematics have to understand the fundamental meaning and justification of ideas and also, they have to be able to make connections between their topics (Ball et al., 2005).

In conclusion, Bray (2011) and Ross et al. (2003) pointed out that mathematics reform has since turned away from the traditional way of mathematics education where the "teacher is mainly focused on lecturing, modeling and drill" towards focusing more on students, solving problems, and hands-on approach. This new movement, according to TIMSS is centered more on students and learners (TIMSS, 2010).

2.3.1 The Mathematics Assessment Reform

Recent changes had emphasised the constructivist approach on mathematics education (NCTM, 2005; Shepard, 2005). It has advocated the need for the problem-solving approach to promote students' reasoning and communication skills. As a result, these changes focused on the need for new approaches of assessment in mathematics. New approaches of assessment emphasised the significant mathematical concepts, introduced an all-inclusive perspective of mathematics, and comprised a full domain of mathematical operations which reflect the critical role of problem solving (NCTM, 2005; Shepard, 2005).

The National Council of Teachers of Mathematics (NCTM, 2000) established nationwide standards for assessments in the mathematics classroom. NCTM also discussed the concept of an assessment-centered learning environment, meaning teachers strongly incorporate assessments into teaching, continually assessing students' progress and learning. The progress of student learning is the focal point of the learning environment. There are two components where teachers need to understand about assessments: *what* and *how* to assess (NCTM, 2000). It simply means, teachers need to consider what they want to assess and how they have planned to assess when designing the unit and lessons. Indeed, the focus of what teachers are assessing is not only knowledge of mathematical procedures but also the mathematical concepts.

Further, NCTM (2000) stated that "assessment should support the learning of important mathematics and furnish useful information to both teachers and students". It is clear that assessment has always been an important aspect of mathematics instruction and is a major contributor to students' mathematics learning. Furthermore, assessment also supports teachers in making instructional decisions. In choosing a task for assessment, teachers tend to impress on students' what level of mathematical knowledge and performance are valued. In the NCTM (2000) report, it was mentioned that any response from the assessment tasks will undoubtedly aid students in forming goals, make students responsible for their own learning and become open-minded learners. Therefore, in this manner, assessment could be categorised into four extensive types: keeping check on students' progress on what their learning goals are, forming instructional decisions, appraising the progress of students from time to time and assessing the programmes (NCTM, 2005).

Therefore, with the new approaches, assessment in mathematics is required to proceed beyond focusing on students' making good use of a memorised algorithm or procedure and must also elicit, evaluate as well as respond to students' understanding of mathematical skills as well as their problem-solving skills. To implement this and as an attempt to reduce memorising in a mathematics class, the evaluation strategies must move away from tradition by not just relying only on final year tests and mid-unit quizzes. Shepard (2005) and Stobart (2008) unanimously agreed that both concentrated on knowledge recall and procedural learning. The main purpose of assessment is to stimulate the students' thinking system regularly and for this to be carried out the assessment must be filled with instructions. On this note, the teacher not only has to be a good listener but also be responsive to what the learner is thinking. Stobart (2008) further pointed out that assessment can be perceived as a constructive activity which inspires active lifelong learning.

The National Council of Teachers of Mathematics have outlined observations, interviews, performance tasks, reflective journals, projects, portfolios, presentations as well as self-assessments as some of prerequisite activities that are necessary for developing a mathematical understanding and insight into mathematical thinking of the students. All these elements are deemed essential when implementing the current methods in the teaching and evaluation of mathematics (NCTM, 1995).

Consequently, both assessment and mathematics education reforms were of the opinion that one single short test is not sufficient to assess the complex nature of students' mathematical thinking. Still, there is a need for teachers to be more mindful about students' abilities to reflect on mathematics and also to concentrate on assessment that will improve learning and to upgrade their teaching. One aspect of encouragement to teachers is to include assessment in their instructions, be more centered on listening and responding to students' thinking as they share their problem-solving solutions and provide a detailed feedback on the day-to-day classroom work.

2.3.2 Assessment Reforms Around the World

In this subsection, the Assessment Reform Movement (ARM) in higher-achieving countries is briefly explained in order to have an insight into the possible factors behind reform success. The researcher has chosen two cases of reform in different continents, namely Finland and Singapore.

Between 1970s and 1990s, the Finland curriculum system was reformed. The practice of tracking students into different streams using test scores and offering them a common core curriculum was abolished. There were changes in the curriculum system which were projected to balance out the educational system and initiate a wider access to higher education. With these changes underway, Finland has since moved away from a highly centralised system based on external testing to a more localised system applying multiple forms of assessments (Sahlberg, 2009).

This might seem counterintuitive for the people who are used to external assessment as a way to ensure accountability. Government leaders in Finland publicize the use of learner-centered, open-ended tasks which are school-based and rooted in the curriculum as a key factor in the country's remarkable success in international examinations. Policymakers unanimously decided that if they invested heavily in highly competent teachers, the local schools would have a freehand in deciding on what and how to teach - this response is in contradiction to the highly centralised system they were hoping to restructure. The set of documents outlining the present national core curriculum is much slimmer compared to the hundreds of pages of highly specific instructions and descriptions of a limited number of skills and core concepts each year. (For example, only 10 pages describe the full set of maths standards for all the grades). Lavonen (2008) stressed that this will aid teachers to jointly develop the curriculum and assessments locally that can inspire students to actively research, analyse, and utilize information for problem solving in new situations.

The education system in Finland does not practice using external standardised tests for ranking of students or schools. Nevertheless, the students' performance is appraised periodically by the Finnish education authorities. This is usually carried out towards the end of their 2nd and 9th grades to update decisions about curriculum and school investments. Designing and managing other assessments is carried out locally.

A list of recommended assessment criteria for each subject as well as the annual final assessment of students' progress in each grade has been outlined by the national core curriculum. Each teacher receives a copy of these assessment criteria. Then, these guidelines are used by schools, authorized by the Finnish National Board of Education (2008), to create more detailed curriculum and learning outcomes as well as approaches for the assessment of the curriculum standards.

The national standards had stressed that the aim of assessment is to lead and motivate the students' self-assessment and reflection. However, the continuous teacher feedback is also of prime importance. Teachers can then provide students with formative and summative reports that reflect the students' performance level relating to the respective curriculum objectives. The reports issued by the teachers are based on several assessment forms and not just on examinations.

In Finland, assessment is carried out to nurture students' active learning skills by asking open-ended questions and helping students to handle these problems. It is very unlikely to see a teacher in Finland, standing in front of a classroom lecturing to students for fifty minutes. But on the contrary, students can be seen involved in independent or group projects of their choice and setting their own targets with teachers in specific subject areas, who will serve as coaches. The independence and active learning motivate students to nurture analytical thinking, problem solving, and meta-cognitive skills.

Before entering university, majority of the Finnish students sat for a voluntary matriculation examination where they are asked to apply problem solving, analytic and writing skills. In 2008, the Finnish National Board of Education passed a rule that teachers should follow the official guidelines when grading matriculation examinations locally and samples of the grades must be re-examined by professional raters appointed by the Matriculation Exam Board.

Eventually, Finland top the list in mathematics, science and reading among all the OECD nations on the PISA assessments. It was evident that the contributing factor to the reform success and improvements in Finland's education was the intensive investment in teacher education. In view of this, all teachers were rewarded with a three-year high-quality graduate level preparation entirely at state expense plus a complete overhaul of the curriculum and assessment system. Majority of the students who possess a Master's degree in education are ready now to teach all types of learners including those with special needs. Laukkanen (2008) emphasised that to teach special needs students, the teachers must have an intense understanding and a strong focus on how to apply formative performance assessments in the service of student learning.

A study was carried out by Tan (2011) who highlighted that between 1965 to 1967 Singapore has gone through three educational reforms. The last reform won Singapore the title as the 'most academically successful nation in the world." After the launching of the "thinking schools, learning nation" initiative in 1997 by the country's Prime Minister, Singapore has clearly focused its reforms on curriculum, assessment, and teaching with the aim of developing a creative and critical thinking culture within schools. This can be achieved by imparting these skills to students and by creating an inquiry culture among teachers as well. Thus, teachers are coerced to undertake research on their teachings and to continuously revise their teaching strategies in response to what they have learned (K. Tan, 2011).

Like any other reform movements, this reform encountered similar challenges. Koh & Luke (2009) reported that during 2004-2005 in one of the projects, teachers evaluated students based on the memorisation of factual and procedural knowledge for a high-stakes examination. Because of this negative effect of examination practices on student learning, Singapore reviewed the affiliation between assessment and learning in primary schools and examinations for Primary 1 and 2 which were eventually substituted by the school-based holistic assessment practices to support learning (K. Tan, 2011). With less emphasis on semester examinations, teachers have more opportunities to pay more attention to assessment for learning.

The new qualitative approach to assessment as expounded by Tan (2007) concentrates on various aspects of assessment and learning. It can be said that this new trend of assessment will encourage open-ended responses (as opposed to standardised instruments) allowing for student participation (as opposed to unilateral testing) which will spread over a period of time (as opposed to controlled environments) in order to prompt, judge and enhance holistic understandings.

Although there is still room for improving the quality of assessment in the education system of Singapore, it is quite obvious that during the last ten years, the education system of Singapore has always been ranked at the top or near the top with most of the major world education systems. Singapore has always viewed education to be extremely vital in building the economy and the nation with the aim to serve as the engine of human capital to drive economic growth (OECD, 2011).

The OECD (2011) report outlined other success factors that have contributed to the success of Singapore's education system. These are a clear belief and vision that education is central for students and the country, persistence in political leadership and coordination between policy and practice, focusing on building teacher and leadership capacity to carry out reforms in schools, ambitious assessments and standards, and a continuous improvement and future orientation culture that measures educational practices against the best in the world (OECD, 2011).

In summing up and upon review, the two assessments reform movements of these two countries appear to be in line with the global assessment reform that consider assessment beyond the traditional or accountability purposes where assessment is used as a major tool for boosting the learning and teaching process. To implement this, focus should be on the success of these countries regarding the improvement in the quality of teachers through professional development programmes, substantial investment in teacher education at graduate level for all the teachers, and their reliance on well-trained teachers to implement their curricula and assessment systems.

However, this matter has not been taken seriously by educators and professionals in Iran.

2.4 Significant of Assessment and its Purposes

In this subsection, we will briefly elaborate on the importance of assessment and its purposes based on several studies.

Assessment besides being an inevitable fact of life is also considered a vital aspect of educational system. It is also an essential task that is needed in the teaching profession and is an essential part of the learning process (Black & Wiliams, 1998). Simpson (2016) and Vik (2013) asserted that assessment is viewed as a key part of teaching and learning processes since it makes evaluating students' learning and utilising the information to enhance learning and teaching possible. Assessment encompasses all the activities performed by teachers and students to collect information which can be utilized to maximise learning and teaching (Reutzel & Cooter, 2016). Harris, Irving and Peterson (2008) further debated that assessment carried out in classrooms will, undoubtedly, assist teachers to regulate the teaching process and how they can adapt their lesson plans to cater to the needs of their students.

Rowntree (1987) cited five essential dimensions of assessment suitable for any educational setting that teachers need to comprehend. The first and the most important stage of conducting an assessment is, why assess? This dimension decides on the purpose of the assessment and what results to expect. Chappuis et al. (2009) defined this as a "clear purpose" and further elaborated that "the assessor must begin with a clear picture of why he or she is conducting the assessment". The remaining four dimensions are "what to assess" which involves deciding the learning objectives of the assessment, "how to assess" which refers to the selection of the appropriate assessment strategy, “how to interpret” which involves interpreting the test results and analysing the assessment data and drawing accurate conclusions, and “how to respond”, which involves communicating the assessment results to the student and other stakeholders such as parents and school or district administration. However, it should be made known that without sufficient knowledge about the first dimension (purpose of assessment) teachers would find it difficult to continue with the remaining four dimensions mentioned above.

In an article entitled “Assessment Literacy for the 21st Century” written by Stiggins (1995), the author highlighted that “assessments should arise from and serve clear purposes; arise from and reflect clear and appropriate achievement targets; rely on a proper assessment method, given the purpose and the target, sample students’ achievement appropriately; and control all relevant sources of bias and distortion.” Furthermore, Newton (2010) quoted an example to explain the significance of the purpose of assessment:

“Anyone planning to invest in a new vehicle ought first to have a good idea of what they are going to use it for. If the foremost desire is to cruise in style along winding roads, then a Harley Davidson motorcycle might fit the bill. If, instead, the major concern is to transport colleagues, tools, and the odd animal across country fields, then a four-wheel-drive utility vehicle might be much more appropriate. In just the same way, those who design vehicles need to start by considering transportation purposes, as do those who review them.”

In addition, Newton (2010) pointed out that designers, users and evaluators of educational assessments should reflect on the purpose before starting an assessment. Obviously, trying to use specific assessment types without specific purposes would be pointless. It may seem odd that in educational assessment, this point is not often understood and appreciated.

Thus, full attention must be given to the purpose of the assessment irrespective whether it is a large-scale or just a classroom assessment as the findings of the assessment will be translated and used appropriately for the purpose. Koch (2013) emphasised that the findings of the assessment should be used only for the purpose for which the assessment was designed, and the inferences need to be appropriate to the assessment design. Nevertheless, it must be borne in mind that the purposes of assessment are sometimes blurry. For example, teachers often use summative classroom assessments for formative purposes when they use the summative results to understand students' misconceptions and design future instruction accordingly, or when they use tasks from large-scale assessments for instructional purpose (Parke et al. 2003).

2.5 Definition of Assessment in Literature

This subsection discusses the definitions of assessment based on the review of literature. Through this review, assessment has been differently defined. Most of them are associated with the process of decision making that is the main part of the teachers' job during the teaching and learning and consist of methods which are used to recognise to what extent students have achieved the intended learning outcomes of instruction. Furthermore, assessment involves four steps which are, (a) measuring students' achievement, (b) defining student performance, (c) summarising what students can do, (d) concluding what students could do (Gronlund, 1998).

Later, Huba and Freed (2000) stated that assessment is an essential component of teaching and learning which contains the process of collecting, interpreting and recording data connected to student development in learning and the efficiency of the teaching approaches. They elaborated that assessment as a process of collecting and debating data from several sources is able to expand a profound understanding of what students know, realise and can do with their knowledge as a result of the educational experiences. In an almost similar view, assessment is seen as a process of collecting information purposefully using different methods/strategies and tools for the purpose of notifying decision (Susuwele-Banda, 2005).

Dhindsa, Omar, and Waldrip (2007) defined assessment as a systematic process of collecting data about student achievement and also viewed assessment as an essential component of teaching. Later, McMillan (2008) explained that assessment is the process of gathering, incorporating, and inferring information to contribute to making a decision.

Recently, Morris and Adamson (2010) defined assessment as "those actions we undertake to obtain information about students' knowledge, attitudes or skills." Later, Smith (2011) said that assessment is a group of processes which teachers use when they try to understand and draw conclusions on the students' learning process, progress and learning outcomes. In fact, assessment is a significant tool to document and improve learning. Simply, assessment is seen as an umbrella term which includes diverse practices such as deciding, collecting and making judgements about evidence relating to the goals of the learning being assessed (Harlen, 2012).

Consequently, in this study, assessment is understood as any act of interpreting data concerning the performance of students where these data are collected through a multitude of means or practices (Brown, 2002, 2004).

2.6 Types of Assessment

2.6.1 Definition of Classroom Assessment in Literature

This subsection presents a brief discussion of the definitions of classroom assessment in the literature as well as the definition of classroom assessment in this research.

Black and Wiliam (1998) described classroom assessment as "the most notable element of the teaching and learning process." Gallagher (1998) further expounded that "classroom assessment refers to the procedure used in the classroom by the teacher to gather information about students' performance on assessment tasks, either as a group or individually, using an extensive range of assessment methods, to dictate the extent to which students are achieving the target instructional outcomes".

According to Zhang and Burry-Stock (2003), classroom assessment is not only restricted to developing assessment methods. It also embraces a wide spectrum of activities, namely paper-and-pencil tests, performance measures, administration, grading and finalising the results of assessment, perfecting grading procedures, conveying assessment results, and using them to make educational decisions (Zhang & Burry-Stock, 2003).

In the view of McMillan (2007), classroom assessment is seen as the "collection, evaluation and also utilisation of the data that enables teachers to make decisions in order to enhance student learning". Further, McMillan explained that the information of classroom assessment qualifies teachers to make decisions in structural practices and also make a valuable situation for learning, besides notifying students concerning the level of their learning.

According to Phelan and Phelan (2010) assessment in the classroom is made up of a set of activities including observing, collecting, analysing, and interpreting evidence so that deductions can be made, and more information could be provided for better decision-making by helping them answer fundamental questions. Furthermore, paying specific attention to the questions helps to show that it is a question that can (and has to) be answered in different ways. Classroom evaluation is not a single activity that fits all forms.

In addition to that, researchers pointed out that there is only one way to collect evidence of assessment. Written tests are seen as the stereotypical assessment method in the classrooms, whereas many other options are available. These options are the more traditional classroom assessments, namely quizzes, homework, written tests, final exams, etc. Also, there are many other assessments including portfolios, performance evaluations, open-ended questions, online evaluations, and computer simulations – often considered as alternative assessments (Phelan & Phelan, 2010).

Keaikitse (2012) defined assessment in the classroom as tests or teacher-made assessment built and managed by teachers as a formative assessment of students' learning. In fact, classroom assessment is applied with the purposes of monitoring the learning level of students as well as their feedbacks.

In addition, she explained that in order to understand what students know or do not know, teachers need assessment. Hence, classroom evaluation is perhaps the first and the most important part of the teaching and learning process which includes measurement, feedback, reflection, and change. It means that, classroom assessment plays a critical role because it is crucial for data creation used in making educational decisions (Keaikitse, 2012).

Recently, Baird et al. (2014) said that classroom assessment contains a series of 'teacher-selected' or 'teacher-made' assessments that are most fruitful once they are thoroughly associated with what and how the students have learned. Using classroom assessment, teachers collect data and prepare feedback to support individual student's learning and enhancement in teaching practice. In fact, this is the current view of classroom assessment that inspires teachers to use a variety of assessment approaches, tools, and formats, providing numerous chances for students to show their learning, making a strong use of formative feedback on a timely and regular basis, and including students in the assessment process (NCTM, 2014).

In summary, from the definitions presented in the above review, classroom assessment in this research refers to a variety of 'teacher-selected' or 'teacher-made' assessments that provide opportunities for students to be active in the learning process. In addition, it helps to acquire information and formative feedback for decision-making within the classroom context with the intent of enhancing teaching and learning.

2.6.2 Definition of Alternative Assessment in Literature

This subsection presents a brief discussion on the definitions of alternative classroom assessment or new approaches to assessment in the literature and ends with the definition of alternative assessment in this research.

According to Law and Eckes (1995) alternative assessment has contributed to innovative methods to encourage students and motivate them to discover and use their inner abilities and also the dimensions of the world outside. As scientists of education believed, it is true that remembrance of the subject matter or the contents can be tested easily, but it is difficult to assess critical thinking and creativity. According to these scientists, experts should concentrate on the long-term retention of knowledge and information and also focus on its transfer. Alternative assessment gives the teachers the opportunity to find out the weak and strong points of their students in different situations (Law & Eckes, 1995).

Later, Gronlund (1998) viewed alternative assessment as assessment methods that can provide an alternative to the traditional paper-and-pencil tests. Prior to that, Puhl (1997) noted that alternative assessments allow students to demonstrate their understanding and personal meaning of what they have learnt in class.

Indeed, this definition was in contrast with the view of alternative assessment by Rahim and Sulaiman (2006) who stated that alternative assessment techniques encouraged a higher-order creative and critical thinking where students have more control over their learning. At the same time, Gronlund (2006) defined that alternative assessments such as portfolios, observations, and other performance-based assessments are higher both in the realism and complexities of the tasks assessed and require more time to use and score than the traditional assessments.

Later, Wu and An (2007) argued that alternative assessment refers to assessment methods that are 'non-standardised' and 'non-traditional'. For instance, alternative assessment refers to the use of non-traditional forms of assessment such as portfolios and projects (Nasab, 2015). She clarified that portfolios include students' work along with the display of mastery of skill in terms of the current tasks. She illustrated that as elements of formative assessment, portfolio assignments insist on the construction of knowledge for the final product via appropriate mechanisms. Apart from using authentic experiences, portfolios are helpful because of the fact that they are able to reproduce processes that require problem-solving approaches. In this way, a project is a goal-oriented task that can be realised in any form of the developing plans, research proposals as well as art works. They require learners to make use of their personal skills and strategies to solve problems (Nasab, 2015).

Mussawy (2009) specified that the term "alternative assessment" is also interchangeably used in formative assessment or in assessment for learning. In some cases, it is noted that formative assessment is another name for alternative or performance assessments. There are several types of formative assessment and one of the most important is alternative assessment. Alternative assessment includes several evaluation techniques, most recognised them as: information gathered through an examination of learning-progress (e.g. portfolios), of realised projects (e.g. expressed in writings, drawings or oral presentations), or of collaborative group performances. Similarly, different kinds of alternative assessment techniques are tools that make it possible for students to self-assess their current knowledge or competence level, during a continuing educational experience (Mussawy, 2009).

Recently, Stobart and Gipps (2010) explained that alternative assessment refers to activities ranging from changing the format of doing something similar to describing completely different approaches to assessment. These alternative approaches are performance, diagnostic, authentic, and formative assessments that reflect varying approaches to learning. This contributes to clarifying and questioning the way the term is used.

At the same time, Frey and Schmitt (2010) clarified that alternative assessment or performance test is an assessment that requires the performance of a skill or the production of a product. This assessment is usually designed to measure a skill or ability. Scoring often requires subjective judgment.

Consequently, according to the discussion above, alternative assessment in this study refers to various forms of non-traditional assessment that concentrates on the process and product of learning to support and improve the teaching and learning process.

2.6.3 Definition of Traditional Assessment in Literature

This subsection presents a brief discussion of the definition of traditional assessment in the literature and the definition of traditional assessment in the present study.

From a traditional point of view, many teachers equated assessment with a test or quizzes. However, assessment and testing are remarkably different from one another. Whereas testing is prepared formally and is usually standardised, an assessment is prepared on the basis of data collection about the knowledge of students and what they are capable of doing. It means that the exact procedures are provided for teachers to carry out administering and scoring in testing. However, in assessment, there are a variety of ways and methods for data collection at different times and contexts (Law & Eckes, 1995).

Bailey (1998) pointed out that the traditional way of assessment is indirect, unreliable and standardised. Because of these reasons, they are one-shot, speed-based and norm-referenced. Prior to that, Law and Eckes (1995) also reiterated the same issue and cited that traditional assessments are single-event tests. This implies that they measure what learners can do at a particular time. Not only test scores cannot report on the progression of a child but also, they cannot tell the type of difficulties students encounter during the test.

In line with this, Bailey (1998) reported that learners had no response to this type of assessment as the projects are individualised and the assessment procedure is decontextualised. Law and Eckes (1995) further elaborated that most standardised tests assess only the lower-order thinking skills of the learner. In a similar case, Simonson et al. (2000) cited that the traditional assessment frequently concentrates on the learner's ability to memorise and recall which are regarded as the lower level cognition skills.

In addition, in traditional assessment, learners are obliged to display their knowledge in a predetermined way. As such this assessment classified students as winners and losers in the classrooms (Stiggins, 2005) and order the students to compare themselves with the other students. In this way traditional assessment is only used for the accountability of students or schools.

Therefore, not only traditional assessments do not encourage the students to be active in the learning process, but they also do not provide students with the opportunity to apply abstract knowledge to a variety of settings. In fact, students are not given the opportunity to apply their knowledge to a context in which they are interested (Van Duinen, 2005). This is the behavioristic view to assess that the traditional assessment design is globally based on a stimulus/response view of learning (Resnick & Resnick, 1992).

It is noted that the most significant advantages of traditional tests are that they are not time consuming, and they are easy to administer and grade, and can be simplified so that they are easy to score reliably. It is also relatively simple to validate and determine internal consistency for traditional tests. According to Brown & Shavelson (1994) such an assessment is valid only for testing students' factual knowledge, not procedural.

Chappius and Stiggins (2008) clarified that traditional assessments (summative assessment) have satisfied the traditional function of tracking the progress of students as the results from the assessments are assessed, in the form of a report card grade. As a matter of fact, they were assigned with the main task of giving a grade to each student. Britton (2011) opined that when the test is over, there is no turning back no matter how a student has performed. A summative assessment or traditional assessment is also defined as "product-oriented," and is designed to certify a student's mastery of objectives and to gauge the level of acquisition of a specific learning objective or curriculum goal (Reyhani, 2015).

According to the researcher's experience as a mathematics teacher, traditional assessment in mathematics consists of recalling formulas and procedures for problems arising from worksheets and homework. In fact, traditional assessments have been primarily teacher-based with limited student involvement, meaning that the students are not given opportunities to express their opinions about the assessments they perform.

Moreover, these assessments promote the value of extrinsic rewards, rely on competition, and encourage academic dishonesty such as cheating or plagiarism. With traditional assessments as the primary means of assessment, students lack initiative in their learning; therefore, students lack autonomy, which is the "ability for making all the decisions concerning all aspects of their learning" (Adeyemi, 2012).

In this study, traditional assessment refers to the definition by Frey and Schmitt (2010) as follows: "traditional assessment is referred to as the paper-and-pencil tests made up of multiple-choice, matching, true/false, short answer/filling-the-blank or essay questions. They are usually designed to measure understanding or knowledge (lower-order thinking) and scoring is often objective."

2.7 Definition of Conception

In order to fully understand the model used in this study and the research questions, it is important to define conceptions and clearly understand the teachers' beliefs about assessment.

Ausubel (1968) described conceptions as “something that appears to us as a consequence of our concrete-empirical experiences with that very thing. “Researchers have debated that although conceptions appear to be abstract, they are not existential in nature. Further, it is deliberated that “conceptions are not knowledge or truth as philosophers or experts might look for, rather conceptions are psychologically true representations of the world, individuals or groups hold about some domain” (Ausubel, 1968).

Brown (2008) asserted that “conceptions are sometimes incomplete but are relative to each individual. Conceptions represent what individuals understand, know, believe, think, or feel about a domain at any one time.” He added that conceptions of various educational processes tend to change or impact the educational practices and outcomes of students’ achievement (Brown 2008).

Pratt (1992) inferred that conceptions refer to numerous groups of thoughts held by teachers behind their explanation of how pedagogical things are experienced. As a matter of fact, teachers see the world through the lenses of their conceptions, interpreting and acting according to their understanding of the world.

Initially, the term “conception” was introduced by Thompson (1992). In the introduction, he elaborated further that conception is “the general mental structure, encompassing beliefs, meanings, concepts, propositions, rules, mental images, preferences, and the like.” In other words, conceptions are consolidated into unitary construct knowledge and beliefs, which represent a sub-category of the conceptions.

In the national assessment context, conceptions described the teachers’ general assessment perception and awareness (Barnes, Fives & Dacey, 2017). Brown (2002, 2004) used the term “conception” as an “organising framework with which an individual understands, responds to, and interacts with a phenomenon.” Brown’s impression of the term "conceptions" sees beliefs and conceptions in a rather harmonious and non-conflicting manner.

Therefore, the term "conception" refers to having an understanding of the nature and goal of a pedagogical process and practice (Thompson, 1992). Moreover, the term "conception" refers to a person's general, and implied, knowledge about the nature of a phenomenon (Brown, 2008, Thompson, 1992). In this manner, conceptions refer to people's thoughts, values, and attitudes about the nature and structure of something and its purposes.

Regarding significance of studying teachers' conceptions, Brown (2000) clarified that there is a relevance in the teachers' conceptions of assessment as it comprises teaching, learning and curricula which have a strong impact on how the teaching is conducted and what students have attained (Pajares, 1992; Thompson, 1992). Brown further elaborated that the perceptions of teachers and the assessment of students' behaviour and performance that are incorporated in all academic studies are influenced by learning, assessment and curriculum. According to Brown (2002, 2004) the conceptions of assessment represent the teachers' understanding of the way how students' learning is examined, tested, evaluated or assessed.

In summary, in this thesis the word "conception" is used to refer to the organized structure of ideas and beliefs about the nature and purpose of assessment (Brown, 2002).

2.8 The Conceptions of Assessment Model

The theoretical perspectives of this study is drawn from Brown (2002, 2004). The link between teachers' conceptions, the nature of assessment and the methods of assessment that teachers utilize based on that conceptions can be explained through this model. It is used in this study to support the investigation of teachers' conceptions about the purpose of assessment and the practices of assessment and thus guided in forming the conceptual framework of this study.

Brown's model (2004, 2006) contains a schematic conceptualization of assessment purposes in teaching and learning process on how teachers' conceptions about assessment play a crucial role in the teachers' assessment techniques and practices. This model addressed several important questions to consider in the investigation of teachers' conceptions of assessment. Based on this model, teachers have conceptions on every aspect of their assessment practices.

Brown (2002, 2004) has used the term 'conception' to describe the organizing framework by which an individual understands, responds to, and interacts with a phenomenon. The term conceptions are considered to reflect "a more general mental structure, encompassing beliefs, meanings, concepts, propositions, rules, mental images, preferences, and the like" (Thompson, 1992). The use of conceptions allows for knowledge and beliefs to be synthesized into a single construct.

This model shows how conceptions represent varying teacher beliefs and ideas behind their descriptions of the way educational experiences occur. Brown believed that conceptions are like filters for teachers to interpret and experience national or school-based policies and practices as well as professional development activities related to assessment.

Brown (2002), using three major purposes for assessment, namely improvement of teaching and learning, certification of students' learning, and accountability of schools and teachers (Torrance & Pryor, 1998; Warren & Nisbet, 1999; Webb, 2012) and adding a fourth purpose of assessment as irrelevance (Shohamy, 2001), proposed the four factor inter-correlated model of conceptions of assessment. Indeed, this model is designed to identify the teachers' level of agreement on the four inter-correlated, major conceptions of assessment based on one logical assumption:

"Since teachers' conceptions of instruction influence how they teach (Brown, 2002; Pajares, 1992), their conceptions of assessment should help us understand how they assess and how we might remediate it."

This model is chosen as it strongly proposes a strong approach to nature and purposes of assessment and teachers' conceptions. This model was also employed by other researchers all around the world to study and interpret the teachers' conceptions in different contexts and different fields of study. Azis (2015), Barnes, Fives, and Dacey (2017), Brown, Chaudhry, and Dhamija (2015), Brown, Hui, Yu, and Kennedy (2011), Pishghadam and Shayesteh (2012), Reza Pishghadam, Adamson, and Shayesteh (2014), Vardar (2010) and Segers and Tillema (2011) all adapted the four-factor model as their theoretical framework to study the teachers' conceptions of assessment in Indonesia, USA, Australia, India, China and Hong Kong, Iran, Turkey and Netherlands, respectively.

In some cases, however, several items of each category in this model were re-examined (Brown & Remesal, 2012) and in some other, new constructs and items were developed to fit the assessment context of their study better (Brown, Hui, Yu, & Kennedy, 2011). However, several researchers agreed that all four factors in Brown's model are reflected by teachers' assessment conceptions. Thus, this model greatly supports the multidimensionality of teachers' assessment conceptions (Benson, 2014; Pishghadam, Adamson, & Shayesteh, 2014; Pishghadam & Shayesteh, 2012; Vardar, 2010).

The model proposed by Brown is a comprehensive model, as Brown (2008) discussed the purpose of another model such as Newton's model (2007) which also falls into the four-factor conceptions model. Newton (2007) suggested a model to clarify the purposes of assessment. Indeed, Newton (2007) proposed 18 uses of assessment and classified them at three levels based on the purpose of assessment distinguished, including: (a) the judgment level, (b) the impact level, and (c) the decision level.

The judgment level looks into the technical aim of an assessment event. At this level, assessment is used to obtain a specific kind of judgment, outcome or result such as a pass or a fail, a grade assigned to a percentile rank or a performance descriptor.

The second level known as the impact level involves the intended impacts of running an assessment system. At this level, a specific kind of educational or social consequence or impact may be introduced such as motivating hard work among learners, or to encouraging aligning teaching with a particular curriculum among teachers.

The third level is the decision level which applies the assessment judgment, namely the decision, action or process when recruiting the best candidate for a job. At this level, it provides guidance for teaching, learning, or monitoring educational standards over a period of time.

All these levels could come under the accountability and improvement conceptions of Brown's model. Finally, in this study, Brown's model is more relevant to this research compared to Newton's model. This is because Brown's model is suitable in guiding the collection of data by answering the research questions. The data consists of teachers' conceptions of the assessment, concerning different demographic information of teachers.

2.9 Conceptual Framework

In addressing the research questions of the study, Brown's model (2000) which incorporates teachers' conceptions of the purposes of assessment, was used in forming a conceptual framework to guide this study. The diagram below (Figure 2.1) displays the connection between teachers' conception factors from Brown's Model (2000) and several variables in the formulation of the conceptual framework.

The model used for this study is based on the importance of teachers' conceptions in conducting assessment as mentioned previously. For instance, a teacher with a conception aligned with the improvement purpose of assessment would be expected to conduct teaching and learning with characteristics associated with that conception.

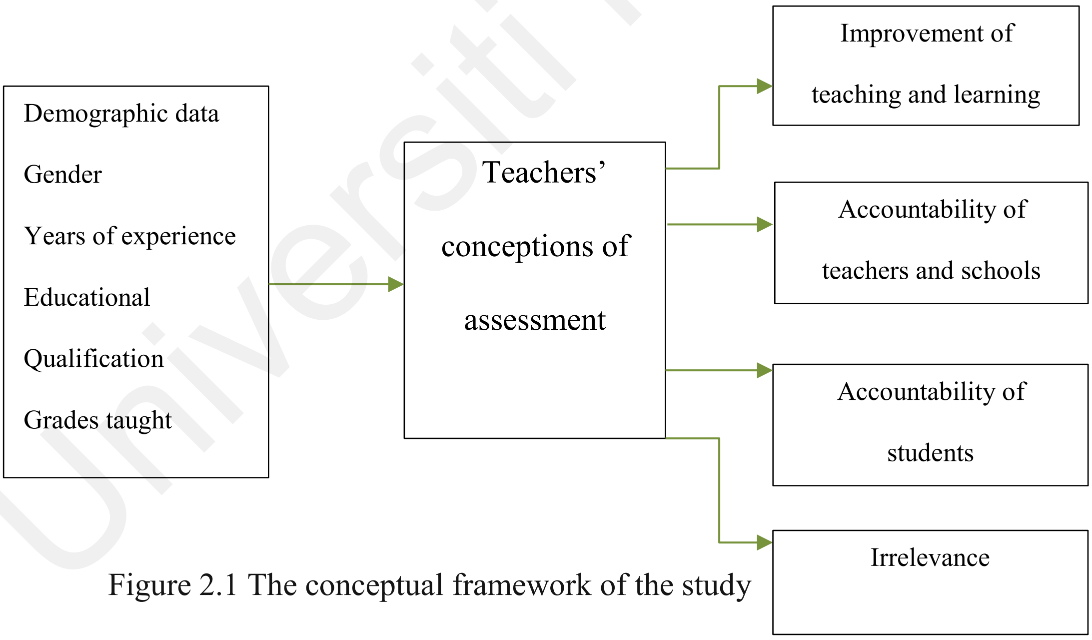

Figure 2.1 The conceptual framework of the study

The conceptual framework shows the teachers' conceptions of assessment in four categories. According to Brown (2002, 2004, 2006) teachers' conceptions of the purposes of assessment fall into one of the three main purposes and an "anti-purpose". As shown in the column on the right, these are: 1) improvement: assessment improves learning and teaching process, 2) school accountability: assessment makes teachers and schools accountable for learning, 3) student accountability: assessment makes students accountable for their learning, and 4) irrelevance: assessment is irrelevant or pernicious to the work of teachers and the life of students. These categories are completely explained as follows:

2.9.1 Conception 1: Improvement of Teaching and Learning

In the first conception, assessment is seen as the means of improving teaching and student learning which is an essential goal of assessment in educational improvement. According to Brown (2002, 2008) assessment for improvement has two important values: a) assessment must describe or diagnose the nature of student learning and b) the information must be valid, reliable, and an accurate description of student performance (Brown, 2008).

According to Brown, in this perspective, teachers use many different instructional techniques. These techniques include informal teacher-based intuitive judgments, formal evaluation tools, and identifying the content and processes of students' learning system, which, in turn, include obstacles to learning and unexpected powers with the clear-cut goal of improving instruction quality and students' learning.

In fact, it can be said that this view of conception was promoted from the view of assessment in the study by Black and Wiliam (1998). This was considered as an assessment for learning or formative assessment in contrast to the assessment of learning or summative assessment which is associated with the accountability conception. They have proven that formative assessment conveys positive impacts on education and provides proper feedback on the results of students' learning and the educational activities of teachers. It is noted that if used properly, formative assessment can result in unparalleled improvements in learning (Popham, 2011). The remarkable advantage of the improvement concept is that evaluation improves both students' learning ability as well as the teaching quality of teachers through preparing advantageous feedback, and by getting students to be involved in the evaluation process (Black & Wiliam, 1998).

Brown (2002) stated that in this conception, assessment is considered a mechanism that teachers can use to improve students' learning as well as their own teaching. In this way, improvement conception rejects the traditional view of assessment which is associated with lower order thinking such as multiple-choice tests. In other words, as improvement conception is aimed at fully diagnosing what students have learned, therefore, using a variety of assessments and tools are recommended for teachers such as performance assessment and alternative assessment that require teachers to have an adequate knowledge and concept of purpose of the assessment. Explicitly, in the improvement notion, in order to create valid changes in the teaching activities or in the learning process of students, an evaluation of students' knowledge, skills, performance as well as their way of thinking is necessary. In this way, their achievements can be facilitated and improved (Brown, 2002).

2.9.2 Conception 2: Accountability of Teachers and Schools

The second conception refers to the idea that assessment can be used to explain the use of society's resources by teachers, schools or the system (Brown, 2004). In fact, the accountability conception uses the assessment results in order to publicly show that teachers or schools are doing very well (Butterfield, Williams, & Marr, 1999).It also imposes the outcomes on the schools or teachers whether or not they have achieved the required standards

As mentioned by Brown (2004), there are two rationales for this notion: one is showing that the schools and teachers deliver quality instruction publicly, and the other one is to improve instruction quality. Brown went on to explain that the first viewpoint insists on the fact that schools and teachers must have the ability to show that they are delivering product quality which is the society's right to receive, and this is done by virtue of funding the educational processes. However, the second viewpoint insists on the role that testing can have in betterment of teachers' as well as students' work.

Brown (2008) asserted that even though the participants of his study did not endorse school accountability as an effective factor, it had a positive relationship with improvement, implying that learning would be promoted by good schools.

2.9.3 Conception 3: Accountability of Students

Student accountability refers to the idea that students independently become accountable by assessment for what they have learned by getting scores or marks, having their performance checked against the criteria, reporting their marks to their parents, to their future employers and also to other educators (Brown, 2002).

These assessments have critical consequences for students. For example, they track students into ability streams, graduation, letting them select entries to higher educational levels, retention in grade, and also the ability to have remedial education tracks. The practice of granting certificates and publicising students' performance in this way is generally taken as legitimate and crucial (Guthrie, 2002).

According to some teachers, this type of assessment imposes a necessary and motivating pressure on the students (Kahn, 2000); however, other teachers believe that the critical testing system has an adverse emotional effect on younger students and creates unnecessary extra worry and anxiety (Guthrie, 2002). In addition, some believe that testing the accountability of students was biased against certain groups, especially socially and economically lower classes of people or ethnic minority groups (Neill, 1997). Therefore, the process of marking or assessment of students creates a mixture of responses including rejection of evaluation.

2.9.4 Conception 4: Irrelevance

The fourth conception is about the fact that assessment is irrelevant. It shows that when assessment is formal, a systematic process of measuring students' performance does not have any legitimate position in teaching and learning or has no goal in relation to 'tests' or 'examinations' (Brown, 2002). This concept is founded on this opinion that in order to judge students' achievements precisely, it is unnecessary for teachers to pass these formal processes. As the teachers have interactions with students daily, they get to observe students in many different ways (they question them, observe, and discuss with them). This helps them to know how much students have progressed and how far they are from reaching the learning goals. When disqualifying assessment as irrelevant, teachers make use of terms such as tests or examinations; they refer to critical tests with an accountability purpose which, according to them, do not always show the learning and teaching activities in the classroom.

More exactly, according to Brown (2002), teachers believed that assessment endangers their everyday work as being dangerous in supporting students on the way of development, which is regarded by them as an important objective. Lastly, Brown (2002) believed that assessment might unfairly be influenced on certain students, and teachers could be made to do an assessment. They should also be entitled to choose to ignore it or assessment might be so inaccurate that it becomes unreliable. According to Brown (2004) this conception is negatively associated with improvement conception. It means that teachers who think assessment is about improvement then it is unlikely they will consider assessment as irrelevant.

As shown in the Figure 2.1, the conception of teachers may be influenced by the experiences gained through their schooling years and the academic qualification of teachers or the grade level the teachers have taught and finally by their gender. Therefore, in this study, first, the researcher aims to explore the mathematics teachers' conceptions of assessment in four categories as shown in Figure 2.1, and second, determine if there are significant differences among teachers' conceptions in the four categories based on the demographic characteristics. More explanation about teachers' characteristics is provided in the next section.

2.10 Teacher Characteristics and Assessment

This subsection presents a discussion on literature related to teachers' characteristics and assessment. It is mentioned that the process of assessment involves many activities. Teachers' involvement and use of assessment practices may differ because of different teaching responsibilities they have. Therefore, teachers may have different conceptions and teachers with different characteristics may explain varying results of assessment practices differently (Adams & Hsu, 1998).

Along with this, many researchers conducted different studies to investigate teacher-related characteristics, namely gender, years of teaching experience, academic qualification and grade level taught which may influence teachers' understanding about assessment and the methods of assessment that teachers utilize inside classroom as well as teachers' assessment skills. The following subsection describes characteristics of teachers related to assessment.

2.10.1 Teachers' Conceptions of Assessment and years of experience

Sevral studies on years of experience and teachers' conceptions of assessment showed that years of experience is one of the factors influencing teachers' beliefs about assesment, although there are some studies with dissimilar findings. For instance, Barkatsas (2003), in a stduy of secondary mathematics teachers' beliefs about assesment found that mathematics teachers with 7-9 years of experience empahsized on a problem-solving orientation to mathematics assessment more than did teachers with experience at other levels. In another study of teachers' assessment conceptions among 414 teachers in Turkey in 2010, Vardar found that the teachers' years of experience and the undergraduate institution they had graduated from led to significant differences in their conceptions about assessment (Vardar, 2010).

In another study, Keaikitse (2012) examined 691 Botswanan elementary and secondary school teachers' thoughts about classroom assessment and attempted to identify effective classroom assessment practices in teachers' views and the most used practices by these teachers. The study showed that teaching experience is important in classroom assessment practices and is practically valuable for school administrators in working with both experienced and inexperienced teachers.

Similarly, a study in 2012 among Iranian teachers in Mashhad found a significant relationship between teachers' assessment conceptions and their years of experience. Using "Teachers' Conceptions of Assessment (TCoA)" questionnaire as the study instrument, the study involved 103 teachers as the participants. In fact, findings revealed that more experienced teachers conceived of assessment as more of school accountability, which means that the school performance can be displayed by assessment (R. Pishghadam & Shayesteh, 2012). In the same way, in 2015, Moiinvaziri in a study of university teachers' conceptions of assessment showed that there was a relationship between teaching experience and school and student accountability factors. Surprisingly, in this study, as teachers' years of experience increased, they had higher values for the conception of school and student accountability (Moiinvaziri, 2015).

On the other hand, the findings of a study by Brown (2004) indicated that years of experience of the respondents did not lead to any significant differences in their conceptions about assessment. Likewise, another study by Barnes, Fives, and Dacey (2017) found that there were no significant differences between clusters on teachers' years of experience. At the same time, another study in the United State by Snyder (2017) assessed the relationship between teachers' conceptions, literacy and practices of assessment and their characteristics. The results of this study also found no significant difference between teachers' conceptions of assessment and their years of experience (Snyder, 2017).

2.10.2 Teachers' Conceptions of Assessment and Academic Qualification and Grade Level Taught

Previous research presented that academic qualification as well as grade level taught are factors that can have effect on teachers' conceptions of assessment. Calveric (2010) studied teachers' assessment conceptions and practices among 79 teachers and showed that significant relationships existed between demographics, conceptions and practices. The most significant relationship was reported to exist between teachers' assessment conceptions and their academic qualification. The analysis of teachers' assessment conceptions and practices in four levels of academic qualification, i.e. Bachelor's, Master's, postgraduate certificate, and Doctorate showed that there were significant differences between teachers with bachelor's degree and those with postgraduate certificate and Master's degree with respect to their student accountability assessment conception. The study also showed a meaningful relationship between grade level and assessment practices (Sarah B. Calveric, 2010).

Alkharusi (2011) studied teachers' classroom assessment skills in order to find the influence of gender, subject area, grade level, teachers' experience and in-service assessment training among 213 public school teachers in Oman and results showed statistically significant differences in teachers' self-perceived assessment skills across grade levels. The researcher suggested that there might be a need to match classroom assessment courses in teacher education programmes with what teachers should know about classroom practice defined by the particular grade level (Alkharusi, 2011).

In a study by Zhan and Burry -Stock (2003), a wide range of classroom assessment practices in various content areas and grade levels were examined and the findings showed that teachers used more objective type items in higher grades and teachers who had received assessment training possessed higher perceived skills in assessment practices even when teaching experience variable was controlled (Zhang & Burry-Stock, 2003).

On the other hand, a study by Neibling (2014) to explore secondary teachers' (grades 6-12) conceptions toward assessment in Kansas, USA revealed that no overall grade level differences were found between middle school and high school teachers and their conceptions of assessment (Neibling, 2014).

In another study of mathematics teachers' perceptions of classroom assessment by Susuwele-Banda (2005), it was found that academic qualification of teachers was possibly one of the factors that influenced their flexibility to accept new assessment practices. In another study on teachers' conceptions of assessment among Iranian teachers, Pishghadam and Sahebjam (2012) reported that language teachers with a PhD degree viewed assessment as more irrelevant compared to teachers who had an MA or BA degree. This suggests that teachers with a higher level of education are more likely to conceive of assessment as unfair or useless. In addition, their study showed that more experienced teachers conceived assessment as more of school accountability, which means that assessments demonstrated language institutes' performance (Pishghadam & Shayesteh, 2012).

2.10.3 Teachers' Conceptions of Assessment and Gender

Reviewing previous studies related to the association between teachers' conceptions of assessment and their gender revealed that different studies have achieved dissimilar outcomes. The findings of some studies supported the significant effect of gender on the teachers' conceptions of assessment. On the other hand, others reported no effects of gender on conceptions of assessment among the teachers participating in the studies.

The findings of a study by Ndalichako (2015) in Tanzania's secondary schools showed that there was a significant difference in the conceptions of assessment between male and female teachers participating in the study. The results showed that the females had more desirable conceptions of assessment than males.

On the other hand, several studies did not show any association between teachers' gender and their conceptions of assessment. For example, a study by Barnes et al. (2017) investigated the teachers' conceptions of the purposes of assessment in the United States. The results of this study revealed that there was no significant difference between American male and female teachers participating in this study in terms of their conceptions of assessment (Barnes et al., 2017).

Similarly, another study in 2015 among Iranian teachers in Sirjan found no significant relationship between teachers' conceptions of assessment and their gender in Sirjan University. Using “Teachers' Conceptions of Assessment (TCoA)” questionnaire as the study instrument, the study involved 147 teachers as the participants(Moiinvaziri, 2015).

Investigating the differences between teachers' assessment conceptions in terms of the demographic characteristics, a study by Benson (2014) was conducted among middle school teachers in the United States. The findings indicated that gender of the respondents did not lead to any significant differences in their conceptions about assessment.

Thus, this study sought to determine whether the independent variables (gender, years taught, degree earned, and grades taught) are related to the dependent conceptions of assessment among Iranian mathematics teachers.

2.11 Mathematics Teachers’ Conception of Assessment and Its Influence on Assessment Practices

According to Adams and Hsu (1998) research on assessment is significant to the expansion of mathematics instruction and syllabuses supporting student learning. These researchers provide several reasons for the significance of focusing on assessment. First, improving student learning is a current and valuable concentration in mathematics education and assessment is the key instrument for specifying student learning. Second, teachers’ assessment conceptions may be mismatched with variations in assessment practices promoted by the mathematics education leaders. Third, mathematics teacher educators have the potential to support the development of teachers' conceptions. Lastly, what happens in the classroom influences the views, development, and implementation of a large-scale reform (Adams & Hsu, 1998). Thus, an understanding of teachers' conceptions of assessment is important since teaching and student learning are affected by the assessment systems (Crooks, 1988). The following literature review further establishes the significance of studying mathematics assessment conceptions mainly at secondary level.

In 1997, Senk, Beckmann, and Thompson studied the activities of 19 high school teachers who taught mathematics by focusing on their assessment and grading performances. This study showed that tests and quizzes were the most frequently applied methods which built 77% of students' marks. Some 12 out of the 19 teachers went for other methods of assessment such as written projects, interviews with students, and so forth. These different methods of evaluation made up 7% of the students' marks. Additionally, they reported that test items require lower levels of encouraging thinking, questions need lesser reasoning, whereas oral and written assessment items are both significant; nevertheless, compared to a lower order questioning, it is the higher order questions that can potentially develop the thinking skills of students (Senk et al., 1997).

Indeed, the study results openly showed that making use of low-level abstract tasks by teachers did not meet the aims of reform curricula. This mismatch is likely to deprive the students who are more powerful in those aspects of the curriculum which are related to reform, because they restrict their chances to show what they can do. In addition to that, according to the reports of the researchers, knowledge and beliefs of the teachers and also the course textbooks and their contents were influential on features of the test items as well as on the features of other assessment tools (Senk et al., 1997).

Furthermore, they identified three reasons which made teachers not to use multiple-choice assessment methods. First of all, some of the teachers had limited levels of information about different types of assessment. Secondly, teachers did not feel that they had enough time to build different forms of assessment. The third point was that teachers felt that there was very little or even no professional guidelines. As a result, teachers did not have sufficient confidence to go for other types of assessments (Senk et al., 1997).

Later, Adams and Hsu (1998) conducted a study in order to explore the mathematics teachers' conceptions of assessment and their assessment practices in their classrooms at public schools in a southeastern state in the United States through a survey. The study revealed that the mathematics teachers' conceptions of assessment embraced a varied series of assessment methods and approaches. In addition, the significant finding of this study was that selecting the assessment techniques that teachers applied in the classroom and the rate of usage was based on the structures of the teachers' conceptions. In fact, these conceptions show that teachers have their own beliefs concerning which assessment technique is relevant or suitable for use in their classrooms. The findings from this study also show that teachers at an upper level such as level 3 and 4 believed the homework was important to assess students' understanding than teachers in lower levels such as level 1 and 2 (Adams & Hsu, 1998).

In 2001, Pfannkuch investigated primary and secondary mathematics teachers' beliefs and assessment practices in New Zealand. This study revealed that the majority of primary teachers believed the main purpose of assessment was to create data concerning students' development in their learning. These primary teachers employed a diversity of assessment methods. However, the secondary teachers were inclined to limit usage of new assessment approaches to non- examination mathematics classes. In other words, primary teachers perceived the purpose of assessment as an improvement of teaching and learning, whereas secondary teachers believed that assessment was to produce data concerning the students' ability to flourish in national examinations and this is close to associating with the accountability purpose of assessment (Pfannkuch, 2001).

In addition, the secondary level teachers perceived the importance of informing students about the level of their mathematical ability by using assessment methods. This study also showed that primary and secondary teachers both believed that the new assessment approaches or improvement approaches of the students have a sense of achievement, whereas the outcome of traditional assessment approaches or accountability approaches create a sense of failure in students (Pfannkuch, 2001).

In 2005, Susuwele-Banda conducted a study to investigate mathematics teachers' perceptions of classroom assessment and their current classroom assessments practices. Specifically, the study was aimed at understanding the range to which teachers practiced several classroom assessment approaches and tools to support both the learning and teaching processes. In this study, the researcher collected data among 6 mathematics teachers including three females and three males in two primary schools in Malawi through questionnaire, observation and interview. According to the data, three teachers, who believed that the purpose and nature of classroom assessment could only be fulfilled through tests, utilised teacher-centered methods to teach and their lessons had no evidence of embedded classroom evaluation. The study findings showed that teachers had confined reflective skills and also had problems in the assessment of their lessons. One of the teachers mentioned that she did not believe in self-evaluation (Susuwele-banda, 2005).

Additionally, teachers showed limited skills to analyse the works of students or their behaviours for drawing inferences. In fact, it was not in line with the curriculum. Furthermore, this study showed that only a single teacher who believed classroom assessment is a process for the progress of students' level used learner-centered approaches in classroom assessment alongside his lessons as well as some factors of classroom assessment methods. The interesting finding in this study was that the teacher who viewed assessment as a tool used it for enhancing learning and teaching. However, his lessons were teacher centered and they lacked the factors of classroom assessment practices. In this case, perhaps the teacher had the theoretical knowledge, but failed to put it into practice. Lastly, according to the findings of this study, it seemed that teachers' academic qualification influenced their flexibility to accept new ideas (Susuwele-Banda, 2005).

At the same time, Raboijane (2005) studied mathematics teachers' understanding of alternative assessment in junior secondary schools in Botswana. The study was aimed at investigating whether or not teachers at junior secondary schools in Botswana were using formative assessment when teaching mathematics. Using classroom observations and interviews among three experienced mathematics teachers, the researcher found that few teachers had attempted to use alternative assessment, partly due to their ignorance in the value of this kind of assessment. Indeed, despite the effort to improve the quality of education in Botswana including an emphasis on alternative forms of assessment, mathematics teachers' assessment strategies are still traditional rather than alternative forms of assessment. In other words, teachers insist to serve assessment for accountability purposes and only attempt to help students to pass the tests and public examinations (Raboijane, 2005).

Moreover, this study revealed that major limitation to adopt alternative classroom assessment was inadequate resources, time and large class size. Besides, the nature of education system is another factor that have influenced adopting new forms of assessment. In fact, Teachers' classroom practices are informed by their conceptions and understanding of teaching, learning and assessment of the context within which these have evolved (Raboijane, 2005).

Again in 2005, Watt examined mathematics teachers' views on alternative assessment methods in secondary schools in Sydney, Australia. The researcher utilized a questionnaire among 11 schools in order to answer three research question as follows: what kind of alternative assessment methods mathematics teachers use? what is the view of mathematics teachers about using alternative assessment methods?, and what struggle they perceived for using alternative methods? The researcher utilized a survey including eight questions, three of which were quantitative and the rest were open-ended questions (Watt, 2005).

The results of this study showed that mathematics teachers were familiar with alternative assessments but did not feel comfortable using alternative assessments in their classrooms. This study indicated that mathematics teachers believed that alternative assessments were either too unstructured, too subjective, or that they did not have the resources required to use them properly in the classroom. It was also found that teachers with more years of experience tended to stick to the traditional forms of assessment. Teachers with fewer years of experience tended to use a wider variety of assessments in their classrooms. Teachers with fewer years of experience were the least satisfied with traditional assessments; this finding is believed to be linked to how the teachers were trained. Teachers with the most years of experience were not introduced to alternative assessments when they became teachers, whereas the newer teachers were introduced to alternative assessments in their schooling. Finally, this study suggested changing the mathematics curriculum including a focus on the match between instructional goals and assessment tasks and using more alternative methods of assessment (Watt, 2005).

In 2009, Aydın, Kö, and Yıldız studied mathematics teachers' beliefs about assessment in order to examine whether teachers' beliefs about assessment were consistent with the new curriculum in Turkey. In this study, the researcher used a questionnaire and 5 interviews among 28 teachers. findings of the study showed that mathematics teachers impartially fulfilled the philosophical notion of the brand-new Turkish mathematics curriculum. No additional explanation was provided in this study (Aydın et al., 2009).

Later, Dandis (2013) conducted a case study aimed at finding out what methods secondary school mathematics teachers used to assess students. Dandis interviewed two mathematics teachers from public secondary schools in Granada (Spain) and found that 70% of student's grades were assigned based on written examinations. For the remaining 30% of student's grades, teachers marked the students' notebooks every day and checked if students did the homework or not. Teachers repeatedly held the exams when they finished a unit or a group of units, and in order to grade these exams, they used numbers (0-10). Additionally, the final marks given to the students were calculated as the average of the marks achieved in these exams (Dandis, 2013).

However, teachers stated that they were not content with the traditional evaluation methods because their results were unsatisfactory, and a lot of students had failed or had to retake some courses. In their opinion, something had to be done to solve this problem. They suggested using some alternative assessments from time to time, such as group work, competition, inviting parents to take part in the group work, and presenting the works orally. But, these forms did not affect the marks of the students (Dandis, 2013).

Moreover, teachers mentioned that they planned to use other forms of alternative assessment such as self-assessment or performance assessment in the next academic year, but they pointed out that professional development seminars need to give more information concerning alternative assessment. This study showed that although assessment reform in Spain began around 20 years before the study was carried out, the characteristics of the evaluation system the teachers were following in 2013 were similar to the characteristics of the evaluation system before the reform movement (Dandis, 2013).

Suurtamm and Koch (2014) conducted a study to provide information on how the grade 7–10 Ontario mathematics curriculum is understood and taught. Researchers aimed at showing the current thinking of teachers on new assessment to find out whether there is a mismatch between curriculum goals and teachers' current practices that could be helpful to show the need for further support. They used questionnaires consisting of 44 items and interviews to collect data. The findings of the study showed that the current teachers' assessment practices were aligned with the assessment reform aims; teachers used a variety of assessment forms to assess students' mathematics thinking regarding real world problems that were able to support students' learning. More specifically, teachers believed that to assess the students' understanding, one form of assessment is not enough and therefore they utilised more diverse methods of assessment than the time they used to assess students only for the purpose of grading (C. Suurtamm & Koch, 2014).

Indeed, the study showed that the participants had a strong belief about the importance of all students engaging in mathematical activities. In other words, they conceived assessment as a significant tool for improving mathematical thinking. In addition, teachers mentioned that the reason of their success in new assessment approaches is their involvement in cooperative professional development that assisted them in linking with other teachers. One of the teachers stated that he joined various local plans to generate and lead workshops for new educational resources related to the curriculum. Finally, the sample of the study reported that an assessment in three ways and purposes are able to support students' learning, first, by providing feedback on the students' mathematical thinking, second, by developing students' ability by using self-assessment and perform on assessment data, and third, by providing teachers with information to guide their classroom practices (C. Suurtamm & Koch, 2014).

In another study, Earl and Timperley (2014) compared two groups of mathematics teachers in a secondary school as the school recently adopted new assessment and evaluation policy in order to see how mathematics teachers present their assessment practices. The researchers found that the first group of teachers conceived of the purpose of assessment at a surface level and based on changes in their practices. Mathematics teachers used new assessment strategies and made more information about the learning outcome for students in order to help them and realize what they need to do to "pass the test". In fact, teachers had implemented practices differently but with the same purpose as in the traditional view. In the second group, the situation was different as teachers made a fundamental shift in their thinking about the purpose of assessment; therefore, they focused on students' learning needs and they used the most proper assessment tasks to find about students' understanding. They gave feedback to students in order to help them improve in the learning process (Earl & Timperley, 2014).

Regarding this, the researchers mentioned that to improve the effectiveness of new approaches to assessment practices with students, teachers are asked to explore students' conceptions and misconceptions in order to build deeper understanding of certain ideas. In the same way, professional developers need to identify and, if necessary, confront teachers' conceptions and misconceptions about assessment. In the end, the researchers emphasized that professional development in assessment should move from transmission and presentation to inquiry and conceptual change (Earl & Timperley, 2014).

Again in 2014, Gashaw conducted a study to examine secondary mathematics teachers' perceptions and practices about continuous assessment in secondary and preparatory schools. The researcher used both qualitative and quantitative data gathered through the use of questionnaire, interview, observation and document analysis. The participants of the study were 20 teachers and 152 students from 4 schools. Finding of the study showed that most participants had inadequate understanding about continuous assessment since they focused on summative aspects of assessment. However, some teachers showed a clear understanding of continuous assessment, but they didn't practice their views in their classes. Thus, teachers did utilize a variety of continuous assessment tools in their plans but did not use them in the classroom (Gashaw, 2014).

Moreover, the study showed that class size and time constraint are the biggest problems for the implementation of continuous assessment. Besides, teachers felt that continuous assessment is simply administering series of paper tests. The main reason was the majority of teachers had not received short training about continuous assessment during teaching; therefore, because of the lack of understanding, teachers did not provide enough feedback for the students during teaching and learning process (Gashaw, 2014).

Recently, Simpson (2016) conducted a qualitative study to investigate secondary mathematics teachers' beliefs toward the use of authentic assessments in the classroom. The researcher utilized interviews among nine mathematics teachers at seven high schools in the same public-school system. The study showed that secondary school mathematics teachers believed that using authentic assessments helped students to understand the significance of learning mathematics better and keep them engaged in the classroom, but teachers were reluctant to use them due to the lack of resources and time. This study recommended more professional development and materials of instruction are necessary to help the teachers maximize their use of authentic assessments (Simpson, 2016).

More recently, Arumugham, Abdullah, and Mahmud (2017) studied conceptions of formative assessment among Malaysian secondary school teachers in different subjects including mathematics. using a qualitative approach, the study aimed to explore a group of teachers' conceptions of formative assessment and how they integrated it into teaching and learning. Results of the study showed three categories of teachers' understanding of formative assessment as a testing and measuring method, as a monthly grading and formative assessment as an enrichment activity.

Indeed, from 10 participants in this study only two mathematics teachers conceived that the formative assessment is more about understanding students' learning process rather than giving grades. Indeed, teachers in this study had a lack of information regarding formative assessment. Thus, teachers intend to prepare tests which can measure students' achievement rather than preparing assessment for learning process. To overcome this situation, it is advisable to the policy makers to observe and guide the teachers to make sure that they are in the right path in achieving the objectives of national education philosophy (Arumugham et al., 2017).

Consequently, a review of these studies showed that mathematics teachers utilised diverse types of assessment based on their conceptions about the nature and purpose of assessment. Therefore, the way they conceive of assessment is the way that they interpret the results of the assessment in order to employ them in the teaching and learning process. As studies showed that majority of the teachers conceive of assessment as an accountability purpose, so they frequently use traditional assessment to assess students' performance in mathematics. However, this is against the aim of mathematics assessment reform in the 21st century. Moreover, these studies showed that to use the new assessment forms, teachers' conceptions should be considered as a significant factor having substantial influence on the quality of implementing and reporting assessment inside the classroom. Finally, if there is an expected variation in the teaching methods of the teachers, more positively, the conceptions and activities of the teachers are continuous. However, they are required to be seen as a whole sharing cognitive and emotional section of the subjects.

2.12 Review of Research that Employed Brown's Model

The purpose of this section is to review literature about the findings of studies on teachers' conceptions of assessment in order to develop an understanding of the previous research that employed the same model as the current study.

In a more inclusive view, Brown (2002) conducted comprehensive research before proposing a model with four aspects for teachers' assessment conceptions. Brown's analysis of the teachers' conceptions about learning, teaching, curriculum and teaching efficacy resulted in an important structure interconnecting the conceptions. After reviewing the relationships among conceptions as well as working on validity and reliability of the scale he had developed, Brown used structural equation modeling (SEM) to propose a model for teachers' assessment conceptions (Brown, 2002).

Brown's surveys demonstrated that there were positive relationships between three conceptions, namely student accountability, school accountability and improvement, while the relationship between the fourth conception, i.e. irrelevance and improvement was negative, it did not have any relationship with school accountability, and there was only a positive relationship between irrelevance and student accountability conception. Brown did his surveys on primary school teacher trainees, students of education at undergraduate level, and practicing primary school teachers to have a more inclusive and concrete picture for the model of teachers' evaluation conceptions, and so that their opinions were similar to those of one another (Brown, 2002).

In addition, the influences of evaluation techniques (classroom assessments held by teachers, formal exams, oral evaluations, and portfolios), evaluation activities (deep cognitive processing, informal classroom evaluation, and formal evaluation), roles of teachers, personal features of teachers (gender, role, experience, evaluation training), school characteristics (size SES), as well as teaching level (primary and secondary) were investigated and it was surprising that there were no statistically significant differences among the examined conceptions about assessment (Brown, 2002).

Brown continued his studies on teachers' conceptions of assessment in order to have a valid model of conceptions in 2006 and suggested that an abridged version of his model proposed in 2002 could serve as a valid alternative to a full-length inventory. Some of the results of his studies on teachers' conceptions of assessment are discussed below.

In order to assess Teachers' Assessment Practices, another qualitative study (based on the original Brown model) was done by Harris and Brown (2008). Overall, 169 teachers from all primary, intermediate and secondary schools in Auckland participated in this study. The results showed that teachers generally use a several types of assessment to reach different goals, regarding to their conception of assessments. The teachers participating in this study indicated that the main priority was the students' needs. Therefore, they decided to reject or change some of the mandated assessment practices based on their student abilities (Harris & Brown, 2008).

Brown (2011) also conducted a study in New Zealand to compare primary and secondary teachers regarding their conceptions of assessment using the TCoA-IIIA self-administered scale. The outcomes of this study showed that the four factors including improvement, irrelevance, school accountability, and student accountability (reported in the hierarchical model by Brown in 2002 for primary teachers) were significantly suitable for both primary and secondary groups of teachers with an exception.

The findings showed that there was a significant difference in the mean score of the student accountability conception. In fact, accountability conception was more endorsed by the participants in the secondary schools, which was consistent with their role in administering the New Zealand qualifications system. According to the results of the study, teachers preferred the conceptions of assessment that allowed them to successfully function within their own policy or legal framework (Brown, 2011).

Furthermore, based on the outcomes of another study, Brown (2012) suggested that teachers' conceptions of assessment mirrored the policies and priorities of the educational system. To design a suitable pattern for assessments and school accountability he suggested that teachers use their own beliefs and conceptions of assessment and monitor their own work to improve schooling regarding their professional capacity (Brown, 2012).

Three years later, in another study based on the original Brown's model, Brown and Gao (2015) studied Chinese teachers' conceptions of assessment with the aim to synthesize recent research on Chinese teachers' assessment conceptions and create a framework characterizing those conceptions. Secondary data from other studies was used in this study. The researchers discovered that Chinese teachers conceived of assessment as facilitating student learning and nurturing personal development and they attempt to provide these benefits for students beside their exam-oriented preparation. They specified that student accountability as well as school accountability is strongly related to the examination system in Chinese context. This study also investigated the factors that affected teachers' assessment conceptions and showed that some teacher characteristics and work environment factors significantly impacted on teachers' conceptions of assessment. Finally, they interpreted that Chinese teachers viewed assessment as playing an important role in management and control of schools and holding them accountable. Assessment was used not only to control school curriculum and teaching, but also to evaluate teachers and schools. Teachers also believed that assessment had a similar strong controlling impact on students. Success in examinations was conceived of not only as a factor forcing students to learn or a means of achieving institutional goals, but it was viewed more importantly as a duty of teachers and schools to improve students' personal and moral character (Gavin T L Brown & Gao, 2015).

Following Brown's model of teachers' conceptions of assessment, several studies utilized his model in different countries such as Iran. Some of the most recent studies are reviewed in the following section.

Recently, Barnes, Fives, and Dacey (2017) conducted a study to investigate 179 American K-12 teachers' assessment conceptions using a person-centered approach. They found that most of the teachers strongly believed that assessment was ignored, inaccurate, and bad and that the purpose of assessment was accountability. Their findings suggest that most teachers view assessment as irrelevant or interfering with teaching and question its quality. Besides, the findings showed that there were no significant differences between clusters on gender, years of teaching and level of education. However, teachers differed by cluster with regard to the grade level they were teaching. Researchers interpreted that this difference might be due to the cultural context of assessment at the district or national level. It might also reflect the quality of teacher preparation and continuing professional development with respect to classroom assessment (Barnes et al., 2017).

Based on Brown's model (2002), a study was conducted by Hidri (2016) among 336 secondary school and 206 university teachers to assess their conceptions of assessment using teachers' conceptions of assessment (TCoA) inventory. Results of the study determined a three-factor model including accountability, improvement and irrelevance. A significant relationship was found between accountability and improvement factors. The findings also revealed the inappropriate conception about assessment among the participants (Hidri, 2016).

In 2015, Moiinvaziri conducted a study to investigate Iranian university teachers' conceptions of assessment using "Teachers' Conceptions of Assessment (TCoA)" inventory (Brown, 2006). Besides, the study aimed to investigate the relationship between teachers' assessment beliefs and their teaching experience and gender. The participants of this study were 147 university teachers (86 males, 61 females) teaching in different universities and a mixed method design was used in this study (Moiinvaziri, 2015).

The results showed that the subcategory “assessment as improvement of teaching and learning” had the highest mean score of the four subcategories, while “assessment as irrelevant in teaching” was in the second place. The other two subcategories “assessment as making school and teachers accountable for their effectiveness” and “assessment as making students accountable for their learning” with very close mean scores could be categorized as the third place. However, the results of interview showed that most teachers believed in the unfairness and inaccurateness of the exams. In addition, some of them stated that they made little use of the exam results. These kinds of beliefs may have resulted in having assessment as irrelevance in the second place. Besides, results showed that there were no differences according to gender of teachers and years of experiences; however, the study showed that there was a positive relationship only between teaching experience and school and student accountability factors. Surprisingly, in this study as teachers’ years of experience increased, they would have higher values for the conception of school and student accountability. The researcher explained that this situation may originate from a change in teachers’ views as they acquire more experience through the years. As the system of education in Iran is mostly a teacher-centered system, teachers have the most authority as well as the most responsibility. Classes are usually held in the form of teacher having lectures and students are mostly passive. In universities, the authorities do not usually have any direct supervision over each class, its students or the reliability and validity of the taken exams. Therefore, the assessment results of the students may not be a good indicator of the level or quality of the university (Moiinvaziri, 2015).

Later in 2014, Benson investigated assessment conceptions and practices among middle school teachers in the United States. The study also attempted to identify differences between teachers' assessment conceptions based on their demographic characteristics. The results of this study showed that Improvement and Student Accountability were the highest perceived assessment conceptions among the participants of the study, followed by Irrelevance and School Accountability. The results also showed no significant differences between teacher characteristics such as grade level taught, academic qualification and gender. Benson believed the reason for these results was that teachers, regardless of their gender, grade level or educational background had to adapt to the local, state, or federal accountability guidelines used for student performance (Benson, 2014).

In 2013, Ludwig conducted a quantitative correlation study to explore teachers' conceptions of assessment, their self-perceived confidence to implement sound classroom assessment practices, the relationship between the two constructs, and the influence of professional characteristics among 160 American Kindergarten through 12th-grade teachers in three public schools. The results of the study showed that the participants had the highest collective self-perceived agreement on the student accountability purpose. In addition, in terms of classroom assessment confidence, the participants rated themselves highest in effective communication of results. Moreover, results revealed that having multiple assessment training opportunities was associated with a significantly higher mean factor score for Assess What than those who had no assessment training (Ludwig, 2013).

In 2012, Pishghadam and Sahebjam conducted a study to explore the assessment conceptions of 103 Iranian English teachers from different language institutes based on "Teachers' Conceptions of Assessment (TCoA)" Inventory (Brown, 2006). In this study, the researchers examined the four aspects of assessment conception with regard to their age, gender, major, academic qualification and teaching experience. The results of the study showed that Student Accountability was the main assessment conception among the EFL teachers who participated in the study. This study also revealed that language teachers with a PhD degree viewed assessment as more Irrelevant compared to teachers who had an MA or BA degree. This suggests that teachers with a higher level of education are more likely to conceive of assessment as unfair or useless. In addition, the results showed that teachers who had majored in TEFL conceived of assessment as Irrelevant more than other majors. Finally, they showed that more experienced teachers viewed assessment as School Accountability, suggesting that assessment demonstrates the performance of language institutes (R. Pishghadam & Shayesteh, 2012).

In a study by Segers and Tillema (2011) the teachers' and students' conceptions of assessment were investigated. The findings showed that there was no significant difference in the mean score of formative and summative purposes of assessment among the teachers participating in this study. Teachers recognized both formative and summative assessment as school accountability. Moreover, they believed assessment as being of negative (in terms of measurement errors) or good quality. On the other hand, students participating in the study believed that the purpose of formative and summative assessment was different (Segers & Tillema, 2011).

Involving 50 primary and secondary teachers in a qualitative study, teachers' conceptions of assessment were investigated in a study by Remesal (2011) in Barcelona, Spain. The study focused on teachers' conceptions about assessment regarding 4 subscales including learning process, teaching process, accreditation of learning, and accountability. Results showed that these factors were hardly separable from one another. In fact, the teachers' conceptions about assessment were made up of all of these factors (Remesal, 2011).

2.13 Summary

From a review of the literature, assessment is at the center of the educational reform movement and teachers are the key change agent, who are critical to success or failure of the reform. Research shows that the personal experience of teachers, as well as their conceptions, are effective on the learning ability of their students and also on their performances in the class. They are also effective on the time they spend in assessing their students in the classroom. Therefore, teachers' beliefs and conceptions about the nature of assessment should be taken into account in any mathematics educational reform. Brown (2004) emphasised that if reformers aim to succeed in implementing any new assessment practices or tools, whether at local or national level they should consider the multifaceted structure of teachers' conceptions related to assessment (Brown, 2004). Besides, if teachers' conceptions remains unchanged, it may reduce effectiveness in implementing the new assessment policy (Brown, 2004). Moreover, policy reform may fail because of teachers' misconceptions in the proposed new usage of assessment (Brown, 2004). Therefore, the current study attempts to answer the questions, "What are the mathematics teachers' conceptions of assessment in the context of Iranian educational reform in lower secondary school from grade 7 to 9?" and, "which demographic characteristics of mathematics teachers such as years of experience, educational qualification and gender have an effect on mathematics teachers' assessment conceptions? "The research design, population and sampling, data collection procedure, instrumentation, and data analysis will be discussed in Chapter Three.

Chapter 3

Research Methodology

3.1 Introduction

Chapter Three involves eight main sections, that is introduction, research design, population and sample, data collection, instrumentation, pilot study, data analysis, and summary. In the first section, the overview of the chapter is elaborated and in the second section, the research design used for conducting the research is described. In the third section, sampling method used for selecting the sample is described. The data collection method used for conducting the study is described in the fourth section. A description of the instrument and also an explanation of validity of the instrument are contained in the fifth section, in the sixth section, the pilot study used to ascertain the reliability of the instrument is described. In the seventh section, data analysis procedure is outlined and, in the eighth section, concluding remarks about the methodology are made.

3.2 Research Design

A quantitative research design was used for this research. According to Gay, Mills, and Airasian (2009) a quantitative research focuses on the collection, investigation, and explanation of numerical data. Indeed, a specific and planned research design is a significant factor within any study as it should support and align both the purpose and goal of the research. Research designs provide reasons and direct meaning to a study (Gay et al., 2009). Therefore, in this study, according to the nature of the research questions, research problem and purpose of the study the more suitable design was the quantitative research design (Creswell, 2012). The purpose of this study is to investigate in-service Iranian mathematics teachers' conceptions of assessment at three grades of lower secondary school (7th, 8th, 9th grades) in seven educational districts. Another purpose of this study is to determine whether there are differences in teachers' conceptions of assessment according to three demographic characteristics of mathematics teachers including years of experience, educational qualification and gender. Therefore, the purpose of this study is well suited for quantitative research methods.

To collect and analyse data in order to answer the research questions, this study utilized a causal-comparative research design (Creswell, 2012). Since this study did not include any treatment, there was an inherent need for a nonexperimental research design. The causal-comparative design is more appropriate for determining the differences between groups of subjects through comparing the pre-existing differences in the independent variable and measuring the result on the dependent variable (Fraenkel, Wallen, & Hyun, 2015; Gall, Gall, & Borg, 2006). Moreover, this design is suitable for this study because of the possible cause-and-effect relationship between the independent variables and the dependent variable in this study. Four groups of teachers were formed based on the independent variables (gender, years of experience, educational qualification, and grades taught) and were compared on the dependent variable (mathematics teachers' conceptions). No variables were manipulated by the researcher in this study.

3.3 Population and Sampling

The target population in this study was lower secondary mathematics teachers in public schools (female schools and male schools) in Razavi Khorasan Province, Mashhad, Iran. As little attention has been paid to mathematics teachers in Mashhad compared to Tehran, the capital city of Iran, while Mashhad is the second biggest city in Iran with a population of more than 3 million people; therefore, the researcher selected this city and its mathematics teachers as the target population in the current study. In addition, the researcher worked as a mathematics teacher in Mashhad, so she is familiar with the educational environment in this area.

This city of Mashhad is located in the northeast of Iran and it is 889 km from the capital city, Tehran that is shown in Figure 3.1 below. This city in one of the most famous cities in Iran and hosts a substantial number of visitors from around Iran and other countries every year. The city is divided into seven educational districts and each district has its own MOE office of education that working under the supervision of the central MOE Office of Razavi Khorasan Province in Mashhad.

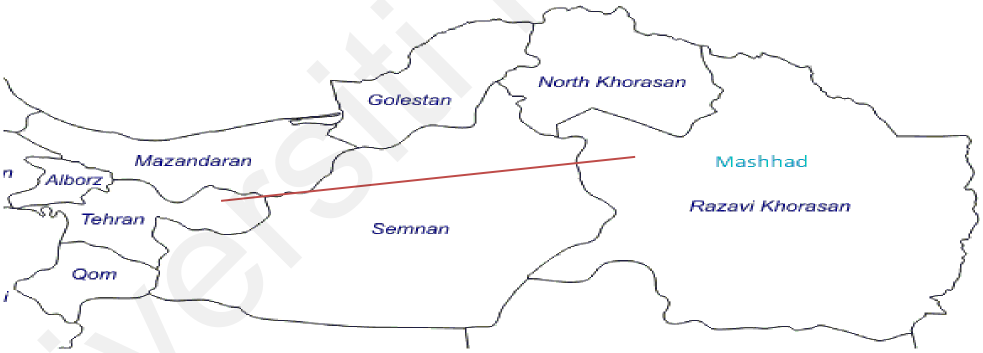

Figure 3.1 Location of Mashhad and Razavi Khorasan Province in Iran (retrieved from https://www.d-maps.com)

The target population in this study consists of mathematics teachers, who were teaching in public lower secondary schools (female and male) through grades 7 to 9 in Razavi Khorasan Province, Mashhad, Iran during the first semester of the 2017-2018 academic year. Teachers who teach in public school in Iran have different levels of academic qualification including Bachelor's, Master's or PhD degrees. Iranian lower secondary school students are mainly assessed by classroom assessments or teacher-made tests. Students take standardized exams only when they complete grade 9 at lower secondary school to move on to the next level. The construction and administration of the standardized examinations are carried out by testing and measurement specialists of the MOE in each city of Iran. However, administration of these exams is the responsibility of teachers in schools.

3.3.1 Sampling

The simple random sampling method was used to select the sample population in the present study. Since surveys attempt to generalize sample statistics (external validity), a probability sample of the population is usually used. The aim of survey sampling is selecting a set of elements from a population in such a manner that those elements give an accurate description of the total population from which they are selected.

The simple random sampling is the most rigorous and popular type of probability sampling from a population. In this method, any individual has an equal chance of being selected from the population and the individuals should be representative of the total population. Thus, biases in the population will be equally distributed among the selected individuals (Creswell, 2008).

In the random sampling method, it is important to know the population' therefore, first, the researcher collected all information about mathematics teachers including number of teachers, their email addresses and also the educational district that they were working during the 1st semester of 2017-2018. So, the population included 567 mathematics teachers. After getting permission from the Ministry of Education Office and creating sampling frame, the researcher used a random number generator to select the sample and finally in order to have a big enough sample size, the researcher selected 352 teachers randomly and sent emails to them and all these teachers accepted to join the survey.

3.4 Instrumentation

The instrument used for collecting data in this study is called Conceptions of Assessment-III (COA-III) Abridged, which is a questionnaire adopted from Brown (2006). The permission to use this instrument was obtained from the questionnaire's developer, Dr. Gavin Brown,via email (Appendix C). This instrument has been used in various studies in the literature (Azis, 2015; Barnes et al., 2017; Pishghadam & Shayesteh, 2012, Pishghadam, Adamson, & Shayesteh, 2014).

The abridged version of COA-III questionnaire consisting of 27 items was used to measure teachers' conceptions related to assessment purposes. This questionnaire has two main sections. The first section contains questions about the teachers' gender, demographic information, academic qualification, teaching experience and grade level and the school district where they teach. The second part contains items about teachers' theoretical assessment conceptions including 12 items for improvement, 3 items for student accountability, 3 items for school accountability and 9 items for irrelevance. Items are evaluated based on a 6-point Likert scale (1-strongly disagree, 2-mostly disagree, 3-slightly agree, 4-moderately agree, 5-mostly agree, and 6-strongly agree). The scale is said to be "positively packed", since there are two negative (mostly and strongly disagree) and four positive (slightly, moderately, mostly, and strongly agree) options for the respondents to choose from for each item (See Appendix A).

The instrument used for the current study was previously validated in the Iranian context and translated into Persian (Pishghadam et al., 2014; Pishghadam & Shayesteh, 2012) in order to study university EFL teachers' conceptions of assessment in Mashhad, Iran (For the link of an online survey including the Persian version of the instrument, see Appendix B).

To ensure the validity of the Persian version, researchers in previous studies conducted confirmatory factor analysis of data and the results showed that the translated version of the instrument was well-matched with the original one. Table 3.2 reports the results of the analysis adopted from the study by Pishghadam et al. (2014). And Table 3.3 gives the breakdown of items in each section of the questionnaire.

Table 3.1

Goodness of fit Indices

	χ^2	Df	χ^2/df	GFI	TLI	CFI	RMSEA
TCoA	5.52	2	2.67	.97	.95	.99	.05

Table 3.2

Instrument

Section	*Items*	*Items number*
Section 1	*4Demographic items*	
Section 2	*improvement 12 items*	*3, 4, 5, 6, 12, 13, 14, 15, 21, 22, 23, 24*
	students accountable 3 items	*2, 11, 20*
	schools accountable 3items	*1,10,19*
	irrelevant 9 items	*7, 8, 9 ,16,17,18, 25, 26, 27*

3.4.1 Validity

Validity is the extent to which any measuring instrument measures what it is intended to measure (Fraenkel et al., 2015). In this study, content validity can be established by the judgment of an expert panel (Radhakrishna, 2007). Although the instrument was validated in Iranian content in the previous studies mentioned above, it was necessary to test the instrument's validity again since this study was carried out on the teachers of a different major and level, i.e. mathematics in lower secondary schools. Thus, an expert panel was utilized for the purpose of validating the instrument used in this research. The panel consisted of three Iranian lecturers in the field of mathematics education with strong knowledge of assessment in mathematics. The instrument, research questions as well as research objectives were sent to these experts via email. They assessed and confirmed the validity of the instrument. Finally, the instrument was used in the pilot study to analyse its reliability.

3.5 Pilot Study

A pilot study is defined as a small-scale trial of the proposed procedures (Fraenkel et al., 2015). According to Mertens (2010) the pilot study is used for several purposes. First, it is used to ascertain the reliability of the instrument, and second, to identify the strengths and weaknesses of a questionnaire with regard to the question format, and the items' wording and order. Third, to help recognize item difficulty, meaning, question variation, and respondents' attention and interest in answering individual items, and aslo to establish item-response relationships, and check item-response reliability (Mertens, 2010).

In this study, the most important function of the pilot study is to analyze the instrument reliability. A link to a pilot study along with an explanation of the intent of the study was sent out via email to a random sample of approximately 86 lower secondary mathematics teachers (male and female). A total of 50 responses were submitted electronically. These respondents were 50% male and 50% female with 44.40% Master's, 46.3% Bachelor's and only 7.4% PhD holders. Some 31.5% of respondents had more than 20 years of experience in teaching mathematics, 25.9% of them had 16 -20 years of experience, 22.2% had 6-10 years of experience, 11.1% had 5-10 years of experience, and 9.3% of them had 11- 15 years of teaching experience. The reliability of the instrument is explained below.

3.5.1 Reliability

Reliability in a research means that the scores from the instrument are stable and consistent when the researcher administers the instrument multiple times and the scores should remain nearly the same (Creswell, 2012). In the view of Twycross and Shields (2004) reliability refers to the consistency, stability and repeatability of results explicitly. The result of a researcher is considered reliable if consistent results have been obtained in identical situations but different circumstances (Twycross & Shields, 2004).

The aspect of reliability that is being stressed in this study is homogeneity or internal consistency. To measure internal consistency, researchers commonly use Cronbach's alpha as an indicator of reliability. According to Gay (1996), Cronbach's alpha is the most appropriate way of establishing reliability. Therefore, the reliability in this study is determined by conducting Cronbach's alpha level calculation using the Statistical Package for the Social Sciences (SPSS Version 24.0). Evidently, it has been stated by Nunnally (1978) and others that an item with alpha value of 0.70 and above is deemed as reliable and acceptable.

Develles (1991) defined Cronbach's α as $\alpha = \frac{K}{K-1}\left(\frac{1-\sum_{i}^{k}\sigma_{yi}^{2}}{\sigma_{x}^{2}}\right)$ where K is the number of components (K-items or testlets), σ_{x}^{2} the variance of the observed total test scores, and σ_{yi}^{2} the variance of component i for the current sample of persons.

The Cronbach's alpha was computed for each factor (subscale) of the instrument. In this study, the instrument has measured the scores successfully, showed consistency and reliability of more than 0.7 for each factor; therefore, the researcher was able to interpret the scores to make a valid conclusion. The reliability value of the instrument for each factor based on the data collected from the pilot study is shown in Table 3.3 below.

Table 3.3

Overall Subscale Reliability

Subscales	**Cronbach's alpha**
Improvement	.89
School accountability	.71
Students accountability	.84
Irrelevance	.78

3.6 Data Collection

This study utilized a quantitative method (questionnaire) to collect data for addressing the problem of the study through descriptive statistics and inferential statistics (Creswell, 2008).

The quantitative method was chosen in the study for several reasons. First, quantitative methods allow the researcher to collect enormous amounts of data from various groups of people in a brief time period and in a moderately profitable way.

Second, questionnaires are a feasible method for gathering information. They can be targeted to groups of your choosing and managed in several ways. They enable researchers to pick and choose the questions asked as well as the format (open-ended , or multiple choice). Questionnaires offer a way to collect massive amounts of information on any issue. Third, questionnaire results are quickly and easily measured by either a researcher or by using a software package. In addition, results can be analyzed more systematically and accurately than with other methods of research. Once information has been measured, it can be used to compare with results from other studies and might be utilized to quantify change. Positivists believe that quantitative data can be used to create new theories and/or test existing hypotheses (Fraenkel et al., 2015).

The researcher gathered the data for this study via an online survey. The survey was administered after getting permission from the Ministry of Education in Mashhad in the new Iranian academic year that was starting from September 2017. After selecting mathematics teachers randomly, the researcher collected all information including the schools where teachers worked and their email addresses and the contact number of each school. In this way, the researcher emailed the link of the online survey to the school and called each school, explained the value of the research to the school principals and asked them to ensure each mathematics teacher selected completed the electronic Teachers' Conceptions of Assessment Abridged Survey (COA-III; Brown, 2006).

3.7 Data Analysis

The statistical techniques used in this study were selected because of the design of the study. Descriptive and inferential analyses are conducted to address each facet of the objectives, research questions and hypotheses of the present study. In this study, question one (see Table 3.4) is analyzed using descriptive statistics such as means, medians, modes, standard deviations, frequencies, and percentile to provide general information about the teachers' conceptions of assessment.

Table 3.4

Data Analysis Method

	Research Question	Statistical Technique
1-	What are the mathematics teachers' conceptions of assessment?	*Descriptive*
2-	Do mathematics teachers' conceptions of assessment have significant difference in terms of years of experience?	*MANOVA*
3-	Do mathematics teachers' conceptions of assessment have significant difference in terms of academic qualification?	*MANOVA*
4-	Do mathematics teachers' conceptions of assessment have significant difference in terms of gender?	*MANOVA*
5-	Do mathematics teachers' conceptions of assessment have significant difference in terms of grade taught?	*MANOVA*

One-way Multivariate Analysis of Variance (MANOVA) was run in order to answer questions 2 to 5. According to Pallant (2007) the MANOVA test functions just like the one-way ANOVA test. It means that both tests answer the same research question on the difference between groups; however, MANOVA test is used to analyse two or more dependent variables, while the ANOVA test is only used for analyzing one dependent variable.

Using the MANOVA is more useful compared with a series of ANOVAs for each independent variable because in this way, there is chance to adjust the risk for Type 1 error when there is more than one independent variable (Pallant, 2007), also once the dependent variable is interval (teachers' conception of assessment). Four separates Multivariate Analyses of Variance (MANOVA) were run to determine how teacher characteristics (years of experience, academic education, gender, and grade level) are related to their "conception of assessment" factors (Improvement, School accountability, Students accountability, and Irrelevance). However, before running the Multivariate Analysis of Variance (MANOVA) it is necessary to check the assumptions of One-way Multivariate Analysis of Variance (MANOVA). According to Pallant (2007) Multivariate Analysis of Variance (MANOVA) has several assumptions which are sample size, normality, and homogeneity of variance and covariance matrices. If the data of the study do not fulfill these assumptions, the alternative test of data analysis will be Kruskal-Wallis and Mann-Whitney test.

3.8 Summary

This Chapter discussed the methodology of the present study. The research design, sampling method, instrumentation, data collection and analysis procedure had been clearly explained in this Chapter. Besides, the validity and reliability of the study were also explained. Furthermore, the process of conducting the pilot study was clarified. In Chapter Four, the actual analysis of the data and the results of the study will be presented. Chapter Five will give a summary of the study, summary of research results, discussion, conclusion, the implications of these findings for the model used, and recommendations for future studies.

Chapter 4

Findings

4.1 Introduction

This chapter presents the findings of the quantitative study on Iranian secondary school mathematics teachers' conceptions of assessment. The main purpose of the study was to investigate in-service Iranian mathematics teachers' conceptions of assessment regarding the purpose of assessment at three grades of lower secondary school (7th, 8th, 9th grades) in seven educational districts in Mashhad, the second biggest city in Iran. Specifically, the study examined teachers' conceptions related to four major assessment purposes: school accountability, student accountability, improvement in teaching and learning, and irrelevance of assessment. The abridged *Conceptions of Assessment Inventory* (CoA-IIIA) with 27 items (Brown, 2002) was used in the study to determine the teachers' level of agreement using a 6-point Likert scale (1= *strongly disagree* and 6 = *strongly agree*). Collecting demographic information made it possible for the researcher to consider years of experience, academic qualification (highest degree), gender, and grade level variables in the study as well.

Furthermore, this chapter contains the results of hypothesis-testing procedure used for each of the hypotheses related to the research questions. Descriptive statistics was used to analyze research question one, and One-way Multivariate Analysis of Variance (MANOVA) was employed in order to answer research questions 2,3,4, and 5. Prior to testing any hypothesis related to the research questions of this study, the assumptions required to run the hypothesis tests were examined. All hypotheses were evaluated at the 5% level of significance.

Following are the results from demographic information and survey responses which are presented within the framework of the following research questions.

1. What are the mathematics teachers' conceptions of assessment?

2. Do Iranian mathematics teachers' conceptions of assessment have significant difference in terms of years of experience?

3- Do Iranian mathematics teachers' conceptions of assessment have significant difference in terms of academic qualification?

4- Do Iranian mathematics teachers' conceptions of assessment have significant difference in terms of grade taught?

5- Do Iranian mathematics teachers' conceptions of assessment have significant difference in terms of gender?

4.2 Response Rate of Survey

The target population of the study consists of 567, 7th through 9th grade mathematics teachers in seven educational districts in Mashhad. Seven educational districts agreed to participate in the research, so 352 out of the 567 mathematics teachers participated in the study with an overall response rate of 62.08%.

4.3 Demographic Data of Respondents

The demographic data of the respondent are obtained in the first section of the questionnaire. These include the years of teaching experience in mathematics, academic qualification (highest qualification), gender and grade level that teachers teach. The next subsection will elaborate on the frequency and percentage of each demographic factor.

4.3.1 Teaching Experience

This subsection describes the respondents' teaching experience in mathematics. The data collected shows that 34 (9.7%) of the respondents have less than 5 years of experience in teaching mathematics. There are about 51 (14.5%) of the respondents who have 6 to 10 years of teaching experience, 66 (18.8%) of them have 11 to15 years of experience in teaching and 84 (23.9%) of the respondents have 16 to 20 years of teaching experience (Table 4.1). Majority of the respondents which are 117 (33.2%) of the teachers have more than 20 years of teaching experience. This means that most of the respondents are relatively considered as senior mathematics teachers.

Table 4.1

Respondents' Years of experience in teaching mathematics (N=352)

Category	Frequency	Percentage
1-5 years	34	9.7
6-10 years	51	14.5
11-15 years	66	18.8
16-20 years	84	23.9
More than 20 years	117	33.2
Total	352	100.0

4.3.2 Academic Qualification

Table 4.2 below shows that the highest qualification of the respondents in this study. It shows that most of the teachers that are 194 (55.1 %) in schools have a Bachelor's degree as their highest qualification, followed by Master holders who are 126 (35.8 %) of them. There are a small number of respondents with a Doctoral degree, i.e. 32 (9.1

%) of the respondents. This shows that most of the mathematics teachers have a minimum qualification of Bachelor's degree in order to teach in a school.

Table 4.2

Highest Academic Qualification of Respondents (N=352)

Category	Frequency	Percentage
Bachelor's Degree	194	55.1
Master's Degree	126	35.8
Doctoral Degree	32	9.1
Total	352	100.0

4.3.3 Grade Level Taught

Table 4.3 shows the frequency and percentage of grades taught by mathematics teachers in their schools at the academic year 2017-2018 in Mashhad.

Table 4.3

Respondents' Grade Level of teaching at school (N=352)

Category	Frequency	Percentage
Level 7	98	27.8
Level 8	153	43.5
Level 9	101	28.7
Total	352	100.0

As shown in Table 4.3, most of the teachers who are 153 (43.5 %) of the respondents teach in grade 8, 101 (28.7 %) of them teach in grade 9 and 98 (27.8 %) of the teachers teach in grade 7.

4.3.4 Gender

The summary of the distribution of male and female participants is presented in Table 4.4, where it shows that from the total of 352 respondents, there are 198 (56.3 %) female respondents and remaining 154 (43.8 %) are male respondents. This statistic shows the majority of responses are female teachers compared to male mathematics teachers.

Table 4.4

Distribution of female and male respondents (N=352)

Category	Frequency	Percentage
Female	198	56.3
Male	154	43.8
Total	352	100.0

4.4 Result of Analysis for Research Question 1

RQ1. What are mathematics teachers' conceptions of assessment?

The mean score and standard deviation for four factors and items in the current survey were used to identify the mathematics teachers' conceptions of assessment among Iranian secondary school mathematics teachers in Mashhad.

Table 4.5

Descriptive Statistics of mathematics Teachers' conceptions of assessment (MTCoA)

Mathematics Teachers' conceptions of assessment	*M*	*S D*
Student Accountability (STACC)	4.02	.72
School Accountability (SACC)	3.89	.74
Improvement (IMP)	3.57	.45
Irrelevance (IRR)	3.49	.51

M= Mean
SD= Standard Deviation

Table 4.5 shows that student accountability (*M*= 4.02, *SD*= .72) had the highest agreement level followed by school accountability (*M*=3.89, *SD*=.74). The two conceptions were not widely different, and participants tended to answer "Moderate Agreement" level among other variables. On the other hand, the Irrelevance conception (*M*= 3.49, *SD*= .51) had the smallest mean and the other remaining conception is improvement (*M*= 3.57, *SD*= .45), so these two conceptions could be considered as being around "Slight Agreement" level among other variables. Therefore, as seen in Table 4.5, mathematics teachers mostly conceived that assessment makes school and students more accountable.

4.5 Descriptive Statistics for School Accountability Items

Table 4.6 presents descriptive statistics for school accountability items that were three items worth up to 6 points. As shown in Table 4.6, Question 1, "Assessment provides information on how well schools are doing" (M=4.11, SD=1.10) has the highest level of agreement for school accountability whereas the other two items which are Question 10, "Assessment is an accurate indicator of a school's quality" (M=3.67, SD= 1.19), and Question 19, "Assessment is a good way to evaluate a school"

(M=3.92, SD=1.14) have lower level of agreement for school accountability. Therefore, it can be concluded that most of the teachers conceived that assessment provides information on the performance of schools.

Table 4.6

Descriptive Statistics for School Accountability Items

Items		M	SD
1	Assessment provides information on how well schools are doing	4.11	1.10
10	Assessment is an accurate indicator of a school's quality	3.67	1.19
19	Assessment is a good way to evaluate a school	3.92	1.14

M= Mean
SD= Standard Deviation

Table 4.7 presents a clear breakdown of the frequency and percentage of School Accountability items that were three items worth up to 6 points. As shown in Table 4.7, 251 teachers (70.7%) moderately to strongly agreed that Assessment provides information on how well schools are doing. Additionally, 187 teachers (53.2%) indicated moderate to strong agreement with the idea that Assessment is an accurate indicator of a school's quality. Finally, with regards to Assessment is a good way to evaluate a school, 219 teachers (62.3%) expressed moderate to strong agreement.

Table 4.7

Frequency and Percentage of School Accountability Items

Items	1	10	19
Response choice		Frequency (%)	
strongly disagree	4 (1.1%)	12 (3.4%)	6 (1.7%)
mostly disagree	20 (5.7%)	40 (11.4%)	24(6.8%)
Slightly agree	77 (21.9%)	113 (32.1%)	103(29.3%)
moderately agree	122 (34.7%)	97 (27.6%)	109 (31.0%)
mostly agree	91 (25.9%)	68 (19.3%)	78 (22.2%)
strongly agree	38 (10.8%)	22 (6.3%)	32(9.1%)
Total	352 (100.0%)	352 (100.0%)	352(100.0%)

4.6 Descriptive Statistics for Student Accountability Items

Table 4.8 illustrates descriptive statistics for student accountability items that were three items worth up to 6 points. As shown in Table 4.8 Question 2, "Assessment places students into categories" (M=4.16, SD=1.24) has the highest level of agreement for student accountability, whereas the other two items which are Question 11, "Assessment is assigning a grade or level to student work" (M=3.99, SD=1.10) and Question 20, "Assessment determines if students meet qualification standards" (M=3.89, SD=1.14) has lower levels of agreement for student accountability. Therefore, the information with the mean value of 4.16 and SD =1.24 shows that most of the teachers conceived of assessment as a tool to place students into categories.

Table 4.8

Descriptive Statistics for Student Accountability Items

Items		*M*	*SD*
2	Assessment places students into categories	4.17	1.24
11	Assessment is assigning a grade or level to student work	4.03	1.10
20	Assessment determines if students meet qualifications standards	3.89	1.14

M= Mean
SD= Standard Deviation

Table 4.9 presents a clear breakdown of the frequency and percentage of Students Accountability items that were three items worth up to 6 points. As shown in Table 4.9, 251 teachers (71.3%) moderately to strongly agreed that Assessment places students into categories. Additionally, 239 teachers (67.9%) indicated moderate to strong agreement with the idea that Assessment is assigning a grade or level to student work. Finally, with regards to Assessment determines if students meet qualifications standards 219 teachers (62.3%) expressed moderate to strong agreement.

Table 4.9

Frequency and Percentage of Student Accountability Items

Items	2	11	20
Response choice		Frequency (%)	
strongly disagree	12 (3.4%)	6 (1.7%)	5 (1.4%)
mostly disagree	18 (5.1%)	17 (4.8%)	34(9.7%)
Slightly agree	71(20.2%)	90 (25.6%)	94(26.7%)
moderately agree	96 (27.3%)	119 (33.8%)	108 (30.7%)
mostly agree	106 (30.1%)	89 (25.3%)	84 (23.9%)
strongly agree	49 (13.9%)	31 (8.8%)	27(7.7%)
Total	352 (100.0%)	352 (100.0%)	352(100.0%)

4.7 Descriptive Statistics for Improvement Items

Table 4.10 depicts descriptive statistics for improvement items that are twelve items worth up to 6 points. As shown in Table 4.10 Question 4, "Assessment provides feedback to student about their performance" (M=4.44, SD=.95) and question 5, "Assessment is integrated with teaching practice" (M=4.15, SD=1.02) have the highest level of agreement for improvement whereas item 15, "Assessment results are consistent" (M=3.15, SD=1.17) according to the Table below has the lowest level of agreement for improvement.

Table 4.10

Descriptive Statistics for Improvement Items

Items		M	SD
3	Assessment is a way to determine how much students have learned from teaching	3.91	.91
4	Assessment provides feedback to students about their performance	4.44	.95
5	Assessment is integrated with teaching practice	4.15	1.02
6	Assessment results are trustworthy	3.76	.87
12	Assessment establishes what students have learned	3.37	1.15
13	Assessment feeds back to students their learning needs	3.30	1.03
14	Assessment information modifies ongoing teaching of student	3.51	.99
15	Assessment results are consistent	3.15	1.17
21	Assessment measures students' higher order thinking skills	3.43	.781
22	Assessment helps students improve their learning	3.41	1.07
23	Assessment allows different students to get different instruction	3.22	1.12
24	Assessment results can be depended on	3.24	1.13

M=Mean
SD= Standard Deviation

Table 4.11 presents a clear breakdown of the frequency and percentage of Improvement Items that were twelve items worth up to 6 points. As shown in Table 4.11, 210 teachers (59.7%) moderately to strongly agreed that Assessment is a way to determine how much students have learned from teaching, 287 teachers (81.6%) indicated moderate to strong agreement with the idea that Assessment provides feedback to students about their performance, 251 teachers (71.3%) indicated moderate to strong agreement with the idea Assessment is integrated with teaching practice. 192 teachers (54.5%) moderately to strongly agreed that Assessment is a way to determine Assessment results are trustworthy, this data following by 151 teachers (42.8%) indicated moderate to strong agreement with the idea Assessment establishes what students have learned. As table shown that only 121 teachers (34.3%) moderately to strongly agreed that Assessment feeds back to students their learning needs. Moreover, only 152 teachers (43.25%) indicated moderate to strong agreement with the idea Assessment information modifies ongoing teaching of student. This data following by 110 teachers (31.3%) moderately to strongly agreed that Assessment results are consistent. Additionally, only 107 teachers (39.4 %) indicated moderate to strong agreement with the idea Assessment measures students' higher order thinking skills. Data from table showed that 144 teachers (40.5%) provides agreement on Assessment helps students improve their learning. This is following by 121 teachers (34.2%) indicated agreement on this idea that Assessment allows different students to get different instruction. Finally, with regards to Assessment results can be depended on 131 teachers (37.2%) expressed moderate to strong agreement.

Table 4.11

Frequency and Percentage of Improvement Items

Items	3	4	5	6	12	13	14	15	21	22	23	24
Response choice						Frequency (%)						
strongly disagree	0	0	0	1 (.3%)	16 (4.5%)	7 2.0%)	3 (.9%)	26 (7.4%)	0	1 (.3%)	0	1 (.3%)
mostly disagree	1 (.3%)	0	9 (2.6%)	5 (2.6%)	65 (18.5%)	62 (17.6%)	37 (10.5%)	67 (19.0%)	6 (1.7%)	71 (20.0%)	108 (30.7%)	111 (31.5%)
Slightly agree	141 (40.1%)	65 (18.5%)	92 (26.1%)	154 (43.8%)	120 (34.1%)	162 (46.0%)	160 (45.5%)	149 (42.3%)	239 (67.9%)	136 (38.6%)	123 (34.9%)	109 (31.0%)
moderately agree	115 (32.7%)	119 (33.8%)	128 (36.4%)	119 (33.8%)	80 (22.7%)	66 (18.8%)	90 (25.6%)	59 (16.8%)	63 (17.9%)	82 (23.3%)	71 (20.0%)	75 (21.3%)
mostly agree	77 (21.9%)	116 (33.0%)	82 (23.3%)	63 (17.9%)	67 (19.0%)	49 (13.9%)	52 (14.85%)	40 (11.4%)	37 (10.5%)	50 (14.2%)	34 (9.7%)	44 (12.5%)
strongly agree	18 (5.1%)	52 (14.8%)	41 (11.6%)	10 (2.8%)	4 (1.1%)	6 (1.7%)	10 (2.8%)	11 (3.1%)	7 (2.0%)	12 (3.4%)	16 (4.5%)	12 (3.4%)
Total	352 (100.0%)	352 (100,0%)	325 (100.0%)	352 (100.0%)	352 (100.0%)	352 (100.0%)	352 (100.0%)	352 (100.0%)	352 (100.0%)	352 (100.0%)	352 (100.0%)	352 (100.0%)

4.8 Descriptive statistics for Irrelevance Items

Table 4.12 presents descriptive statistics for irrelevance items that are nine items worth up to 6 points. As shown in Table 4.12, Question 18, “Teachers should take into account the error and imprecision in all assessment” (M=4.05, SD=1.25) and Question 9, “Assessment results should be treated cautiously because of measurement error” (M=4.00, SD=1.15) have the highest level of agreement for irrelevance, whereas Item 7, “Assessment forces teachers to teach in a way against their beliefs” (M=3.13, SD=1.03) has the lowest level of agreement for irrelevance conception.

Table 4.12

Descriptive Statistics for Irrelevance Items

Items		*M*	*SD*
7	Assessment forces teachers to teach in a way against their beliefs	3.13	1.03
8	Teachers conduct assessments but make little use of the results	3.38	1.06
9	Assessment results should be treated cautiously because of measurement error	4.00	1.15
16	Assessment is unfair to students	3.51	1.26
17	Assessment results are filed & ignored	3.28	1.22
18	Teachers should take into account the error and imprecision in all assessment	4.05	1.25
25	Assessment interferes with teaching	3.38	1.31
26	Assessment has little impact on teaching	3.20	1.28
27	Assessment is an imprecise process	3.55	1.28

M= Mean
SD= Standard Deviation

Table 4.13 presents a clear breakdown of the frequency and percentage of Irrelevance Items that were nine items worth up to 6 points. As shown in Table 4.13, 262 Teachers (74.5%) moderately to strongly agreed that Assessment forces teachers to teach in a way against their beliefs. Teachers should take into account the error and imprecision in all assessment.147 Teachers (41.7%) indicated moderate to strong agreement with the idea that Teachers conduct assessments but make little use of the results, 225 teachers (63.9 %) indicated moderate to strong agreement with the idea Assessment results should be treated cautiously because of measurement error. 179 teachers (50.9%) moderately to strongly agreed that Assessment is unfair to students, this data following by 143 teachers (40.6%) indicated moderate to strong agreement with the idea Assessment results are failed and ignored. As table shown that 235 teachers (67.1%) moderately to strongly agreed that Assessment interferes with teaching. Moreover, 171teachers (48.5%) indicated moderate to strong agreement with the idea Assessment has little impact on teaching. Finally, with regards to Assessment is an imprecise process,175 teachers (51.7%) expressed moderate to strong agreement.

Table 4.13

Frequency and Percent of Irrelevance Items

Items	7	8	9	16	17	18	25	26	27
Response choice					Frequency (%)				
strongly disagree	--	14 (4.0%)	4 (1.1)	18 (5.1%)	27 (7.7%)	10 (2.8%)	29 (8.2%)	31 (8.8%)	18 (5.1%)
mostly disagree	19 (5.4%)	49 (13.9%)	23 (6.5%)	62 (17.6%)	66 (18.8%)	29 (8.2%)	68 (19.3%)	77 (21.9%)	56 (15.9%)
Slightly agree	71 (20.2%)	142 (40.3%)	100 (28.4%)	93 (26.4%)	116 (33.0%)	78 (22.2%)	84 (23.9%)	109 (31.0%)	103 (29.3%)
moderately agree	145 (41.2%)	89 (25.3%)	107 (30.4%)	101 (28.7%)	75 (21.3%)	97 (27.6%)	98 (27.8%)	79 (22.4%)	92 (26.1%)
mostly agree	78 (22.2%)	54 (15.3%)	77 (21.9%)	57 (16.2%)	61 (17.3%)	93 (26.7%)	57 (16.2%)	38 (10.8%)	56 (15.9%)
strongly agree	39 (11.1%)	4 (1.1%)	41 (11.6%)	21 (6.0%)	7 (2.0%)	45 (12.8%)	16 (4.5%)	18 (5.1%)	27 (7.7%)
Total	352 (100.0%)	352 (100,0%)	325 (100.0%)	352 (100.0%)	352 (100.0%)	352 (100.0%)	352 (100.0%)	352 (100.0%)	352 (100.0%)

4.9 Assumptions of MANOVA (Multivariate Analysis of Variance)

Multivariate Analysis of Variance was carried out to answer research questions 2 to 5 and also test the hypotheses in the study. Before performing MANOVA, the assumptions of the Multivariate Analysis of Variance (MANOVA) were tested and no violation for conducting MANOVA was found. According to Pallant (2007) there are several assumptions for Multivariate Analysis of Variance (MANOVA): sample size, normality, outliers and homogeneity of variance-covariance matrices.

4.10 Sample Size

According to Pallant (2007) when using MANOVA, it is important to have more cases than dependent variables (DVs) in every cell. In other words, based on rule of thumb the number of participants in each cell should be more than the number of DVs. In this study the number of the dependent variables were four, this is the minimum number we need in each cell. There are sufficient reliable participants in each cell in this study (with the small effect size with approximately 30 participants per cell). Therefore, by conducting G*Power a sufficient sample size was determined using an alpha of 0.05, a power of 0.80, and a small effect size ($f = 0.10$) (Faul, Erdfelder, Buchner, & Lang, 2013). Based on the assumptions, the desired sample size is 280 for the current study.

4.11 Normality Test

In checking assumptions of normality, the Kolmogorov-Smirnov and the skewness and kurtosis z-scores were carried out with SPSS 24. In addition, the histogram, normal Q-Q plots and box plots for four factors were visually inspected and they indicated that the data were approximately normally distributed.

Table 4.14

Kolmogorov-Smirnov Test of Teachers' conceptions of assessment in four dimensions

	Kolmogorov-Smirnov[a]		
	Statistic	*Df*	*Sig*
Student Accountability	.099	352	.000
School Accountability	.114	352	.000
Improvement	.091	352	.000
Irrelevance	.067	352	.001

a. Lilliefors Significance Correction

Table 4.14 shows the Kolmogorov-Smirnov values for teachers' conceptions of assessment in four factors (dimensions). It can be observed that the significant value is less than 0.05; therefore, it will be assumed that the data are not normally distributed. However, Table 4.15 indicates that the skewness and kurtosis values for teachers' conceptions of assessment in four factors (dimensions) were between -3 and +3 range, which is accepted for normal distribution.

Table 4.15

Skewness and Kurtosis Test of Teachers' conceptions of assessment in four dimensions

	Skewness		*Kurtosis*	
	Statistic	*Std. Error*	*Statistic*	*Std. Error*
Student Accountability	*.06*	*.13*	*-.51*	*.25*
School Accountability	*-.12*	*.13*	*-.58*	*.25*
Improvement	*-.21*	*.13*	*-.60*	*.25*
Irrelevance	*.24*	*.13*	*-.58*	*.25*

As displayed in Figure 4.1, the histogram of mathematics teachers' conceptions of assessment in terms of student accountability conception shows that the data exhibit normal distribution since the distribution follows a bell-shaped density curve described by means and standard deviations.

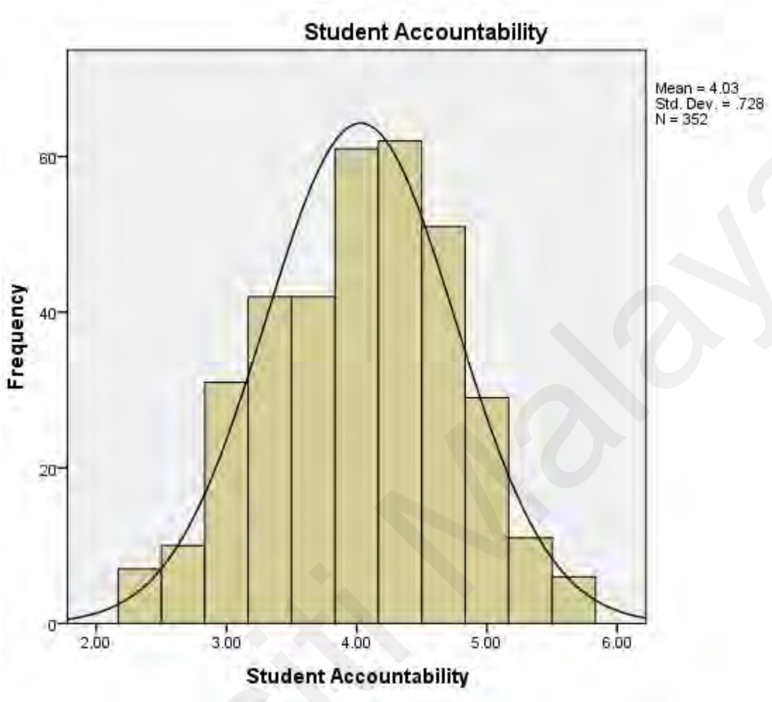

Figure 4.1. Histogram of Student Accountability Conception

Figure 4.2 shows that the circles all lie close to the straight line; thus, it can be concluded that the data were normally distributed without outliers.

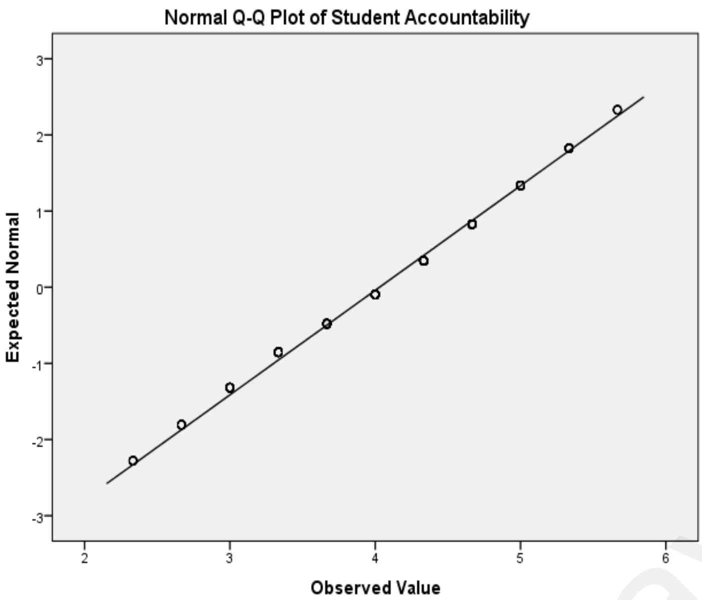

Figure 4.2. Normal Q-Q Plot of Student Accountability Conception

Figure 4.3 shows that the box-plot for Student Accountabilty Conception is perfectly symmetric; therefore, it can be concluded that the data is normally distributed.

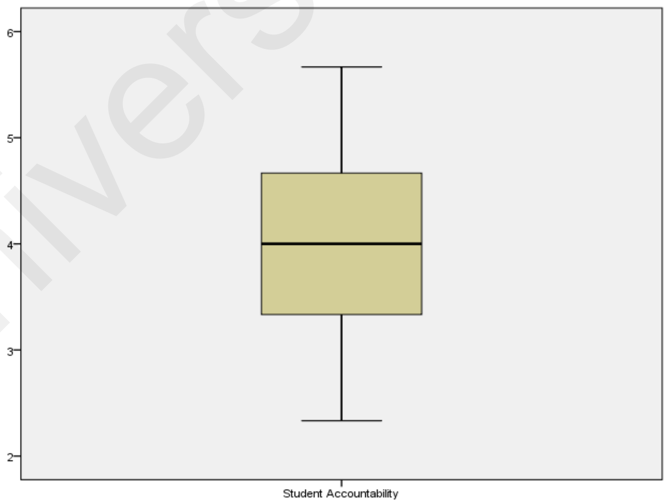

Figure 4.3. Box-plot of Student Accountability Conception

As displayed in Figure 4.4, the histogram of mathematics teachers' conceptions of assessment in terms of school accountability conception shows that the data exhibit normal distribution since the distribution follows a bell-shaped density curve described by means and standard deviations.

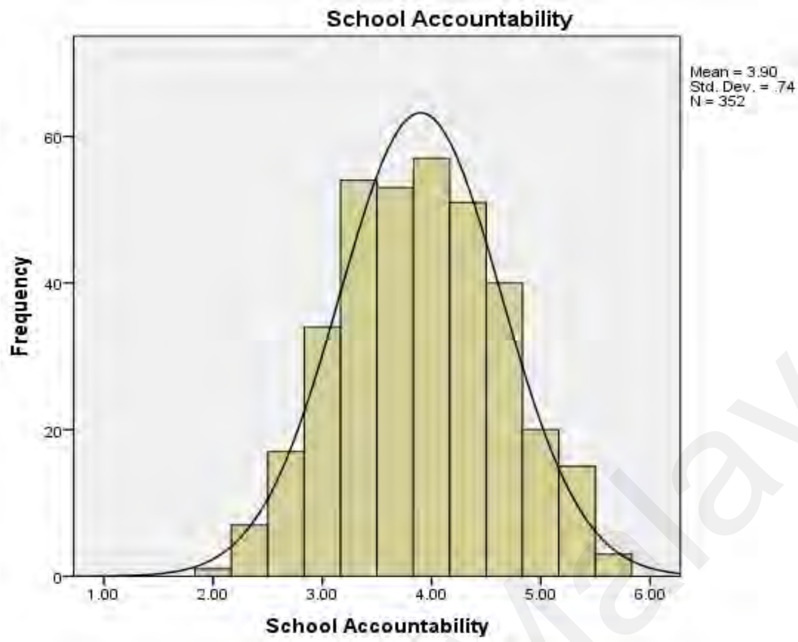

Figure 4.4. Histogram of School Accountability Conception

Figure 4.5 shows that the circles all lie close to the straight line; thus, it can be concluded that the data were normally distributed without outliers.

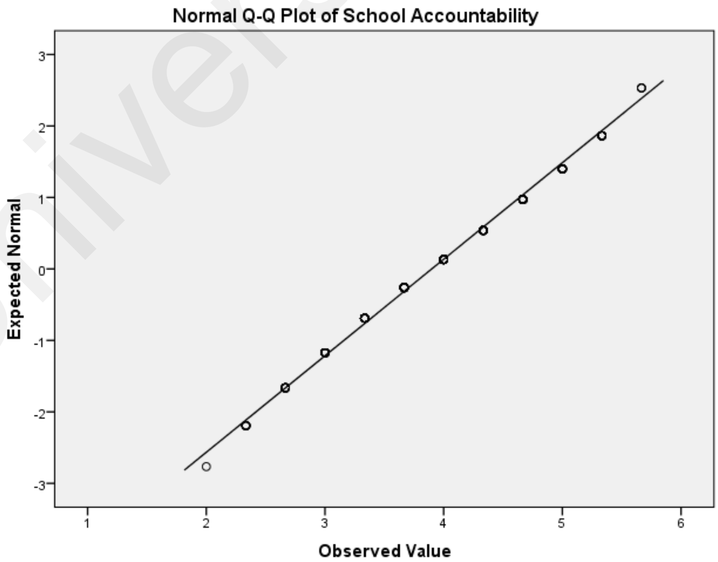

Figure 4.5. Normal Q-Q Plot of School Accountability Conception

Figure 4.6 shows the box-plot for School Accountabilty Conception. Although the box-plot is not perfectly symmetric, there is no clear violation of normality. As a result, it can be concluded that the data is normally distributed.

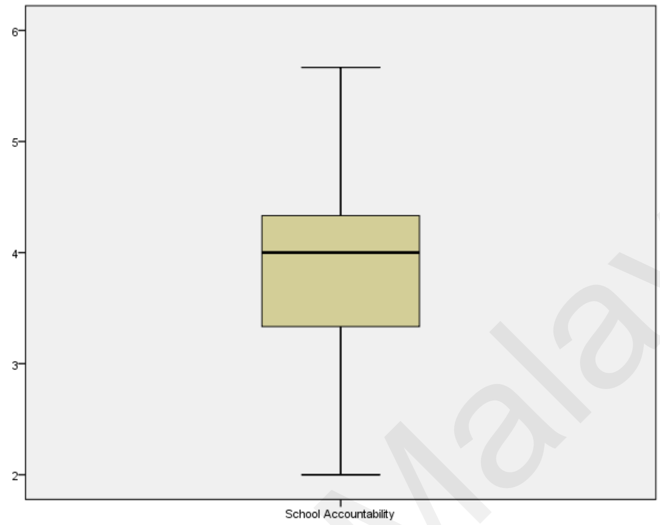

Figure 4.6. Box-plot of School Accountability Conception

As displayed in Figure 4.7, the histogram of mathematics teachers' conceptions of assessment in terms of irrelevance conception shows that the data exhibit normal distribution since the distribution follows a bell-shaped density curve described by means and standard deviations.

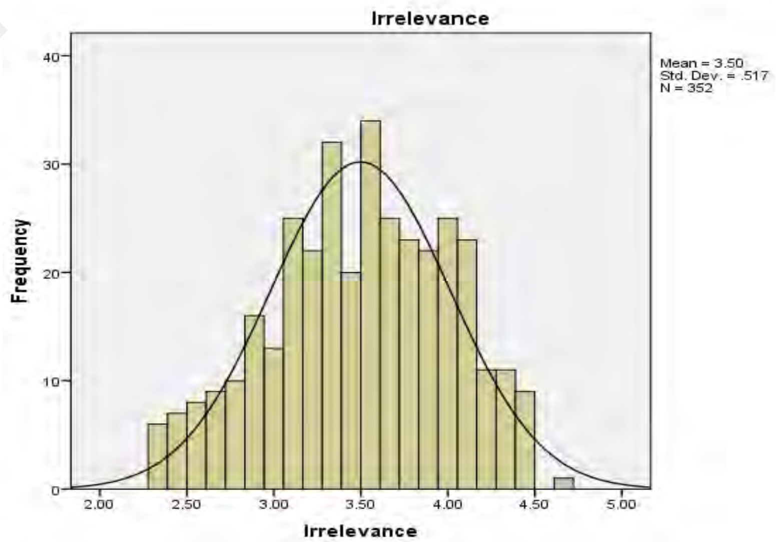

Figure 4.7. Histogram of Irrelevance Conception

Figure 4.8 shows that the circles all lie close to the straight line; thus, it can be concluded that the data were normally distributed without outliers.

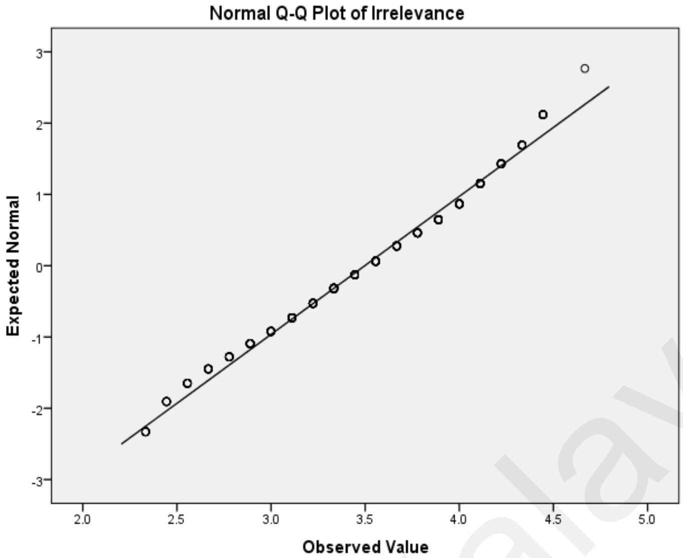

Figure 4.8. Normal Q-Q Plot of Irrelevance Conception

Figure 4.9 shows the box-plot for Irrelevance Conception. Although the box-plot is not perfectly symmetric, there is no clear violation of normality. As a result, it can be concluded that the data is normally distributed.

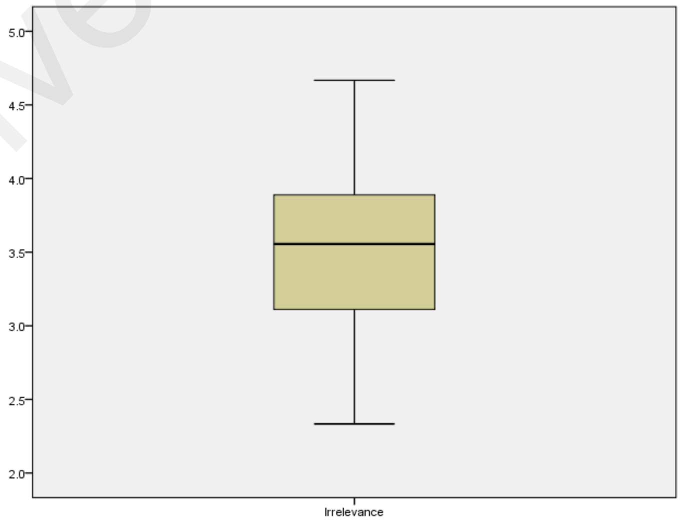

Figure 4.9. Box-plot of Irrelevance Conception

As displayed in Figure 4.10, the histogram of mathematics teachers' conceptions of assessment in terms of improvement conception shows that the data exhibit a right-skewed distribution. However, it is still considered as a normal distribution.

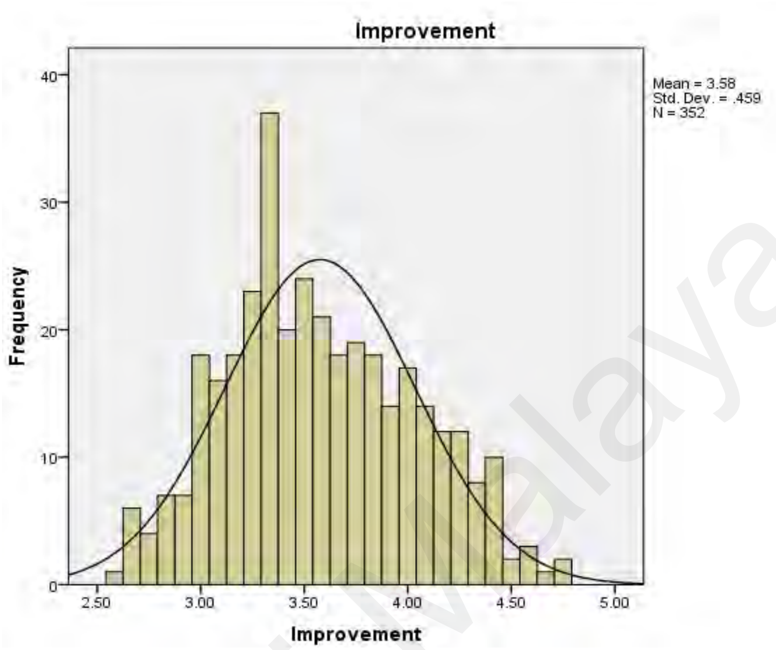

Figure 4.10. Histogram of Improvement Conception

Figure 4.11 shows that the circles all lie close to the straight line; thus, it can be concluded that the data were normally distributed without outliers.

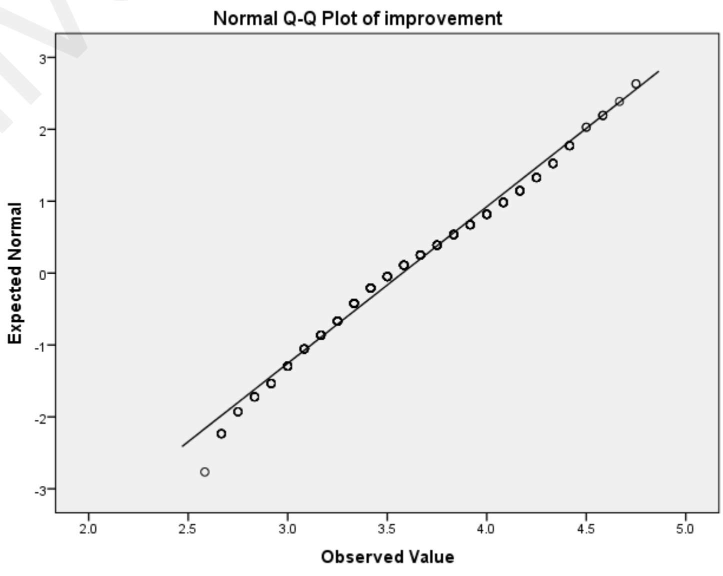

Figure 4.11. Normal Q-Q Plot of Improvement Conception

Figure 4.12 shows the box-plot for improvement Conception. Although the box-plot is not perfectly symmetric, there is no clear violation of normality. As a result, it can be concluded that the data is normally distributed.

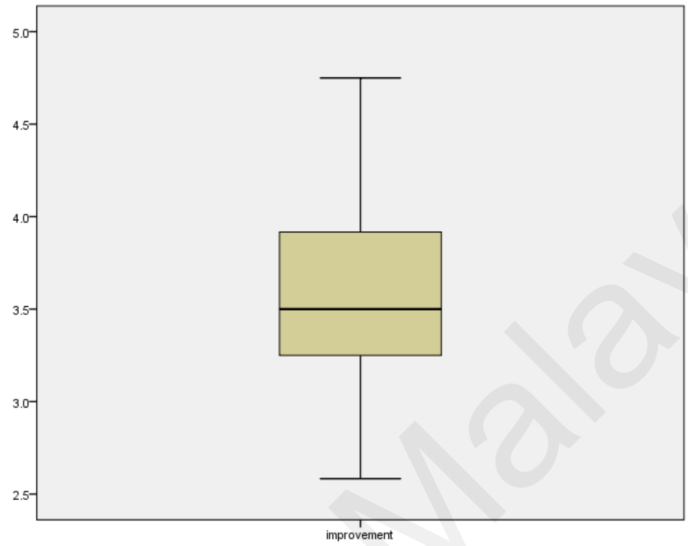

Figure 4.12. Box-plot of Improvement Conception

4.12 Results of Analysis for Research Question 2

RQ2. Do mathematics teachers' conceptions of assessment have significant difference in terms of years of experience?

To answer the research question two, multivariate analysis of variance (MANOVA) was conducted to determine whether there were any significant differences among different conceptions of assessment with respect to teachers' years of experience. Prior to that, the tests of assumptions were conducted. The results of testing assumptions in Figure 4.13 indicated that there were no univariate or multivariate outliers in terms of years of experience in four categories of conceptions (School Accountability, Student Accountability, Irrelevance, and Improvement) as assessed by boxplot.

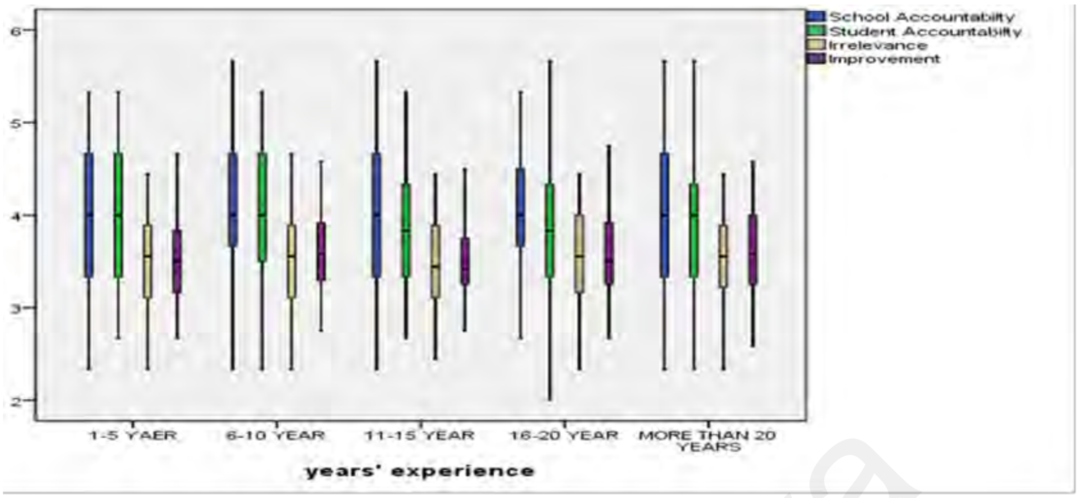

Figure 4.13. Assumption of No Univariate Outliers

Although Table 4.16 shows that Kolmogorov-Smirnov values for years of experience in some categories of conceptions of assessment is less than 0.05, meaning that it is assumed not normally distributed, Table 4.17 shows that skewness and kurtosis values for years of experience in each categories of conceptions of assessment were between -3 and +3, which are acceptable, so the assumption of normality has been met.

Table 4.16.

Kolmogorov-Smirnov[a] Test of Normality for Teachers' Conceptions in terms of years of experience

	1-5 Y EAR			6-10 YEAR			11-15 YEAR			16-20 YEAR			MORE THAN 20 YEARS		
	Statistic	df	Sig.	Statistic	df	Sig.	Statistic	df	Sig.	Statistic	df	Sig.	Statistic	df	Sig.
SCACC	.142	34	.078	.127	51	.038	.124	66	.013	.111	84	.013	.124	117	.000
STACC	.167	34	.025	.102	51	.200*	.134	66	.005	.116	84	.007	.123	117	.000
IMP	.101	34	.200*	.108	51	.196	.117	66	.024	.26	84	.002	.080	117	.065
IRR	.096	34	.200*	.089	51	.200*	.117	66	.025	.119	84	.005	.073	117	.175

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Table 4.17.

Assumption of Multivariate Normality

	1-5 YEAR		6-10 YEAR		11-15 YEAR		16-20 YEAR		MORE THAN 20 YEARS	
	Skewness	Kurtosis	Skewness	Kurtosis	Skewness	Kurtosis	Skewness	Kurtosis	Skewness	Kurtosis
SCACC	.28	-.84	-.42	.48	1.00	-.30	.12	-.35	.07	-.60
STACC	-.17	-.94	.05	-.33	-.28	-.48	-.10	-.58	-.09	-50
IMP	.30	-.45	.22	-.42	.47	-.38	.41	-.55	-.04	-.71
IRR	-.20	-.72	-.05	-.36	.217	-.63	-.51	.56	-.25	-.57

Moreover, the homogeneity of variance-covariance matrices was calculated by Box's M test (Box's M=39.67, F=.96, p=.53) (see Table 4.18). The result showed that the variances were approximately equal and there was homogeneity of variance of dependent variables among the groups.

Table 4.18.

Box's Test of Equality of Covariance Matrices

Box's M	39.67
F	.96
df1	40
df2	97700.73
P	.53

Also, Table 4.19 shows that the assumption of homogeneity of variance- covariance was met, as assessed by Levene's test of homogeneity of variance- covariance for School Accountability (F=1.19, p=.31), Student Accountability (F=1.171, p=.32), Improvement (F=1.147, p=.33) and Irrelevance factor (F=.373, p=.82).

Table 4.19.

Assumption of Homogeneity of Variance

	F	*df1*	*df2*	*p*
School Accountability	1.19	4	347	.31
Student Accountability	1.17	4	347	.32
Improvement	1.14	4	347	.33
Irrelevance	.37	4	347	.82

As the results of the preliminary analyses showed none of the assumptions of normality and homoscedasticity were violated. Thus, One-way MANOVA test was performed on the four dependent variables (improvement, student accountability, school accountability and irrelevance conceptions) in terms of years of experience. The results showed no significant differences in the effect of years of experience on dependent variable in four dimensions: F (16, 1051.57) =.39, Wilks Lambda =.98, p =.98 (*p>.05)*.

Table 4.20.

Descriptive Statistics for Teachers' conceptions of assessment in terms of years of experience

Dependent Variable	years' experience	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
School Accountability	1-5 YAER	3.98	.12	3.73	4.23
	6-10 YEAR	4.00	.10	3.79	4.20
	11-15 YEAR	3.80	.09	3.62	3.98
	16-20 YEAR	3.88	.08	3.73	4.04
	MORE THAN 20 YEARS	3.89	.06	3.75	4.02
Student Accountability	1-5 YAER	4.00	.12	3.75	4.24
	6-10 YEAR	4.05	.10	3.85	4.25
	11-15 YEAR	3.97	.09	3.79	4.15
	16-20 YEAR	4.06	.08	3.91	4.22
	MORE THAN 20 YEARS	4.03	.06	3.89	4.16
Improvement	1-5 YAER	3.53	.07	3.38	3.69
	6-10 YEAR	3.59	.06	3.46	3.72
	11-15 YEAR	3.49	.05	3.38	3.60
	16-20 YEAR	3.60	.05	3.50	3.70
	MORE THAN 20 YEARS	3.60	.04	3.52	3.69
Irrelevance	1-5 YAER	3.45	.08	3.28	3.63
	6-10 YEAR	3.50	.07	3.36	3.64
	11-15 YEAR	3.49	.06	3.36	3.62
	16-20 YEAR	3.52	.05	3.41	3.63
	MORE THAN 20 YEARS	3.48	.04	3.39	3.58

Table 4.21.

Results of a one-way MANOVA on difference in the mean scores for conceptions of assessment factors in terms of years of experience

Dependent Variable	Type III Sum of Squares	df	MS	F	*p*	η^2
School accountability	1.306	4	.32	.59	.66	.007
Student accountability	.375	4	.09	.17	.95	.002
Improvement	.657	4	.16	.77	.54	.009
Irrelevance	.126	4	.03	.11	.97	.001

MS= Mean Square
η^2=Partial Eta Squared

Table 4.21 provides a summary of results of the one-way MANOVA analysis based on years of experience. Mathematics teachers' conceptions of assessment showed similar conception in each dimension with respect to years of experience: School Accountability $F= (4,347) = .59$, $p = .66$ ($p > .05$), Student Accountability $F= (4,347) = .17$, $p = .95$ ($p > .05$), Improvement $F= (4,347) = .77$, $p = .54$ ($p > .05$), and Irrelevance $F= (4,347) = .11$, $p = .97$ ($p > .05$).

Partial eta-squared is a measure of variance, and it indicates what proportion of the variance in the dependent variable is attributable to independent variable(Pallant, 2007). The value of partial eta-squared as shown in Table 4.21 for each dimension of teachers' conceptions of assessment is as follows: (School Accountability=.007, Student Accountability=.002, Improvement= .009, Irrelevance= .001). These values show that all factors have low effect size. According to Cohen's (1988) guidelines, the level of the effect size less than 0.3 shows low effect, between 0.3 and 0.8 shows moderate effect size, and more than 0.8 shows high level of effect size.

Along with the result of this analysis, the researcher fails to reject each null hypothesis and reported that years of experience is not a factor affecting the teachers' conceptions of assessment in four categories among Iranian mathematics teachers.

4.13 Result of Analysis for Research Question 3

RQ3. Do mathematics teachers' conceptions of assessment have significant difference in terms of academic qualification?

To answer the research question three, multivariate analysis of variance (MANOVA) was conducted to determine whether there were any significant differences among different conceptions of assessment with respect to teachers' academic qualification. However, prior to the main analysis, the tests of assumptions were conducted. The results of testing assumptions in Figure 4.14 indicated that there were no univariate or multivariate outliers in terms of academic qualification in four categories of conceptions (School Accountability, Student Accountability, Irrelevance, and Improvement) as assessed by boxplot.

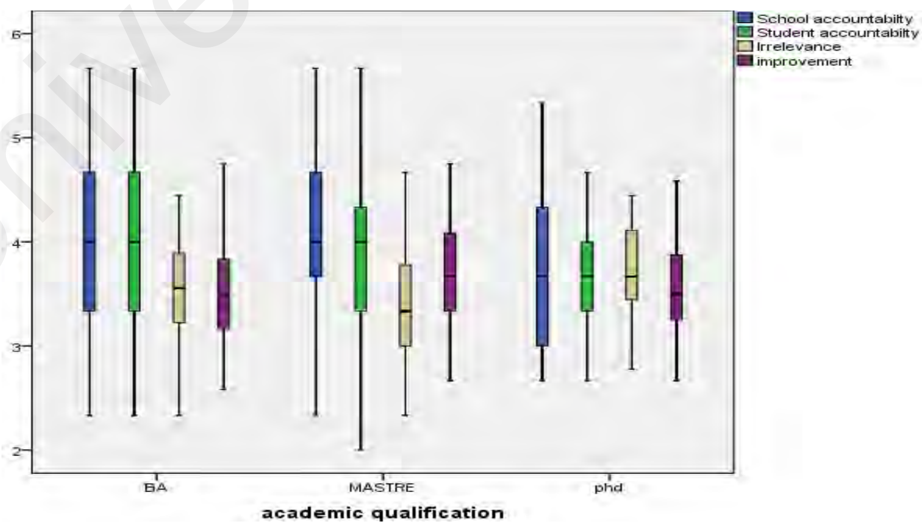

Figure 4.14: Assumption of No Univariate Outliers

Although Table 4.22 shows that Kolmogorov-Smirnov values for academic qualification in some categories of conceptions of assessment are less than 0.05, which is assumed to be not a normal distribution, Table 4.23 shows that skewness and kurtosis values for academic qualification in each category of conceptions of assessment were between -3 and +3, which are acceptable, so the assumption of normality has been met.

Table 4.22.

Kolmogorov-Smirnov Test of Normality for teachers' conceptions in terms of academic qualification

		Kolmogorov-Smirnov[a]		
	Academic Qualification	Statistic	df	Sig.
School Accountability	BA	.10	194	.000
	MASTER	.10	126	.001
	PhD	.16	32	.025
Student Accountability	BA	.12	194	.000
	MASTER	.11	126	.000
	PhD	.15	32	.052
Improvement	BA	.08	194	.001
	MASTER	.08	126	.018
	PhD	.10	32	.200*
Irrelevance	BA	.09	194	.000
	MASTER	.07	126	.060
	PhD	.14	32	.105

*. This is a lower bound of the true significance.
a. Lilliefors Significance Correction

Table 4.23.

Assumption of Multivariate Normality

	Bachelor 's degree		Master's degree		Doctor's degree	
	Skewness	Kurtosis	Skewness	Kurtosis	Skewness	Kurtosis
SCACC	.01	-.61	.09	-.44	-.19	-.62
STACC	-.29	-.40	-.02	-.34	.42	-.58
IMP	.31	-.34	.07	-.86	.29	-.21
IRR	-.33	-.49	.06	-.58	-.39	-.39

Also, the homogeneity of variance-covariance matrices was calculated by Box's M test (Box's *M=17.29, F=.83, p=.66*) (see Table 4.24). Based on this result, it can be concluded that the variances were approximately equal and there was homogeneity of variance of dependent variables among the groups.

Table 4.24.

Box's Test of Equality of Covariance Matrices

Box's M	17.29
F	.83
df1	20
df2	29182.31
P	.66

Moreover, Table 4.25 shows that the assumption of homogeneity of variance-covariance was met, as assessed by Levene's test of homogeneity of variance-covariance for School Accountability *(F=2.39, p=.09),* Student Accountability (*F*=.06, *p=.93),* Improvement (*F*=1.19, *p=.30),* and Irrelevance factor (*F*=.89, *p=.41).*

Table 4.25.

Assumption of Homogeneity of Variance

	F	*df1*	*df2*	*p*
School Accountability	2.39	2	349	.09
Student Accountability	.06	2	349	.93
Improvement	1.19	2	349	.30
Irrelevance	.89	2	349	.41

As the results of the preliminary analyses showed none of the assumptions of normality and homoscedasticity were violated. Therefore, the One-way MANOVA test was performed on the four dependent variables (improvement, student accountability, school accountability and irrelevance conceptions) with respect to academic qualification. The results showed significant differences in the effect of academic qualification on dependent variables in four dimensions: F (8,692) =3.65, Wilks Lambda=.92, $p<.05$; partial η2 = .04

Table 4.26.

Descriptive Statistics for Teachers' conceptions of assessment in terms of academic qualification

Dependent Variable	academic qualification	Mean	Std. Error	95% Confidence Interval Lower Bound	Upper Bound
School Accountability	BA	3.95	.05	3.84	4.05
	MASTER	3.87	.06	3.74	4.00
	PHD	3.69	.13	3.44	3.95
Student Accountability	BA	4.05	.05	3.95	4.15
	MASTER	4.06	.06	3.93	4.19
	PHD	3.74	.12	3.48	3.99
Improvement	BA	3.52	.03	3.45	3.58
	MASTER	3.67	.04	3.59	3.75
	PHD	3.52	.08	3.36	3.68
Irrelevance	BA	3.52	.03	3.45	3.59
	MASTER	3.39	.04	3.30	3.48
	PHD	3.73	.09	3.55	3.91

Table 4.27.

Results of a one-way MANOVA on difference in the mean scores for conception of assessment factors by academic qualification

Dependent Variable	Type III Sum of Squares	df	MS	F	*p*	η^2
School accountability	1.88	2	.94	1.72	.17	.010
Student accountability	2.96	2	1.48	2.83	.06	.016
Improvement	1.84	2	.92	4.47	.01	.025
Irrelevance	3.32	2	1.66	6.42	.00	.036

MS= Mean Square
η^2=Partial Eta Squared

Table 4.27 provides a summary of the results of the one-way MANOVA analysis based on academic qualification. Mathematics teachers' conceptions of assessment showed similar conception in two dimensions of conceptions with respect to academic qualification as follows: School Accountability F= (2,349) = 1.72, p =*.17 (p>.05),* Student Accountability $F= (2,349) = 1.48$, $p = .06$ ($p> .0125$). However, there were significant differences across academic qualification levels regarding Improvement, $F (2,349) = 4.47, p < .05$, partial η2 = .02 and Irrelevance, F $(2,349) = 6.42, p < .05$, partial η2 = .03.

The results of the Scheffe test show that only Ph.D and MA holders are significantly different in irrelevance conception, teachers with Ph.D. degree conceived that assessment is irrelevant and unfair more than teachers with Ma degree did. Besides, results show that only MA and BA holders are significantly different in improvement conception; teachers with MA degree conceived of assessment as improving teaching and learning mathematics more than BA holder teachers did.

Table 4.28.

Results of Post Hoc Scheffe test for Irrelevance and Improvement Conceptions

Dependent Variable	(I) academic qualification	(J) academic qualification	Mean Difference (I-J)	Std. Error	Sig
Irrelevance	BA	MA	.13	.05	.07
		PHD	-.20	.09	.10
	MA	BA	-.13	.05	.07
		PHD	-.34*	.10	.004
Improvement	BA	MA	.13	.05	.07
		MA	-.15*	.05	.01
	MA	PHD	-.004	.08	.99
		PHD	.14	.08	.26

Partial eta-squared is a measure of variance showing what proportion of the variance in the dependent variable is attributable to the independent variable(Pallant, 2007). The value of Partial eta-squared as shown in Table 4.27 for each dimension of teachers' conceptions of assessment were as follows: School Accountability=.01, Student Accountability=.01, Improvement= .02, Irrelevance= .03. Thus, all factors have low effect size.

According to the results, the researcher was able to reject the null hypothesis and reported that academic qualification is a factor affecting the teachers' conceptions of assessment in two categories which are Irrelevance and Improvement categories among Iranian mathematics teachers.

4.14 Result of Analysis for Research Question 4

RQ4. Do mathematics teachers' conceptions of assessment have significant difference in terms of grade taught?

To answer the research question four, multivariate analysis of variance (MANOVA) was conducted to determine whether there were any significant differences among different conceptions of assessment with respect to grade taught. However, prior to the main analysis, the tests of assumptions were conducted. The results of testing assumptions in Figure 4.15 indicated that there were no univariate or multivariate outliers in terms of grade taught in four categories of conceptions (School Accountability, Student Accountability, Irrelevance, and Improvement) as assessed by boxplot.

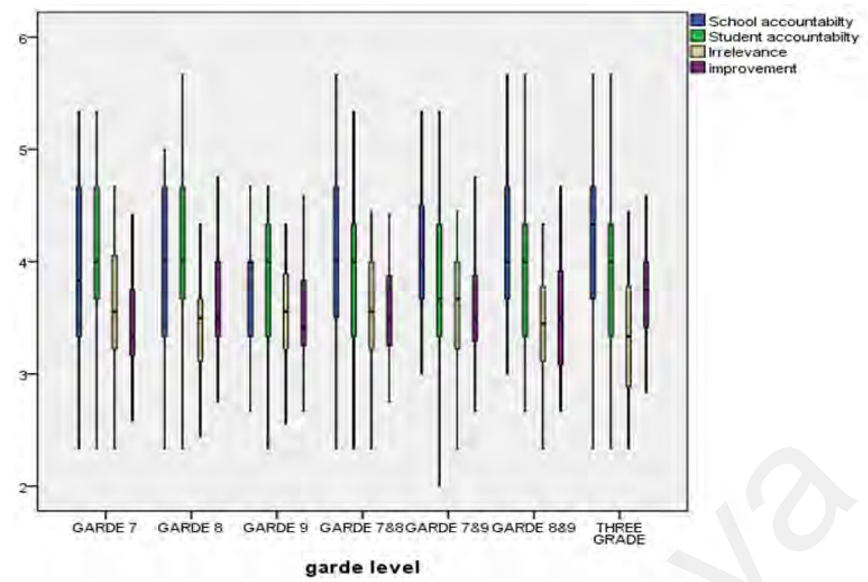

Figure 4.15. Assumption of No Univariate Outliers

Although Table 4.29 shows that Kolmogorov-Smirnov values for grade level taught in some categories of conceptions of assessment are less than 0.05, which is assumed not a normal distribution, Table 4.30 shows that skewness and kurtosis values for grade level taught in each category of conceptions of assessment were between -3 and +3, which are acceptable, so the assumption of normality has been met.

Table 4.29.

Kolmogorov-Smirnov Test of Normality for teachers' conceptions in terms of grade level taught

		Kolmogorov-Smirnov[a]		
	grade level	Statistic	Df	Sig.
School Accountability	GARDE 7	.11	98	.002
	GARDE 8	.11	153	.000
	GARDE 9	.12	101	.000
Student Accountability	GARDE 7	.13	98	.000
	GARDE 8	.11	153	.000
	GARDE 9	.13	101	.000
Improvement	GARDE 7	.08	98	.119
	GARDE 8	.11	153	.000
	GARDE 9	.10	101	.009
Irrelevance	GARDE 7	.10	98	.007
	GARDE 8	.06	153	.200*
	GARDE 9	.09	101	.036

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Table 4.30.

Assumption of Multivariate Normality

	SCACC		STACC		IMP		IRR	
	Skewness	Kurtosis	Skewness	Kurtosis	Skewness	Kurtosis	Skewness	Kurtosis
GARDE 7	.14	-4.75	-.09	.52	.30	-.32	-.40	-.23
GARDE 8	-.09	-.81	-.12	-.61	.12	-.83	-.03	-.67
GARDE 9	-.26	.23	-.06	-.25	-.35	-.30	-.30	-.60

The homogeneity of variance-covariance matrices was calculated by Box's M test (Box's *M*=22.48, *F*=1.10, *p*=.33) (see Table 4.31). Therefore, it can be concluded that the variance was approximately equal and there was homogeneity of variance of dependent variable among the groups.

Table 4.31.

Box's Test of Equality of Covariance Matrices

Box's M	22.48
F	1.10
df1	20
df2	335234.28
P	.33

Further, Table 4.31 shows that the assumption of homogeneity of variance-covariance was met, as assessed by Levene's test of homogeneity of variance-covariance (School Accountability (*F*=2.60, *p=.07),* Student Accountability (*F*=2.16, *p=.11),* Improvement (*F*=1.01, *p=.36)* and Irrelevance (*F*=.44, *p=.64)).*

Table 4.32.

Assumption of Homogeneity of Variance

	F	*df1*	*df2*	*p*
School Accountability	2.60	2	349	.07
Student Accountability	2.16	2	349	.11
Improvement	1.01	2	349	.36
Irrelevance	.44	2	349	.64

As the results of the preliminary analyses showed none of the assumptions of normality and homoscedasticity were violated. Thus, One-way MANOVA test was performed on the four dependent variables (improvement conception, student accountability, school accountability and irrelevance conception) with respect to grade taught. The results showed non-significant differences in the effect of grade taught on dependent variable in four dimensions: $F (8, 692) = .923$, Wilks Lambda = .979, $p = .49$ ($p > .05$).

Table 4.33.

Descriptive Statistics for Teachers' conceptions of assessment in terms of grade level taught

				95% Confidence Interval	
Dependent Variable	grade level	Mean	Std. Error	Lower Bound	Upper Bound
School Accountability	GARDE 7	3.93	.07	3.78	4.08
	GARDE 8	3.92	.06	3.81	4.04
	GARDE 9	3.82	.07	3.67	3.96
Student Accountability	GARDE 7	4.13	.07	3.98	4.27
	GARDE 8	3.99	.05	3.87	4.10
	GARDE 9	3.98	.07	3.84	4.12
Improvement	GARDE 7	3.56	.04	3.47	3.65
	GARDE 8	3.60	.03	3.53	3.67
	GARDE 9	3.54	.04	3.45	3.63
Irrelevance	GARDE 7	3.55	.05	3.45	3.66
	GARDE 8	3.45	.04	3.37	3.53
	GARDE 9	3.50	.05	3.40	3.60

Table 4.34.

Results of a one-way MANOVA on difference in the mean scores for conception of assessment factors by grade taught

Dependent Variable	Type III Sum of Squares	df	MS	F	*p*	η^2
School accountability	.86	2	.43	.78	.45	.004
Student accountability	1.45	2	.72	1.37	.25	.008
Improvement	.23	2	.11	.55	.57	.003
Irrelevance	.65	2	.32	1.22	.29	.007

MS= Mean Square
η^2=Partial Eta Squared

Table 4.34 provides a summary of results of the one-way MANOVA analysis based on grade taught. Mathematics teachers' conceptions of assessment showed similar conception in three dimensions with respect to grade taught. School Accountability F= (2,349) = .78, p = .45 (p> *.05),* Student Accountability F= (2,349) = 1.37, p = .25 (p> *.05),* Improvement F= (2,349) = .11, p = .57 (p> *.05)* and Irrelevance Factor F= (2,349) = 1.22, p = .29 (p> *.05).*

The values of partial eta-squared as shown in Table 4.34 for each dimension of teachers' conceptions of assessment are as follows: School Accountability=.004, Student Accountability=.008, Improvement= .003, Irrelevance= .007). which indicates that all factors have low effect size.

Based on the analysis results, the researcher fails to reject the null hypothesis and reported that grade taught is not a factor affecting the teachers' conceptions of assessment in four categories among Iranian mathematics teachers.

4.15 Results of Analysis for Research Question 5

RQ5. Do mathematics teachers' conceptions of assessment have significant difference in terms of gender?

To answer the research question five, multivariate analysis of variance (MANOVA) was conducted to determine whether there were any significant differences among different conceptions of assessment with respect to gender. However, first, the tests of assumptions were conducted. The results of testing assumptions in Figure 4.16 indicate that there were no univariate or multivariate outliers in terms of gender in four categories of conceptions (School Accountability, Student Accountability, Improvement, Irrelevance) as assessed by boxplot.

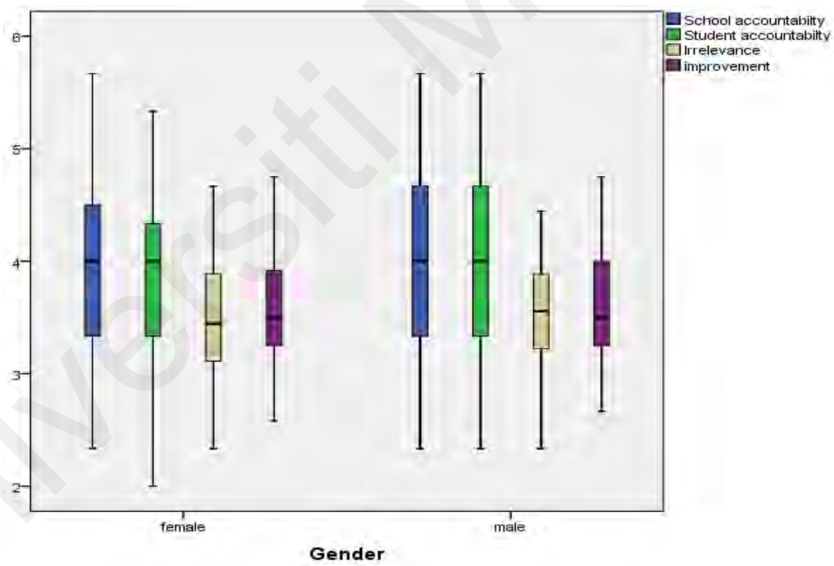

Figure 4.16. Assumption of No Univariate Outliers

Although Table 4.34 shows that Kolmogorov-Smirnov values for gender in some categories of conceptions of assessment are less than 0.05, which is assumed not a normal distribution, Table 4.35 shows that skewness and kurtosis values for gender in each category of conceptions of assessment were between -3 and +3, which are acceptable, so the assumption of normality has been met.

Table 4.35.

Kolmogorov-Smirnov test of normality for teachers' conceptions in terms of gender

		Kolmogorov-Smirnov[a]		
	Gender	Statistic	Df	Sig.
School Accountability	female	.09	198	.00
	male	.11	154	.00
Student Accountability	female	.11	198	.00
	male	.11	154	.00
Improvement	female	.09	198	.00
	male	.10	154	.00
Irrelevance	female	.08	198	.00
	male	.06	154	.200*

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Table 4.36.

Assumption of Multivariate Normality

	Female		Male	
	Skewness	Kurtosis	Skewness	Kurtosis
SCACC	-.03	-.50	.13	-.63
STACC	-.13	-.36	-.13	-.58
IMP	.21	.27	-.49	-.70
IRR	-.25	-.65	-.15	-.49

In addition, the homogeneity of variance-covariance matrices was calculated by Box's M test (Box's $M=11.58$, $F=1.14$, $p=.32$) (see Table 4.37). based on this result, it can be concluded that the variances were approximately equal and there was homogeneity of variance of dependent variables among the groups.

Table 4.37.

Box's Test of Equality of Covariance Matrices

Box's M	11.58
F	1.14
df1	10
df2	510716.80
P	.32

Also, Table 4.38 shows that the assumption of homogeneity of variance- covariance was met, as assessed by Levene's test of homogeneity of variance-covariance (School Accountability *(F=.32, p=.56),* Student Accountability (*F*=.61, *p=.43),* Improvement (*F*=.28, *p=.59),* and irrelevance factor (*F*=.43, *p=.51).*

Table 4.38.

Assumption of Homogeneity of Variance

	F	*df1*	*df2*	*p*
School Accountability	.32	1	350	.56
Student Accountability	.61	1	350	.43
Improvement	.28	1	350	.59
Irrelevance	.43	1	350	.51

As the results of the preliminary analyses showed none of the assumptions of normality and homoscedasticity were violated. So, One-way MANOVA test was performed on the four dependent variables (improvement conception, student accountability, school accountability and irrelevance conception) with respect to gender. The results showed no significant differences in the effect of gender on dependent variables in four dimensions: F (4, 347) =1.24, Wilks Lambda =.98, p = .29(*p>.05).*

Table 4.39.

Descriptive statistics for teachers' conceptions of assessment in terms of gender

Dependent Variable	Gender	Mean	Std. Error	95% Confidence Interval Lower Bound	Upper Bound
School Accountability	female	3.83	.05	3.72	3.93
	male	3.98	.05	3.87	4.10
Student Accountability	female	3.98	.05	3.88	4.08
	male	4.08	.05	3.97	4.20
Improvement	female	3.55	.03	3.49	3.62
	male	3.60	.03	3.52	3.67
Irrelevance	female	3.50	.03	3.43	3.58
	male	3.48	.04	3.40	3.56

Table 4.40.

Results of a one-way MANOVA on difference in the mean scores for conception of assessment factors by Gender

Dependent Variable	Type III Sum of Squares	df	MS	F	*p*	η^2
School accountability	2.196	1	2.19	4.03	.04	.011
Student accountability	.897	1	.89	1.69	.19	.005
Improvement	.171	1	.17	.81	.36	.002
Irrelevance	.060	1	.06	.22	.63	.001

MS= Mean Square
η^2=Partial Eta Squared

Table 4.40 provides a summary of results of the one-way MANOVA analysis based on gender. Mathematics teachers' conceptions of assessment showed similar conception in three dimensions with respect to gender. Student Accountability F= (1,350) = 1.69, p = .19 ($p > .05$), Improvement F= (1,350) = .17, p = .36 ($p > .05$), Irrelevance factor F= (1,350) = .22, p = .63 ($p > .05$). However, the results showed that there were significant differences across the gender groups in School Accountability $F= (1,350) = 2.16$, $p = .04$ ($p< .05$).

The comparison of the mean scores for School Accountability conception between male and female respondents showed that the male mathematics teachers (Mean=4.08) conceived of assessment as a process of School accountability more than female mathematics teachers did (Mean=3.98) (see Table 4.38).

Partial eta-squared is a measure of variance that indicates what proportion of the variance in the dependent variable is attributable to independent variable(Pallant, 2007). The value of Partial eta-squared as shown in Table 4.27 for each dimension of teachers' conceptions of assessment as follows: School Accountability=.011, Student Accountability=.005, Improvement= .002, Irrelevance= .001., suggesting that all factors have low effect size. According to Cohen s' (1988)guidelines, the level of the effect size less than 0.3 shows low effect, between 0.3 and 0.8 shows moderate effect size and more than 0.8 shows high level of effect size.

According to the results of this analysis, the researcher fails to reject the null hypothesis and reports that overall gender in not a factor affecting the teachers' conceptions of assessment among Iranian mathematics teachers. However, gender is a factor for the category of school accountability.

4.16 Summary

This chapter provides the findings of research on mathematics teachers' conceptions of assessment with respect to demographic factors including gender, years of experience, academic qualification and grade level taught among Iranian mathematics teachers in four educational districts in Mashhad. The data were analysed using descriptive statistics including frequency, percentage, mean and standards deviation and inferential statistics which is Multivariate Analysis of Variance (MANOVA) and the results were presented in tabular form as well.

The findings of the study show that mathematics teachers mostly conceived that assessment makes schools and students more accountable. Besides, it was found that years of teaching experience, gender of mathematics teachers as well as grade level taught by teachers are not factors influencing conceptions of assessment among Iranian mathematics teachers. However, findings of the study regarding gender showed that male mathematics teachers conceived of assessment as a process of student accountability more than female teachers. Moreover, findings revealed that there were significant differences in assessment conceptions across the academic qualifications. However, it was shown that there were significant differences across the academic qualifications on only two dimensions of conceptions, i.e. irrelevance and improvement.

Discussion of findings, implications of the study and recommendations for future research and conclusion of the study will be presented in Chapter 5.

Chapter 5

Discussion

5.1 Introduction

This chapter consists of seven sections, namely overview of the study, summary of the findings, discussion of findings, implications of the study, recommendations for future research, and contribution and conclusion of the study. In the first section, the research problem, research purpose, research questions and the methodology are elaborated on and in the second section, the summary of the findings is presented. In the third section, the discussion of the research findings is provided. A description of the implications of the study are contained in the fourth section and, in the fifth section, the recommendations for future studies are described. In the sixth section, contribution and finally in the seventh section concluding remarks for the study are described.

5.2 Overview of the Study

Reforms in the education of mathematics in Iran have recently been under the influence of theories such as constructivism, that have led to deconceptualization of learning and teaching and also assessment as the bridge between these two processes. The disparity between teachers' assessment conceptions and ideas characterizing reforms has often been considered as a big hindrance in the way of implementing educational reforms and new methods of assessment.

The purpose of this study was to investigate in-service Iranian mathematics teachers' conceptions of assessment at three grades of lower secondary school (7th, 8th, 9th grades) in seven educational districts in Mashhad, Iran. In addition, an attempt was made to determine whether there are differences in teachers' conceptions of assessment with regard to three demographic characteristics of mathematics teachers including years of experience, educational qualification and gender. This study was conducted using the model of conceptions of assessment proposed by Brown (2002; 2004).

A quantitative design was utilized in this study. The target population of the study consists of male and female mathematics teachers, who were teaching in public or government lower secondary schools grade 7 through 9 in Razavi Khorasan Province, Mashhad, Iran during semester 1 of the 2017-2018 academic year. They were selected by cluster random sampling method.

The instrument used for this study is Conceptions of Assessment-III (COA-III) Abridged questionnaire adopted from Brown (2006). Therefore, data was collected via an electronic questionnaire that included two sections. The first section asks teachers to provide demographic information, their educational background, gender, teaching experience and grade level (7th, 8th, 9th) taught and the district of education where they teach. The second section which is COA consists of 12 items for conception of improvement, 3 items for student accountability, 3 items for school accountability, and 9 items for the conception of irrelevant. Items were assessed on a 6-point positively packed scale (1-strongly disagree, 2-mostly disagree, 3-slightly agree, 4-moderately agree, 5-mostly agree, and 6-strongly agree). "Positively packed" means that for each statement, there are two negative (mostly disagree and strongly disagree) and four positive (slightly, moderately, mostly, and strongly agree) options to choose from. In the end, the quantitative data was analyzed using SPSS version 24.

5.3 Summary of Findings

The following are the major findings of this study:

1- The results of the analysis of the means and standard deviations related to the level of agreement in mathematics teachers' conceptions of the purposes of assessment revealed that student accountability conception had the highest agreement levels followed by school accountability conception. The two conceptions were not widely different, and participants tended to answer "Moderate Agreement" level among other variables. On the other hand, the irrelevance conception that had the smallest mean and the improvement conception could be considered as being around "Slight Agreement" level among other variables. Therefore, mathematics teachers mostly conceived that assessment makes school and students more accountable.

2- The results of the one-way MANOVA analysis based on years of experience indicated that there were no significant differences between among the four types of assessment conceptions with respect to years of experience.

3- The results of the one-way MANOVA analysis based on academic qualification showed that there were significant differences between among the four types of assessment conceptions with regard to academic qualification. Results showed that Ph.D. and MA are significantly different in irrelevance conception, i.e. teachers with Ph.D. degree conceived that assessment is not fair and relevant more than teachers with MA degree. Besides, the results showed that only MA and BA are significantly different in improvement conception, so that teachers with MA degree conceived that assessment improves teaching and learning mathematics more than BA degree holders did.

4- The results of the one-way MANOVA analysis based on gender indicated that mathematics teachers' conceptions of assessment have similar conception in three dimensions with respect to gender and only there were significant differences across the gender groups in school accountability. The comparison of the mean scores for school accountability conception between male and female teachers showed that male mathematics teachers conceived assessment as a process of school accountability more than female mathematics teachers did.

5- The results of the one-way MANOVA analysis based on grade level taught indicated that there were no significant differences between among the four types of assessment conceptions with respect to grade level taught.

5.4 Discussion of the Findings

This section discusses the interpretation of research findings based on the research questions presented in Chapter One. There are five research questions altogether. The discussion of research findings is presented here.

5.4.1 Research Question 1: What are mathematics teachers' conceptions of assessment?

The results of the current study showed the highest score for student accountability conception of mathematics teachers, followed by school accountability as having the second highest value of agreement. This reveals that mathematics teachers who teach mathematics at lower secondary government schools in Mashhad conceived of the purpose of assessment as more of accountability (student and school) rather than improvement. Initially, this was expected given the culture of accountability under Iranian education system.

This finding might be due to the fact that nature and culture of Iranian education system is more grading-oriented rather than learning-oriented (Saad, Sardareh, & Ambarwati, 2013) particularly in mathematics which has long been considered a top priority in Iranian school curricula, from antiquity to today. The competitive nature of the education system encourages this idea and belief that a good student is one who obtains good scores because examination results reflect the quality and worth of the individual.

As summative assessment approach or traditional assessment approach has been mostly used in this education system for a long time; therefore, the final score becomes the ultimate goal in the education system. In fact, Iranian society considers assessment as marks or grades. It is difficult to assess students in secondary school because only marks are important, so if a student gets a good grade, everyone suppose that this student is eligible to enter university (Saad, Sardareh, & Ambarwati, 2013). This encourages students to make more effort to obtain higher grades at school as well as in nationwide examinations. Moreover, high stake assessments have a big impact on how students shape their future plans; teachers undertake the responsibility to prepare their students for such examinations seriously, hence helping their students to obtain higher scores than others.

Besides, the findings of this study might be in line with the report by the International Mathematics and Science Study (TIMSS), which. has showed a continuous decline from 1995-2011 in mathematics among the Iranian students specifically in lower secondary schools. In fact, Iran's performance in all periodic studies 1995-2011 was lower than the international average. Local researchers agree that one of the reasons considering the result of TIMSS is that teachers with the current practices only emphasize the final score in assessment and examination and this has caused damage to the students' creativity and has led to the lack of exposure to higher cognitive skills (Kakia & Almasi, 2008; Porahmadi, 2008).

In other words, in the traditional (accountability) system before the reform in 2011, teachers had limited practices in the teaching and learning process. Although the reform was launched in 2011 in Iranian education and emphasized learner-centered instead of teacher-centered education, unfortunately, the result of TIMSS 2015 after the reform also indicated the poor performance and lack of growth by Iranian students, whereas many other countries continued to improve. This suggests that most Iranian teachers still use assessment to assign grades to students rather than to provide feedback to the students and their families to improve their learning.

Next, the improvement conception had the third highest value which was lower than mean value of accountability conception. This convinced the researcher that although new mathematics curriculum changed the view about mathematics from behaviorism to constructivism in order to improve mathematics teaching and learning process and all mathematics textbooks were rewritten as a reflection of this change, still the nature of Iranian education system is based on behaviorism in teaching and learning mathematics and accountability of students and schools is a basic value that teachers and school authorities are concerned about. Besides, according to previous studies (Brown, 2002; Calveric, 2010; Moiinvaziri, 2015) although improvement conception has positive relationship with school and student accountability, as teachers tended to answer "slight agreement" level to this conception among other variables in this study, it appears they agreed more with accountability purpose rather than improvement.

The last but not least, the fourth mean score was for irrelevance conception which indicated that the participants of the study did not think assessment was irrelevant in teaching and learning environment. It is necessary to mention that according to previous research and Brown's model, irrelevance conception has a negative relationship with improvement conception of assessment. It means that if teachers view assessment as a means for Improvement, they are not likely to see it as Irrelevant (Brown, Hui, Yu, & Kennedy, 2011; Barnes, Fives, & Dacey, 2017; Brown, 2002).

The findings of this study are in contrast with the outcomes of Brown's studies (from 2002 to 2008) as the theoretical framework of the current study. According to Brown (2004), teachers viewed assessment as a means for improvement with improving school quality rather than giving grades to students. These differences may be because of distinctions in the system of education and the context of the original study in New Zealand (Brown's studies were conducted in New Zealand) and the present study in Iran. This is because in New Zealand, the national policy of the Ministry stresses school-based and voluntary assessment in order to foster achievement and improve teaching programmes' quality. There is no mandatory assessment regime, so all assessment practices are voluntary and low stakes (Brown, 2002; 2004).

The finding of a qualitative study in Malaysia by Arumugham, Abdullah and Mahnud (2017), among Malaysian secondary school teachers showed that the participants mostly intended to provide the assessments in order to evaluate students' scores and assessing the level of student learning was far less important than their academic achievements and scores. Therefore, the researchers suggested that Malaysian teachers may need more information and training to guide them to have better conceptions of an effective assessment. These findings have greatly supported the results of the current study among Iranian teachers. As Iran and Malaysia are both developing countries, similar results in these researches could be due to the similarity of the educational system and the dominant culture of the communities (Arumugham, Abdullah, & Mahmud, 2017).

The findings of the current study are also consistent with the results of a study of teachers' conceptions of assessment in the Chinese context (Brown & Gao, 2015). Researchers found that teachers conceived accountability purpose more than improvement. Indeed, in high-stakes environment like China and Iran where examination results determine placement into different levels of education and acceptance into high-quality institutions and moreover, teachers use summative assessment frequently to motivate students as well as to inform instruction in the classroom, it is expected that teachers endorse assessment for the purpose of accountability of schools and students.

On the other hand, the study by Earl and Timperley (2014) among two groups of secondary mathematics teachers in the United States also revealed that first group of mathematics teachers participating in the study followed the traditional conception of assessment (test-oriented assessment). However, the small size of the sample in this study reduced the possibility of generalizing the results achieved to the entire population of mathematics teachers in the United States (Earl & Timperley, 2014).

Similarly, the findings of another study by Ludwig (2013) in the United States are in line with the results of the current study. Ludwig investigated the teachers' conceptions of assessment and their confidence regarding the implementation of sound classroom assessment practices among 160 Kindergarten through 12th-grade teachers in three public school districts. Teachers mostly conceived student accountability as the purpose of assessment. Therefore, it seems that there are several other factors beside the educational system and cultural issues effecting teachers' conceptions of assessment (Ludwig, 2013).

As an example of a similar study with different findings, Irving et al. (2011) conducted a qualitative study investigating teachers' descriptions of assessment in New Zealand. The results showed that improving students' learning was described as the first conception of assessment, unlike the results of the current study. However, similar to the participants in the current study, the New Zealand teachers put the idea of “assessment is detrimental or irrelevant in the student process” in the last place (Irving, Harris, & Peterson, 2011).

In the United states, Calveric (2010) investigated primary teachers’ conceptions about assessment and found that teachers in the study conceived of assessment as a means to improve teaching and learning and accountability was the second conception agreed upon by the participants of the study; no additional analysis was provided making comparisons limited.

Further, in the study of mathematics teachers’ assessment practices in the context of reform in Canada by Suurtamm et al. (2010) the researchers found that mathematics teachers conceived assessment as significant a tool for improving mathematical thinking. In other words, current teachers’ assessment practices were aligned with the assessment reform aims; teachers used a variety of assessment forms to assess students’ mathematics thinking regarding real world problems that were able to support students’ learning. It means that the teachers did not view assessment as the main means of giving grades to students.

In summary, the comparison of the results of the present study with the findings of previous studies shows that several grounded factors such as the educational system and policies, common beliefs about the educational success, environment and culture may change the priorities and conceptions of assessment among teachers. Therefore, the differences between the findings of researches in other countries and the current study among the Iranian mathematics teachers can be due to the differences in the education system and the viewpoints of the Iranian society including educational authorities, teachers, parents and students.

5.4.2 Research Question 2: Do mathematics teachers' conceptions of assessment have significant difference with respect to years of experience?

Despite the importance of mathematics teachers' experiences in teaching mathematics, the results of this study revealed that years of teaching experience made a non-significant difference in four conceptions of assessment. In other words, teachers with different years of experience in teaching hold approximately similar conceptions regarding accountability purpose of assessment (Moiinvaziri, 2015).

As majority of the sample in this stduy have experiences between 11 and more than 20 years, most of mathematics teachers in this study were considered as highly experienced teachers. This shows that as the teachers' years of experience increases, they would have higher values for the conception of school and student accountability (Moiinvaziri, 2015). In other words, even though they have many years of teaching experience, most in-service teachers have a narrow view of assessment purposes.

This finding might be due to the fact that teacher conceptions come predominantly from their experiences as students. Mathematics teachers who had experienced teacher-centered testing methods in their own schooling tended to agree with the use of teacher-controlled testing. More explicitly, as there is a long history of behaviorism view in the mathematics and education system of Iran, mathematics teachers looked incapable or reluctant to entertain any conceptualization of assessment outside their past experience. Besides, mathematics teachers might have experienced limited forms of assessment in schools and are now experiencing such views in their early years of university studies, too.

Regarding the findings of Brown's study, teachers' years of experience have no significant impact on teachers' conceptions of assessment. Therefore, the findings of the current study regarding the association between the teachers' years of experience and their conceptions of assessment are supported by Brown's model (Brown, 2006b).

These findings correspond with the findings of a study by Barnes et al. (2017) among 172 teachers (early childhood to high school level) in the United States. The researchers found that years of experience had no significant relationship with the conceptions of assessments among the teachers participating in the study (Barnes, Fives, & Dacey, 2017).

Another study in the United States by Snyder (2107) assessed the relationship between teachers' conceptions, literacy and practices of assessment and their characteristics. The results of this study also found no significant difference between teachers' conceptions of assessments based on their years of experience (Snyder, 2017).

On the other hand, the findings of several studies indicated the significant impact of teachers' experiences on their conceptions of assessment. For example, based on the Brown's model, a mixed method study by Moiinvaziri (2105) investigated the conceptions of assessment and their associated factors among 417 Iranian university teachers. Results of the study showed that there was significant difference in the teachers' conceptions of assessments regarding their years of teaching experience (Moiinvaziri, 2015). It is notable that this study was conducted in a different location/city (Sirjan) compared with the location of the current study (Mashhad) and used a different study method (mixed method). Furthermore, the study mentioned above investigated the conceptions of assessment among university teachers. These differences may explain the discrepancy between the findings of the current study and those of the above-mentioned study.

As another example, a study by Keaikitse (2012) in Botswana among the primary and secondary school teachers investigated the participants' thoughts of classroom assessment. The findings of this study showed that the years of teaching experience was a significant factor regarding classroom assessment practices (Koloi-Keaikitse, 2012).

Involving 414 Turkish teachers, a study by Vardar (2010) investigated teachers' conceptions of assessment as well. The researcher found that years of teaching experience resulted in significant differences in teachers' assessment conceptions. The study showed that the more experienced teachers had the highest conception of Student Accountability compared to less experienced teachers. According to Vardar, the reason behind this finding was teachers' past experience, that might have been related to more product-oriented strategies of assessment to improve students' scores in examinations in Turkey. Moreover, results of the study showed that teachers with five or less years of experience showed the highest level of irrelevant conception. Vardar (2010) considered the result as surprising and said it was probably related to the low quality of university assessment courses (Vardar, 2010).

Moreover, the results of the current study are not consistent with a study by Barkatsas (2003), that found that the mathematics teachers' years of experience was an effective factor in shaping their beliefs about assesment. Based on the study findings, among the mathematics teachers participating in the study, those having 7 to 9 years of experience empahsized a problem-solving orientation more (regarding the mathematics assessment) compared to the teachers with other years of experience.

5.4.3 Research Question 3: Do mathematics teachers' conceptions of assessment have significant difference in terms of academic qualification?

The findings of this study showed that significant differences existed between mathematics teachers' conceptions of assessment based on academic qualification in two categories (i.e. irrelevance and improvement). According to the results of this study, the mathematics teachers with Ph.D. degree conceive that assessment is not fair and relevant more than mathematics teachers with MA or BA degrees did. Moreover, results showed that teachers with MA degree conceive that assessment improve teaching and learning mathematics more than BA degree or Ph.D holders did.

The outcomes of studies on the impact of teachers' academic qualification on their conceptions of assessment are not the same. Regarding the findings of Brown's (2002) study, teachers' academic qualification has no significant impact on teachers' conceptions of assessment. Therefore, unlike other personal variables in this study (gender, experience and grade taught), the findings of the current study regarding the effects of the teachers' academic qualification on their conceptions of assessment are not supported by Brown's model.

However, most of the previous studies in this area found that teachers' academic qualification is a significant factor regarding the conceptions of assessment among teachers. For example, the results of the current study regarding the impact of teachers' academic qualification on their conceptions of assessment are supported by the outcomes of a study by Pishghadam and Sahebjam (2012) on teachers' conceptions of assessment among the Iranian English teachers in private language institutes. The finding of this study showed that the participants with PhD degree (compared with the teachers with MA or BA degrees) mostly believed that the assessments are irrelevant.

In other words, the more educated participants mostly believed that the assessment is something bad or unfair (Pishghadam & Shayesteh, 2012).

Explaining these results, the researchers pointed to the significance of objective measurement in Iran's educational system. Therefore, many teachers with PhD degree believe that assessment may not be very accurate, and there is always a margin for errors. Another possible reason might be the poor quality of assessment courses in universities or absence of assessment courses in universities (Pishghadam & Shayesteh, 2012). It is notable that the results of the current study may be explained by both of these reasons.

Furthermore, the findings of the study on teachers' assessment conceptions and practices among 79 teachers by Calveric (2010)revealed that significant relationships existed between conceptions and practices related to the level of degree attainment or academic qualification of teachers. The findings showed that there were significant differences between teachers with bachelor's degree and those who had postgraduate certificates and Master's degree regarding student accountability assessment conception. Calveric (2010) reported that this finding was due to teachers' insufficient assessment literacy and professional development in assessment.

However, as it is mentioned above, results of some other studies have not shown any significant relationship between teachers' academic qualification and their conceptions of assessment. For example, using Browns' Model of conceptions, a study by Barnes et al.(2017) among 179 teachers (K-12) assessed the participants' conceptions of assessment from a person-centered perspective. The results showed no significant differences among the teachers in the study clusters regarding their academic qualification.

5.4.4 Research Question 4: Do mathematics teachers' conceptions of assessment have significant difference in terms of grade taught?

The results of this study showed that there were no significant differences between and mathematics teachers' conceptions of assessment with respect to grade level taught, which means that mathematics teachers conceived the purpose and nature of assessment in the same way in all grades (7,8, and9). In other words, mathematics teachers in all three grades conceived of the purpose of assessment as more for accountability purposes instead of using assessment as an improvement tool in mathematics teaching and learning process.

One explanation for this result could be that there might be limited offered or promoted professional development in terms of assessment methods with specific purposes for mathematics teachers across each grade level from the Ministry of Education in Mashhad. Although there is evidence that the ministry of education has offered compulsory courses of in-service training in Mashhad, that is beneficial for new solutions and problem-solving in classrooms (MOEM, 2011), based on the researcher's familiarity with this training course, it seems that such courses make teachers to feel tired and overwhelmed.

Besides, in these courses of in-service training, teachers just gain general knowledge in terms of new mathematics books in three grades, while teachers need to have professional knowledge in each grade level about the goals of mathematics teaching and learning as well as assessment. Moreover, the educators who teach these courses of in-service training are chosen from mathematics teachers; thus, they have limited knowledge about basic principles of each mathematics textbook. In addition, new mathematics textbooks in each grade level place little stress on methods of assessment that teachers need to apply in each lesson.

It was expected that the participants of this study would apply assessment as a tool for improving teaching and learning as there is no final standard examination at lower grades, 7 and 8 and mathematics teachers are completely responsible of evaluation during the academic year, but as results show mathematics teachers in grade 7 and 8 conceived assessment as more of accountability rather than improvement and this is due to the fact that they have poor conceptions about the purpose of the new mathematics books in each level as well as alternative assessment.

Regarding the findings of Brown's study, grade taught has no significant impact on teachers' conceptions of assessments. Therefore, the findings of the current study regarding the effects of the grade taught on their conceptions of assessments are supported by Brown's model.

As an example, to support the findings of the current study, a study by Benson (2014) reported that there was no significant difference in the views and conceptions of teachers regarding their grade level taught. The study was conducted to investigate the purposes of assessment among the school teachers in United States.

Investigating teachers' conceptions of types of assessment and the effects of grade level and state tested content area on the participants' conceptions, another study by Neibling (2014)was conducted among American secondary teachers. According to the results reported by the researcher, no differences were seen among participants in the middle and high schools teachers related to the conceptions of assessment (Neibling, 2014).

However, the findings of this study do not support the results of the study conducted by Barnes et al. (2017)which also examined 179 K-12 teachers' conceptions of the purposes of assessment from a person-centered perspective using Brown's model of conceptions. The findings of this study also showed that there were significant differences between clusters on grade level taught by teachers.

Overall, reviewing the results of the current study and comparing them with the outcomes of other studies in Iran and other countries show that many factors can be effective in determining and changing teachers' conceptions of assessments. According to the findings of different studies, social factors such as the dominant culture and the viewpoint of community individuals, especially those who work in the field of education as well as the educational system and policies seem to be more effective than individual factors. However, more studies are needed to make a definitive comment on this matter.

5.4.5 Research Question 5: Do mathematics teachers' conceptions of assessment have significant difference in terms of their gender?

The results of the present study showed that significant differences did not exist in mathematics teachers' conceptions of assessment regarding their gender, which means that female and male mathematics teachers conceived of the purpose and nature of assessment similarly. However, results showed that there were significant differences across gender groups in terms of school accountability conception. The comparison of mean scores showed that male mathematics teachers conceived of assessment as a process of school accountability more than female mathematics teachers did.

The main reason for not underlining gender as a facet of variables influencing mathematics teachers' conceptions any longer might be the fact that all male and female teachers are working under the same education policy and education culture. Even after reform movement and the introduction of new mathematics curricula the system of education in Iran is still a teacher-centered system in which teachers have the most authority as well as the most responsibility (Moiinvaziri, 2015). Classes are usually held in the form of teachers having lectures and students are mostly passive. So, teachers' gender would not be a concern as all teachers including female and male follow the same goals, accountability of schools and students. In other words, all teachers (female and male) adapt to local guidelines developed for accountability standards in order to learn assessment skills that are necessary to assist them while preparing students for academic success on national assessments.

According to Brown's studies, gender is not a determinant factor in teachers' conceptions of assessment. Therefore, the findings of the current study (regarding the gender difference) is supported by the Brown's model of conceptions of assessment. However, reviewing the studies related to the relationship between gender and teachers' conceptions of assessment have shown that these studies have reached dissimilar findings. Several studies reported no effects of gender on conceptions of assessment among the teachers, while some others supported the significant impact of gender on the teachers' conceptions of assessment.

The findings of the present study also support those of a study by Barnes et al. (2017) in the United State which investigated the teachers' conceptions of the purposes of assessment. The findings of their study showed no significant difference between male and female participants in terms of their conceptions of assessment (Barnes et al., 2017). In another example, involving 147 Iranian teachers in Sirjan universities, a study by Moiinvaziri (2015) reported no significant relationship between conceptions of assessment and gender among the participants (Moiinvaziri, 2015).

Moreover, studying the differences between teachers' conceptions of assessment in terms of the demographic characteristics, Boston (2014) conducted a study among middle school teachers in the United States. The findings indicated that gender of the respondents did not make any significant difference in their conceptions of assessment. According to the researcher, one of the possible reasons for this finding could be the necessity of following the federal, state, or local accountability guidelines by teachers (regardless of their gender) for student performance.

On the other hand, as it was mentioned above, some studies reported that teachers' conceptions of assessment was influenced by their gender. As an example, the findings of a study by Ndalichako and Joyce (2015) in Tanzanian secondary schools showed that there was a significant difference in the conceptions of assessment between male and female teachers participating in the study. The results showed that the females had more desirable conceptions of assessment than males (Ndalichako, 2015).

5.5 Implications of the Study

Based on the data, the findings of this study have implications for policy makers, teachers, researchers and schools.

5.5.1 Policy Makers

The findings of this study provide essential information for the policy makers. The participants of this study conceived of the purpose and nature of assessment as more of accountability rather than improvement of students' learning. Therefore, this finding raises issues that require policy makers' attention, in that teachers persist in using traditional assessment practices, even though they are cognizant of the policy vision and objectives of the reform. Besides, tensions and contradictions within the policy areas of accountability, teaching, and evaluation need reviewing and resolving. Otherwise, goals and objectives that conflict and compete with the policy intent will result in further divergence from the policy intent. There are several key aspects that policy makers should pay attention to, as follows:

It is important to change the policy in secondary schools from quantitative assessing to qualitative assessing similar to what has been recently done in primary schools in Iran. Policy makers should make an effort to reduce the impact of traditional assessments (summative assessments) on teachers within the existing educational system and encourage teachers to adopt more alternative assessments. More explicitly, instead of a once-off pen-and-paper examination, there could be a school-based component that allows teachers to design alternative tasks that encourage processing skills rather than just "examination learning" (Hogan, 2014). Including a school-based component is an effort to push for a more balanced approach to a highly pressurized education system (Berry, 2011). In this way, policy makers will be able to reduce emphasis on summative assessment, encourage the use of quality feedback in formative assessment, and involve students as active partners in the assessment process (Berry, 2011). In fact, the Iranian educational reforms, like Singapore's and Finland's, should focus on reducing examination pressure (see Chapter 2).

It is vital for policy makers and the ministry of education to ensure that schools do not push teachers to enact teaching and assessment practices that do not reflect the reform intent of preparing students for the real life. The absence of improvement conception in assessment as a first level of agreement in this study suggests that numerous teachers' assessment practices reflect the policy only to the letter. Such enactment is a "cultural morphing" which is a way teachers and stakeholders incorporate reforms within the existing cultural norms (Goh & Wong, 2013). In such instances, it will be necessary for superintendents or policy makers to help the leadership teams of these schools understand the philosophies underlying the policy, to comprehend the reform goals, and to review ways to align their school-based processes and procedures with the reform vision.

More importantly, policy makers and school leaders could provide more open communication channels and revision policies based on consultation with and feedback from teachers. In fact, if policy makers want to bring about changed practices, they will have to work with teacher educators to help teachers comprehend and interpret their beliefs and values prior to and during a reform effort. Because adopting constructivist approaches means allowing a larger student presence in the classroom, this change in classroom dynamics may be uncomfortable for teachers for whom in teaching, "authority is hierarchy," and "classroom talk is teacher dominated" (Hogan, 2014). Teaching practices in Iran are embedded in the deep cultural traditions of the Asian region as a whole. Therefore, to ensure more alignment to the spirit of the reform, teacher educators and policy makers have to help teachers reconcile western-style approaches with traditional Asian views of teaching and learning (Goh & Wong, 2013).

Additionally, professional development needs to be provided for teachers in order for them to enact new curricular and assessment practices well. Brown (2004) declared that implementing new assessment policies, tools, or practices, either at local or national level, should take into consideration the complex structure of teachers' assessment conceptions to guarantee success. Besides, Brown (2004) mentioned that in order to effectively move teachers towards desired conceptions, teachers' pre-existing conceptions should be considered in their professional pre-service preparation and in-service development related to assessment. Pajares (1992) believed that teachers' assessment conceptions stem from their own experience as students, so if their experiences are mainly related to accountability and irrelevance, some teacher trainees and practicing teachers would hold conceptions that need development.

Moreover, while lectures and seminars can be provided to scale up and deepen teachers' assessment knowledge, a more critical concern pertains to the sustainability of the learning. Recent studies report that the quality of teacher assessment increased significantly when teachers participated in ongoing and continual professional learning in assessment literacy (e.g., Koh, 2011). For instance, over a two-year period, Koh (2011) provided a series of professional learning sessions including guiding teachers to develop authentic assessment tasks and rubrics and having them engage in peer critique and discussion. This kind of professional development model embodies the features of effective professional development and is more powerful than short-term, once-off sessions (K. H. Koh, 2011). Drawing on this model of sustained professional development, researchers and teacher educators must collaborate closely with policy makers to review current models of professional development, and to provide more localized and continuous professional learning in the area of assessment practices.

5.5.2 Teachers and Researchers

The findings of this study also have implications for teachers as teachers' conceptions play an important part in mediating how educational reforms are implemented in schools and classrooms. Teachers should know that the way they conceive of assessments shapes the way they make use of the results of assessments and frames selecting and developing assessment tools in their classrooms (Brown, 2008).

Based on the findings of this study, it is essential that teachers change their conceptions in order to help improve student achievement in mathematics. They have to use more alternative assessments that would engage students in the process of learning. It would be useful for researchers to work toward changing the nature of large-scale assessments to enable the assessment of a larger range of skills and cognitive domains. Besides, researchers could explore test formats that require students to demonstrate their learning in more ways than just through multiple-choice questions and short answer constructed-response tasks which have been deemed as requiring only limited responses. Teachers and researchers should help each other in order to change the nature of knowledge and the demands of society and also, they have to provide deeper insights into assessment practices in relation to the policy intent to enhance learning.

5.5.3 Schools

The findings of this study also have implications for principals at schools, given that they make decisions for schools' curricular programs and give approval to the types of support and resources teachers require in order to bring about student learning. In this study teachers showed that they agree with school accountability as a second conception of assessment. It means that assessment holds teachers, schools, and systems accountable for achieving societal goals and expectations. Besides, assessment results are used to demonstrate publicly that teachers and schools are doing a good job. Schools and teachers are rewarded (e.g., pay bonuses) or punished (e.g., dismissal) for exceeding or not reaching required standards. This conception of assessment by teachers provided evidence that school level policies have created a culture that is in misalignment with the reform objectives and vision.

There is therefore an urgent need for principals to be more mindful of the impact of school level policies on teachers' assessment practices, especially if they detract from the policy intent. For instance, to enhance student learning, principals need to move beyond being administrative consumers of assessment data. Rather, they need to be equipped with assessment literacy skills so that they are able to lead their staff in interpreting, evaluating and analyzing assessment data together (Stiggins, 2001).

To bring about fundamental change in teachers' practices, principals can provide assessment leadership in different areas. For instance, they should build the foundation for teachers' assessment literacy in their schools. In view of the reform vision, principals must establish a balanced assessment system in their schools, one that combines formative and summative assessments purposefully and meaningfully (Jakicic, 2009). A balanced assessment system in schools would enable teachers to find out which students are learning, and how to help those who are struggling. Such a system would guide principals in deciding which curriculum and instruction approaches need changing (Jakicic, 2009). Adopting balanced assessment systems in schools would enable principals to steer teachers away from the emphasis on marks and achievement and to focus their efforts on helping students learn.

Another aspect of building an assessment foundation is for principals to establish a strong culture of learning in their schools, thereby leading teachers away from the focus on achievement toward an emphasis on learning. A focus on learning means that principals work with their teachers to identify the intended learning goals and outcomes, and that they encourage their staff to adopt high expectations of students (Huff, 2009). Building a foundation for assessment literacy in the school would require principals to model assessment strategies (e.g., reflection, effective feedback), and to support the teachers who want to try new assessment practices to challenge their students.

5.6 Recommendations for Future Research

This study is conducted to investigate in-service Iranian mathematics teachers' conceptions of assessment at three grades of lower secondary at public schools (7th, 8th, 9th grades) in seven educational districts in Mashhad. Future studies might consider expanding the scope of the study in terms of setting, sample size, research design and data collecting method.

The first limitation of this study was that the sample was selected only from Mashhad city. Future studies could include more teachers from other areas of the country to have bigger samples and increase the generalizability of research findings. Moreover, apart from lower secondary school mathematics teachers teaching 7th, 8th, and 9th grades, mathematics teachers teaching in primary schools (grades 1 to 6), in higher secondary schools (grades 10-12), and in private schools should also be included in order see the main differences among these groups. Furthermore, the assessment conceptions of other stakeholders including students need to be examined as well to make connections between the findings in order to have a more comprehensive picture of the conceptions held by different users of assessments.

Second, this study only focused on mathematics teachers' conceptions in terms of nature and purposes of assessment. Future research might be done to identify mathematics teachers' conceptions of teaching, learning and mathematics curriculum and how they are related to the conceptions of assessment. As Thompson (1992) found teachers who view mathematics as instrumental have a different approach to assessment than those who approach mathematics as conceptual. Thus, the suggested study could focus on research questions such as, how are teachers' conceptions of mathematics teaching related to their conceptions of assessment? How are teachers' conceptions of mathematics curriculum related to their conceptions of assessment? and how are teachers' conceptions of learning mathematics related to their conceptions of assessment?

Third, this study used quantitative research design and questionnaire as the means to collect data. If the purpose of a future study is to identify how mathematics teachers design, analyze, interpret and report the results of assessment based on their conceptions about assessment and also to identify how mathematics teachers give feedback to their students, then, a case study involving interview or other qualitative methods may be used to carry out the study. Thus, the suggested future study may focus on research question such as the following: How do mathematics teachers design assessment tasks based on their conceptions about assessment? How do mathematics teachers analyze the results of the assessments based on their conceptions about assessment?

5.7 Contribution of the Study

This study is one of the first works in the area of mathematics teachers' conceptions and assessment in Iran. This study has contextualized the subject of assessment conceptions in mathematics and contributes to the macro literature that has been conducted mostly in the Global North (Adams & Hsu, 1998; Brown, 2002, 2004, 2008; Barkatsas, 2003). The evidence provided by this study shows that in the Iranian context, after reform movement in mathematics curricula, mathematics teachers still conceive of the purpose of assessment as more related to accountability rather than improvement, suggesting that the traditional assessment is dominant with little use of alternative assessment, which seems to be due the ease and standardized nature of traditional assessment as well as the traditional culture of education system in Iran.

This study reveals a severe contrast in teachers' working conditions, i.e. lack of professional development and support for teachers related to assessment in Iranian education system, weak teacher knowledge and skills in assessment as well as teachers still being burdened by culture of accountability, high-stakes external testing and assessment structure at different levels after reform movement opposed to what was found in the literature of Global North. This suggests that implementation of alternative assessment (new approaches to assessment) regarding reform movement in Iran may have different results from those of Global North.

5.8 Conclusion of the Study

This study has shown that mathematics curriculum reform alone does not guarantee changes in the nature of assessment without changing the conceptions of mathematics teachers regarding the purposes of assessment.

To implement a revised curriculum, the conceptions held by teachers should be matched with the ideas in the curriculum as previously mentioned that teachers play key role in driving education reform movement (chapter 1). Reform often requires teachers to abandon unwanted conceptions and practices. Besides, factors thought to be restricting teachers' assessment should be identified and eliminated. Thus, teachers' conceptions and assessment is an important research area, especially with regard to reforms in the mathematics curriculum.

In Iran, the successful implementation of the recent reform in the mathematics curriculum requires change in assessment and practices: this, in turn requires an understanding of teachers' conceptions and current practices of assessment. This study, therefore, explored and examined these conceptions.

The study revealed that majority of teachers had an accountability conception about the nature and purpose of assessment, that are not aligned with the curriculum. Indeed, in general their conceptions did not reflect improvement conception. It was showed that assessment was very effective on teachers' teaching practice and was associated with most of the other factors such as teacher accountability for students' results, and schools and policy makers' pressure for score-oriented culture. The study suggests that barriers limiting teachers' practice should be removed. It particularly highlights the importance of changing current assessment practices in Iran. The findings of this study also suggest that teachers need to change their conceptions to match the curriculum in order to implement them successfully. Finally, due to lack of data and limited studies on Iranian mathematics teachers' conception respect to assessment, this study has provided information on large scale among Iranian mathematic teachers that would be useful not only for Iranian mathematics researchers but also for all mathematics researchers around the world in research area of teachers' conception and assessment.

REFERENCES

Abdollahi, B., Dadjooye Tavakkoli, A., & Youseliani, G. (2014). Identification and validation of effective teachers' professional competencies. Quarterly Journal of Educational Innovations, 13(1), 7-24 [In Persian].

Adams, T. L., & Hsu, J.-W.Y. (1998). Classroom assessment: Teachers' conceptions and practices in mathematics. *School Science & Mathematics, 98*(4), 174-180.

Adeyemi, A. (2012) Effect of Peer and Self-Assessment on Male and Female Students' Self- Efficacy and Self-Autonomy in the Learning of Mathematics. *Gender & Behaviour, 10,* 4492-4508.

Alkharusi, H. (2011). Teachers' Classroom Assessment Skills : Influence Of Gender , Subject Area , Grade Level , Teaching Experience and In-service Assessment Training, *8*(2), 39–48.

Arani, A. M., Kakia, L., & Karimi, V. (2012). Assessment in education in Iran. *Sa-Educ Journal*, *9*(2), 1–10.

Arumugham, K. S., Abdullah, S. S., & Mahmud, N. (2017). Conceptions of formative assessment among Malaysian secondary school teachers. *International Journal of Multidisciplinary Education and Research*, *2*(2), 27–30.

Azis, Astuti. (2015). Conceptions and practices of assessment: a case of Teachers representing improvement conception. *teflinjournal. v26i2*, 129–154. Retrieved from http://dx.doi.org/10.15639

Ausubel, D. P. (1968). Educational psychology: A cognitive view. New York, NY: Holt, Rinehart & Winston

Aydın, M., Kö, D., & Yıldız, C. (2009). Mathematics teacher educators ' beliefs about assessment, *1*, 2126–2130. Retrieved from https://doi.org/10.1016/j.sbspro.2009.01.372

Bailey, K. M. (1998). *Learning about language assessment: dilemmas, decisions, and directions.* Heinle & Heinle: US.

Baird, J., Hopfenbeck, T. N., Newton, P., Stobart, G., & Steen-Utheim, A. T. (2014). State of the field review: Assessment and learning. Norwegian Knowledge Centre for Education. Oxford, England: Oxford University Centre for Educational Assessment.

Ball, D., Ferrini-Mundy, J., Kilpatrick, J., Milgram, J., Schmid, W., & Scharr, R. (2005). Reaching for Common Ground in K-12 Mathematics Education. Washington, DC: American Mathematical Society.

Barkatsas, A. (2003). Secondary mathematics teachers' beliefs about teaching, learning and assessment and factors that influence these beliefs. Unpublished Ph.D. Thesis. Curtin University of Technology.

Barnes, N., Fives, H., & Dacey, C. M. (2017). U. S. teachers' conceptions of the purposes of assessment. *Teaching and Teacher Education*, *65*, 107–116. doi: 10.1016/j.tate.2017.02.017

Benson, III, T. P. (2014). Sixth through eighth grade teachers' conceptions (beliefs) about assessment and practices (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3629573).

Berry, R. (2011). Assessment trends in Hong Kong: Seeking to establish formative assessment in an examination culture. *Assessment in Education: Principles, Policy & Practice, 18*(2), 199-211.

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

Bray, W. S. (2011). A collective case study of the influence of teachers' beliefs and knowledge on error-handling practices during class discussion of mathematics. *Journal for Research in Mathematics Education, 42*(2), 2-38. doi:10.5951/jresematheduc.42.1.0002

Britton, T. (2011). Using formative and alternative assessments to support instruction and measure student learning. *Science Scope, 34*(5)*,* 16-21. Retrieved from eric.ed.gov/? id=EJ912474

Brown, G. T. L. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice,* 11(3), 301e318. doi: 10.1080/0969594042000304609

Brown, G.T. L. (2002). *Teacher conceptions of assessment.* (Unpublished doctoral dissertation, University of Auckland, New Zealand).

Brown, G. T. L. (2006). Teachers' conceptions of assessment: Validation of an abridged instrument. *Psychological Reports*, 99(1), 166e170. doi: 10.2466/pr0.99.1.166-170.

Brown, G. T. (2007). Teachers' conceptions of assessment: Comparing measurement models for primary and secondary teachers in New Zealand. Paper presented at the Annual Conference of the New Zealand Association for Research in Education (NZARE), Christchurch, NZ.

Brown, G. T. L. (2008). *Conceptions of assessment: Understanding what assessment means to teachers and students*. New York, NY: Nova Science.

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,* 50e64

Brown, G. T. L., Hui, S. K. F., Yu, F. W. M., & Kennedy, K.J. (2011). Teachers' conceptions of assessment in Chinese contexts: A tripartite model of accountability, improvement, and irrelevance. *International Journal of Educational Research, 50,* 307-320.

Brown, G. T. L., Lake, R., & Matters, G. (2011). Queensland teachers' conceptions of assessment: The impact of policy priorities on teacher attitudes. *Teaching and Teacher Education, 27,* 210–220. doi:10.1016/j. tate.2010.08.003

Brown, G. T. L., & Remesal, A. (2012). Prospective teachers' conceptions of assessment: A cross-cultural comparison. *The Spanish Journal of Psychology, 15(1), 75–89. http://dx.doi.org/10.5209/rev_SJOP.2012.v15.n1.37286*

Brown, J., & Shavelson, R. (1994). Does your testing match your teaching style? *Instructor, 103*(7)*,* 86-89

Borko, H. Flory, M and Cumbo, K (1993). Teachers' ideas and practices about assessment and instruction: A case study of the effects of alternative assessment in instruction, student learning, and accountability practices. Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA, April 12-16, 1993.

Butterfield, S., Williams, A., & Marr, A. (1999). Talking about assessment: Mentor-student dialogues about pupil assessment in initial teacher training. *Assessment in Education, 6*(2)*, 225-246*

Calveric, S. B. (2010). Elementary teachers' assessment beliefs and practices. (Doctoral dissertation, Virginia Commonwealth University, Virginia, USA).

Carneiro, R., Looney, J., & Vincent-Lancrin, S. (2015). Learning from the Past, Looking to the Future: issues and agendas in education. European Journal of Education, 50(4), 524-535.

Chappuis, J., & Stiggins, R. (2008). Finding balance: Assessment in the middle school classroom. *Middle Ground, 12*(2)*,* 12-15. Retrieved from http://www.amle.org/portals/0/pdf/mg/oct2008.pdf

Chappuis, S., Chappuis, J., & Stiggins, R. (2009). The quest for quality. Educational Leadership, pp. 14-19.

Chappuis, S., Commodore, C., & Stiggins, R., (2010). Assessment balance and quality: An action guide for school leaders (3rd ed.). NY: Pearson.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd edition). Hillsdale, NJ: Lawrence Earlbaum Associates.

Cooney, T. J. & Shealy, B. E. (1995). Teachers thinking and rethinking assessment practices. Paper presented at the Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education (Columbus, OH, October 21-24, 1995).

Creswell, J. W. (2008). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (3rd ed.). Upper Saddle River, NJ: Pearson.

Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Boston, MA: Pearson Education.

Creswell, J. W. (2013). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.

Dandis, M. A. (2013). The assessment methods that are used in a secondary mathematics class Métodos de evaluación utilizados en una clase de matemáticas en enseñanza secundaria, *4*(2), 133–143.

Delors, J. (2013). The treasure within: Learning to know, learning to do, learning to live together and learning to be. What is the value of that treasure 15 years after its publication? International Review of Education, 59, 319-330.

Develles RF (1991). Scale Development. Sage Publications. pp. 24-33.

Dhindsa, H., Omar, K., & Waldrip, B. (2007, August 1). Upper Secondary Bruneian Science Students' Perceptions of Assessment. *International Journal of Science Education*, *29*(10), 1281-1280

Earl, L., & Timperley, H. (2014). Challenging conceptions of assessment. In C. Wyatt-Smith, V. Klenowski, & P. Colbert (Eds.), *Designing assessment for quality learning* (pp. 325–336). Dordrecht, The Netherlands: Springer.

Ernest, P. (1988). The impact of beliefs on the teaching of mathematics. Paper prepared for ICME VI, Budapest, Hungary.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2013). G*Power Version 3.1.7 [computer software]. Uiversität Kiel, Germany. Retrieved from http://www.psycho.uni-duesseldorf.de/abteilungen/aap/gpower3/download-and-register

Finnish National Board of Education. (2008). “Basic Education.” Retrieved from http://www.oph.fi/english/page.asp?path=447,4699,4847

Firestone, W. A., Mayrowetz, D., & Fairman, J. (1998). Performance-based assessment and instructional change: the effects of testing in Maine and Maryland. *Educational Evaluation and Policy Analysis, 20*(2), 95e113

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. 471e499). Washington: American Psychological Association.

Fives, H., & Buehl, M. M. (2014). Exploring differences in practicing teachers' valuing of pedagogical knowledge based on teaching ability beliefs. Journal of Teacher Education, 65(5), 435e448. http://dx.doi.org/10.1177/0022487114541813.

Fraenkel, J. R., Wallen, N. E., & Hyun, H. (2015). *How to design and evaluate research in education* (9th ed.). New York, NY: McGraw-Hill.

Frey, B. B., & Schmitt, V. L. (2010). Teachers' classroom assessment practices. *Middle Grades Research Journal,* 5(3), 107-117.

Fundamental reform document of education (FRDE, 2011) in the Islamic Republic of Iran. Retrieved from http://www.dres.ir/safeschool/Downloads/FRDE.pdf

Fung, D., Kutnick, P., Mok, I., Leung, F., Lee, B. P. Y., Mai, Y. Y., & Tyler, M. T. (2017). Relationships between teachers' background, their subject knowledge and pedagogic efficacy, and pupil achievement in primary school mathematics in Hong Kong: An indicative study. International Journal of Educational Research, 81, 119-130

Gage, N. L. (2001). The scientific basis of the art of teaching. [M. Mehrmohammadi, Trans.]. Tehran: School Publishing. [In Persian]

Gall, M. D., Gall, J. P., & Borg, W. R. (2006). *Educational research: An introduction (8th ed.). New York, NY: Allyn & Bacon. Retrieved from EBSCOhost.*

Gashaw, T. (2014). Teachers ` Perceptions and Practices of Continuous Assessment in Mathematics Class in Dera woreda General Secondary and Preparatory Schools. (Master dissertation). Addis Ababa University. Retrieved from http://etd.aau.edu.et/dspace/handle/123456789/7208

Gates, April A, (2008) *Wyoming Teachers' Knowledge and Use of Formative Assessment.* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3320747)

Gay, L., Mills, G., & Airasian, P. (2009). Educational research: Competencies for analysis and applications. Upper Saddle River, NJ: Pearson.

Gay, L. R. (1996). *Educational research: Competencies for analysis and application (*5th ed) New York: Macmillan.

Gholamāzad. S. (2011). Mathematical assessment. Magazine Growth of Mathematical Education, No. 105, p.14. (in Persian language).

Gholamāzad. S (2015). Assessment of mathematics curriculum of guidance school in Iran according to teachers' perspectives. *Quarterly Journal of Educational Innovations*14(1) 96-130

Gronlund, N. E. (1998). *Assessment of student achievement*. Boston, MA: Allyn and Bacon.

Gronlund, N. E. (2006). *Assessment of student achievement* (8th ed.). Boston: Pearson.

Gallagher, D., Jo. (1998). *Classroom assessment for teachers*. Upper Saddle River, NJ: Merrill.

Gooya. Z (2011). The evolution and formation of mathematical curriculum in Iran. *Quarterly Journal of Educational Mathematics, 28*(4). (In Persian).

Goh, C. L., & Wong, T. Bridging the education gap, *The Straits Times,* pp. D2-D3.

Guthrie, J. T. (2002). Preparing students for high-stakes test taking in reading. In A. E. Farstrup & S. J. Samuels (Eds.), *What research has to say about reading instruction* (pp. 370–391). Newark, DE: International Reading Association.

Handal, B., & Herrington, A. (2003). Mathematics teachers' beliefs and curriculum reform. *Mathematics Education Research Journal, 15*(1)*,* 59-69.

Hanushek, E. A., & Raymond, M. E. (2005). Does school accountability lead to improved student performance? Journal of Policy Analysis and Management, 24(2), 297-327. doi: 10.1002/pam.20091

Harlen, W. (2012). On the relationship between assessment for formative and summative purposes. In J. Gardner (Ed.), *Assessment and learning* (2nd ed.) London: Sage

Herrington, A., Herrington, J. & Glazer, E. (2002). Authentic approaches to learning assessment strategies: Beginning teachers' practices in classrooms. Proceedings of the Annual Meeting of the North American chapter of the International Group for the Psychology of Mathematics Education, Athens, GA.

Harris, L., Irving, S., & Petterson, E., (2008, December). Secondary teachers' conceptions of the purpose of assessment and feedback. Paper presented at the annual conference of the Australian Association for Research in Education, Brisbane, Australia.

Hogan, D. (2014). Why is Singapore's school system so successful, and is it a model for the West? Retrieved from http://theconversation.com/why-is-singapores-schoolsystem- so-successful-and-is-it-a-model-for-the-west-22917

Huba, M. B., & Freed, J. E. (2000). *Learner-centered assessment on college campuses: Shifting the focus from teaching to learning*. Boston, MA: Allyn and Bacon

Hudaee. M (2009) Twenty Years of Experimenting and Bounce: A Look at General Secondary School Curriculum in Two Decades. First present Quarterly Journal of Education, No.98. (in Persian).

Huff, S. (2009). Build, promote, guide, provide, monitor: Action words for principals as instructional leaders in assessment. In T. R. Guskey (Ed.), *The principal as assessment leader* (pp. 31-51). Bloomington, IN: Solution Tree Press.

Iranian National Curriculum (2010). Iran, Tehran: Ministry of Education.

Jakicic, C. (2009). A principal's guide to assessment. In T. R. Guskey (Ed.), *The principal assessment leader* (pp. 53-70). Bloomington, IN: Solution Tree Press.

Kahn, E. A. (2000). A case study of assessment in a grade 10 English course. The Journal of Educational Research, 93, 276–286. http://dx.doi.org/10.1080/00220670009598719

Kakia, L., & Almasi, A. (2008). *A comparative study of effectiveness's evaluation of qualitative and quantitative assessment methods on learning, anxiety and attitude of primary pupils of Tehran City*. Paper presented at Formative Assessment National Conference, Khoramabad: Lorestan's Education Organization.

Kaiser, J., Südkamp, A., & Möller, J. (2017). The effects of student characteristics on teachers' judgment accuracy: Disentangling ethnicity, minority status, and achievement. Journal of Educational Psychology, 109(6), 871

Keaikitse, S. K. (2012). *Classroom Assessment Practices : a Survey of Botswana Primary and Secondary School Teachers*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3507918).

Kelly, G. A. (1991). The psychology of personal constructs: A theory of personality (Vol. 1). London: Routledge.

Koh, K., & Luke, A. (2009). Authentic and conventional assessment in Singapore schools: An empirical study of teacher assignments and student work. *Assessment in Education: Principles, Policy and Practice, 16*(3)*,* 291–318.

Koch, M. J. (2013). The multiple-use of accountability assessments: Implications for the process of validation. Educational Measurement: Issues and Practice, 32(4), 2–15.

Koh, K. H. (2011). Improving teachers' assessment literacy through professional development. *Teaching Education, 22*(3), 255-276.

Lavonen, J. (2008). "Reasons behind Finnish Students' Success in the PISA Scientific Literacy Assessment." University of Helsinki, Finland. Retrieved from http://www.oph.fi/info/finlandinpisastudies/conference2008/science_results_and_reasons.pdf

Law, B., & Eckes, M. (1995). *Assessment and ESL*. Manitoba, Canada: Peguis.

Leahy, S., Lyon, C., Thompson, M., & Wiliam, D. (2005). Classroom assessment: Minute by minute, day by day. *Educational Leadership, 63*(3)*,* 19–24.

Laukkanen. R (2008). Finnish strategy for high-level education for all. In N. C. Soguel & P. Jaccard (Eds.), *Governance and Performance of Education Systems* (pp. 305–324). Springer, at p. 319.

Ludwig, N. W. (2013). Exploring the relationship between K-12 public school taechers' conception of assesment and their classroom assessment confidence levels. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3579798).

McMillan, J. H. (2007). *Classroom assessment: Principles and practice for effective standards-based instruction* (4th ed.). Boston, MA: Pearson.

McMillan, J. H. (2008). *Assessment essentials for student-based education* (2nd ed.). Thosand Oaks, CA: Crown Press.

Mehran, G. (2003). The Paradox of Tradition and Modernity in Female Education in the Islamic Republic of Iran. Comparative Education Review, 47(3). pp. 269-286, August.

Mertens, D. M. (2010). *Research and evaluation in education and psychology. Integrating diversity with quantitative and qualitative and mixed methods* (3rd ed.). Thousand Oaks, CA: Sage.

Morris, P., & Adamson, B. (2010). *Curriculum, schooling and society in Hong Kong*. Hong Kong: Hong Kong University Press.

Musapour, N. (2011) *Basics of Secondary Education Planning,* Mashhad: Astan Quds Razavi Publication. (in Persian)

Mussawy, S. A. J. (2009). *Assessment practices : Student's and teachers' Perceptions of classroom assessment.* (Master dissertation, University of Massachusetts - Amherst). http://scholarworks.umass.edu/cie_capstones

Nasab, F. G. (2015). Alternative versus traditional assessment. *Ournal of Applied Linguistics and Language Research*, *2*(6)*,* 165–178.

National Council of Teachers of Mathematics. (1989). *Assessment standards for school mathematics.* Reston. VA: Author

National Council of Teachers of Mathematics (1995). *Assessment Standards for School Mathematic*s. Reston, VA: Author.

National Council of Teachers of Mathematics. (2000). *Principles and standards for mathematics*. Reston, VA: Author.

National Council of Teachers of Mathematics (2005). *Curriculum Focal Points for Pre-Kindergarten through Grade 8 Mathematics*. Reston, VA: Author.

National Council of Teachers of Mathematics (NCTM). (2014). *Principles to action: Ensuring mathematical success for all*. Reston, VA: Author

Ndalichako, J. L. (2015). Secondary school teachers' perceptions of assessment. *International Journal of Information and Education Technology, 5*(5)*,* 326–330. doi: 10.7763/IJIET.2015.V5.524

Neill, M. (1997). *Testing our children: A report card on state assessment systems.* Cambridge, MA: Fair Test.

Newton, P. E. (2007). Clarifying the purposes of educational assessment. *Assessment in Education: Principles, Policy & Practice, 14,* 149–170.

Newton, P.E.; (2010) Educational assessment – concepts and issues: the multiple purposes of assessment. In: Peterson, P. and Baker, E. and McGaw, B., (eds.) International Encyclopedia of Education. Third Edition. (pp. 392-396). Elsevier: Oxford.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw-Hill

OECD (2011). Lessons from PISA for the United States, Strong Performers and Successful Reformers in Education, OECD Publishing. doi: 10.1787/9789264096660-en

Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research, 62,* 307–332.

Pallant, J. (2007). SPSS survival manual: A step by step guide to data analysis using SPSS for windows (3rd ed.). New York: Open University Press.

Parke, C., Lane, S., Silver, E. A., & Magone, M. (2003). Using assessment to improve mathematics teaching and learning: Suggested activities using QUASAR tasks, scoring criteria, and student work. Reston, VA: National Council of Teachers of Mathematics.

Pangeni, K. P. (2014). Factors determining educational quality: Student mathematics achievement in Nepal. International Journal of Educational Development, 34, 30-41.

Pfannkuch, M. (2001). Assessment of school mathematics: Teachers' perceptions and practices. *Mathematics Education Research Journal, 13*(3)*,* 185-203.

Phelan, J. P., & Phelan, J. C. (2010). Classroom Assessment Tasks and Tests. *International Encyclopedia of Education*, *3*(3)*,* 209- 219. Oxford: Elsevier.

Pishghadam, R., Adamson, B., & Shayesteh, S. (2014). Assessment in education :

Principles, policy & practice Conceptions of assessment and teacher burnout, 37–41. doi: 10.1080/0969594X.2013.817382

Pishghadam, R., & Shayesteh, S. (2012). Conceptions of assessment among Iranian EFL teachers. *Iranian EFL Journal, 8,* 9–23.

Porahmadi, M. (2008). *Assessment instruments characteristics and information gathering at formative assessment.* Paper presented at Formative Assessment National Conference. Khoramabad: Lorestan's Education Organization.

Popham, W. J. (2009) Assessment literacy for teachers: Faddish or fundamental? Theory into Practice, 48, 4-11. doi:10.1080/00405840802577536

Popham WJ. (2011) Transformative assessment in action: An inside look at applying the process. ASCD.

Pratt, D. D. (1992) Conceptions of teaching. *Adult Education Quarterly*, *42*(4), 203-220.

Puhl, C. A. (1997). Develop, not judge: Continuous assessment in ESL class room. *English Teaching Forum, l35*(2).

Raboijane, B. (2005). Mathematics teachers' understanding of alternative assessment as applied in junior secondary schools in Gaborone (Botswana), Doctoral dissertation. Retrieved from http://hdl.handle.net/10539/1610

Radhakrishna, R. B. (2007). Tips for developing and testing questionnaires/instruments. *Journal of Extension, 45*(1). Retrieved from www.joe.org/joe/2007/

Rahim. S. S. A., & Sulaiman.T (2006). Classroom Assessment: Paper-pencil vs Alternative Humanising Assessment: Compilation of presentation papers (pp. 49-52).

Resnick, L. B., & Resnick, D. P. (1992). Assessing the thinking curriculum: New tools for educational reform. In B. R. Gifford & M. C. O'Conner (Eds.), *Changing assessments: Alternative views of aptitude, achievement, and instruction* (pp. 37-75). Boston, MA: Kluwer Academic.

Reyhani, E (2015). Analysis of policy documents approved, reputable research and resources related to the field of mathematics. Research, development and mathematical education in Persian language.

Reutzel, D. R., & Cooter, R. B. (2016). Strategies for reading assessment and instruction in an era of common core standards (5th ed.). Boston, MA: Pearson.

Romberg, T. (2000). Changing the teaching and learning of mathematics. *AMT, 56* (4), 6-9.

Ross, J. A., McDougall, D., Hogaboam-Gray, A., & LeSage, A. (2003). A survey measuring elementary teachers' implementation of standards-based mathematics teaching. *Journal for Research in Mathematics Education, 34*(4)*,* 344–363. doi:10.2307/30034787

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

Saad, M., Sardareh, S. A., & Ambarwati. A. E. (2013). Iranian Secondary School EFL Teachers' Assessment Beliefs and Roles. *Life Science Journal, 10*(3), 1638–1647.

Sahlberg, P. (2009). A short history of educational reform in Finland. Retrieved from https://www.researchgate.net/file.PostFileLoader.html?id=5427c060d3df3ed1668b457b&assetKey=AS%3A271763508793347%401441804877979

Saif. A. (2011). *Measurement, assessment and evaluation* (5th ed.). Tehran, Iran: Publishing Duran.in Persian language.

Samadi, P. (2001). *To design an optimal religious curriculum for secondary school education, and the level of agreement with its official curriculum*. (Doctoral thesis, Tarbiat Modarres University).

Segers, M., & Tillema, H. (2011). Studies in Educational Evaluation How do Dutch secondary teachers and students conceive the purpose of assessment? *Studies in Educational Evaluation*, *37*(1), 49–54. doi: 10.1016/j.stueduc.2011.03.008

Senk, S., Beckmann, C., & Thompson, D. (1997). Assessment and grading in high school mathematics classrooms. *Journal for Research in Mathematics Education, 28*(2), 187 – 215. Retrieved from http://math.coe.uga.edu/olive/EMAT6990Sp10/JRME1997-03-187a.pdf

Shepard, L. A. (2005) Linking formative assessment to scaffolding. *Educational Researcher, 63*(3)*,* 66–70.

Shohamy, E. (2001). The power of tests: A critical perspective on the uses of language tests. Harlow, UK: Pearson Education.

Simonson M., Smaldino, S, Albright, M., & Zvacek, S. (2000). Assessment for distance education (ch 11). *Teaching and learning at a distance: Foundations of distance education*. Upper Saddle River, NJ: Prentice Hall.

Simpson, A. K. (2016). *The attitude of seconadry mathematics taechers toward authentic assessment in their classroom*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 10032444).

Smith, K. (2011). Professional development of teachers: A prerequisite for AfL to be successfully implemented in the classroom. *Studies in Educational Evaluation, 37*(1)*,* 55- 61.

Statistical Centre of Iran. (n.d.). *Statistical yearbook, Ministry of Education, 2012–2013*. Tehran: Author.

Stiggins, R. (2005). *Assessment for learning defined.* Paper presented at the ETS/Assessment Training Institute's International Conference: Promoting Sound Assessment in Every Classroom, Portland OR

Stiggins, R. J. (2001). The principal's leadership role in assessment. *NASSP Bulletin, 85*(621),13-26.

Stiggins, R. J. (1995). Assessment literacy for the 21st century. Phi Delta Kappan, 77(3), 238-245. http://www.kappanmagazine.org/

Stobart, G. (2008). *Testing times: The uses and abuses of assessment*. London: Routledge.

Stobart, G., & Gipps, C. (2010). Alternative assessment. In B. McGraw, E. Baker, & P. Peterson (Eds.), *International encyclopedia of education* (3rd ed., pp. 202–208). Oxford, UK: Elsevier.

Susuwele-Banda, W. J. (2005). *Classroom assessment in Malawi: Teachers' perceptions and practices in mathematics.* Doctoral dissertation, Virginia Polytechnic Institute and State University. Retrieved from http://scholar.lib.vt.edu

Suurtamm, C., & Koch, M. J. (2014). Navigating dilemmas in transforming assessment practices: Experiences of mathematics teachers in Ontario, Canada. Educational Assessment, Evaluation and Accountability, 26(3), 263–287.

Snyder, M. R. (2017). Investigating the link between current classroom teachers' conception, literacy , and practices of assessment. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 10254508).

Tan, K. H. K. (2007). *The case for qualitative approaches to assessment. Alternative assessment in schools: A qualitative approach*. Singapore: Pearson Education South Asia.

Tan, K. (2011). Assessment for Learning Reform in Singapore: Quality, sustainable or threshold? *Assessment Reform in Education* (pp. 75-87).

Thompson, A. G. (1992). Teachers' beliefs and conceptions: A synthesis of the research. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 127e146). Reston, VA: National Council of Teachers of Mathematics.

Torrance, H., & Pryor, J. (1998). Investigating formative assessment: Teaching, learning and assessment in the classroom. Oxford: Taylor & Francis.

Trends in International Mathematics and Science Study. (2010). Retrieved from http://timss.org/

Trends in International Mathematics and Science Study. (2015). Retrieved from http://timss.org/

Twycross, A. & Shields, L. (2004). Validity and reliability - What's it all about? Part 2 Reliability in quantitative studies. Paediatric Nursing, 16 (10) p. 36

Vardar, E. (2010). Sixth, seventh, and eighth grade teachers' conception of assessment. (master's Thesis, Middle East Technical University, Ankara, Turkey).

Van Duinen, D. V. (2005). Authentic assessment: Praxis with power. *International Journal of Learning, 12*(6), 141-148. Retrieved from http://www.scimagojr.com/journalsearch.php?q=17700156750&tip=sid&clean=0

Vik, L. B. (2013). *The assessment of English in two Norwegian upper secondary schools: A comparative study*, Master dissertation, univesrity of Stavanger. Retrieved from http:// brage.bibsys.no

Warren, E., & Nisbet, S. (1999). The relationship between the purported use of assessment techniques and beliefs about the uses of assessment. Paper presented at the 22nd Annual Conference of the Mathematics Education and Research Group of Australasia, Adelaide, SA.

Watt, H. (2005). Attitudes to the use of alternative assessment methods in mathematics: A study with secondary mathematics teachers in Sydney, Australia. Educational Studies in Mathematics, 58(1), 21-44. doi:10.1007/s10649-005-3228-z

Webb, D. C. (2012). Teacher change in classroom assessment: The role of teacher content knowledge in the design and use of productive classroom assessment. In *Proceedings of the 12th International Congress on Mathematical Education: Topic Study Group 33* (pp. 6773– 6782). Seoul, Korea.

Webb, N. L. (1992). Assessment of students' knowledge of mathematics: Steps toward a theory. In D. A. Grouws (Ed.), Handbook of research on mathematics teaching and learning. (pp. 661–683). New York: Macmillan.

Windschitl, M. (2002). Framing constructivism in practice as the negotiation of dilemmas: An analysis of the conceptual, pedagogical, cultural, and political challenges facing teachers. *Review of Educational Research, 72*(2)*,* 131–175

Woolfolk Hoy, A., Davis, H., & Pape, S. J. (2006). Teacher knowledge and beliefs. In P. A. Alexander, & P. Winne (Eds.), Handbook of educational psychology (2nd ed., pp. 715e738). Mahwah, NJ: Lawrence Erlbaum.

Wu, Z., & An, S. (2007). Using alternative assessment to assess students' learning. In N. Idris (Ed.), *Classroom assessment in mathematics education* (pp. 19-31). Kuala Lumpur, Malaysia: McGraw-Hill

Zhang, Z. R., & Burry-Stock, J. A. (2003). Classroom assessment practices and teachers 'self-perceived assessment skills. *Applied Measurement in Education, 16,* 323-342.