Teachers' Professional Identity Measures in the Kingdom of Saudi Arabia from the Perspective of Rasch Model Framework

AlghamdiHamad Abdullah J[1], Mohd Ali Samsudin[2], ShamimahParveen Abdul Rahim[3], Mohammed Ali H Alqarni[4],Mohd Erfy Ismail[5]

[1,4]Saudi Ministry of Education, Saudi Arabia
[2,3]School of Educational Studies, Universiti Sains Malaysia, Malaysia
[5]Faculty of Technical and Vocational Education, Universiti Tun Hussein Onn Malaysia, Malaysia

ABSTRACT

This study aims to assess the psychometric properties of the professional identity measure of the professional identity of Mathematics teachers in the Kingdom of Saudi Arabia. Professional identity topic has not been widely addressed in the Arab environment. Studying the Mathematics Teachers' Professional Identity (TPI) is important because only few and inadequate body of research has been performed in the Arab environment. Establishing measuring tool on the TPI contributes on the evaluation of the teachers in Kingdom of Saudi Arabia and their roles in the educational process in the country. This study applies descriptive approach. 600 secondary Mathematics teachers were involved. The adapted version of the questionnaire designed by Albaqi'i (2014) used to measure the Mathematics TPI. Delphi techniques in three rounds to gather experts' opinion about the suitability of Mathematics TPI questionnaire used in the Kingdom of Saudi Arabia context. Rasch model as a framework applied to assess the psychometric properties of Mathematic TPI questionnaire. The results are valuable in terms of analyzing the psychometric properties for the first version of Mathematics TPI, to improve Differential Item Functioning (DIF) and the psychometric properties for the revised version of Mathematics TPI Questionnaire.

Introduction

Many factors contribute in forming the professional identity of teachers (Bolívar & Domingo, 2006; Graven & Lerman, 2014), such as the major development in the teacher work and the educational challenges that affect the teacher's performance (Coldron & Smith, 1999). Due to that, it was necessary to measure the professional identity, especially since this topic has not been widely addressed in the Arab environment. Few researchers only tried to highlight the important aspect of the TPI in the educational process(Albaqi'i, 2014; Abdulghani & Abulnaeem, 2016; Hussein, 2017). Thus, this field is highly needed to be addressed in the Arab environment due to the complex environment and culture in this area. Furthermore, Graven (2004) and Graven and Lerman (2014) argued that there are many aspects of the TPI that are integrally connected to the particular nature of the Mathematics subject.

TPI differs only with regard to beliefs about and views of the subject itself, which plays a significant role in the professional identity of the Mathematics teacher. However, many researchers have argued that researchers working with the notion of Mathematics teacher's identity have not clearly defined and operationalized the notion (Graven & Lerman, 2014). Moreover, despite the importance that has been highlights in the literature for studying the Mathematics TPI, only few and inadequate body of research has been performed in the Arab environment.

Problem Statement

This view is supported by the argument that the contribution of Mathematics to scientific and technological development has influenced Mathematics itself, and has become a tool for developing thinking and achieving relationships in an integrated framework in the individual after viewing it as merely acquiring normal mathematical skills. Traditional Mathematics teaching methods is no longer acceptable today. According to Ball (2003), teaching Mathematics requires the necessary skills that must be exercised by the teacher who is able to work effectively, whether it is related to the objectives of teaching Mathematics, or how to prepare, implement and evaluate. The complexity of Mathematics teachers to develop the necessary skills can be varied depending on their interpretations of their ongoing interaction with their context (Beijaard, Meijer, &Verloop, 2004). Kelchtermans (2009) stated that teachers develop an interpretative framework during their careers and that this framework is formed and restructured through interaction between individual teachers and social, cultural and structural working conditions in the context of their work. The measurement of this interpretive framework is called professional which relates to how teachers perceive themselves as teachers, based on their interpretations of their ongoing interaction with their context (Beijaard et al., 2004).

As well as Dabous (2016) and Hussein (2017) highlighted the need for a precise measurement tool to measure the professional identity of teachers that fit to the criteria for objective measurement and the culture in the Arab

environment. This led to the importance of test the validity and the reliability of the measurement based on the modern measurement theories such as the Item Response Theory (IRT). This is what motivated the researcher to prepare an objective measure of the professional identity of teachers in the context of Saudi Arabia. Furthermore, Saudi Arabia had significant transform in the educational process. Several conferences were held to support the development of the teachers and the education in the Kingdom of Saudi Arabia and the first conference was in (1973), the second one in (1995) and the third in (1999) to prepare the teacher of public education in the Kingdom of Saudi Arabia. Recently the education in the Kingdom of Saudi Arabia gain a great attention form the government in the country which led to the important of perform this research in such country. Furthermore, providing a well-tested tool to measure the TPI will provide a trusted profile of the teachers and these profiles will contribute on the evaluation of the teachers in Saudi Arabia and their roles in the educational process in the country.

Literature Review

Psychology researchers were interested in the concept of identity and due to this growing interest and research on this concept, it has been divided into several forms; which are the psychological identity, cultural identity, social identity and professional identity. The institutions of social formation play a vital role in forming and developing the identity in its various forms among individuals. According to Erikson (1968), identity formation is an important dimension of the human personality. It determines the individual's behavior pattern, system of management, self-evaluation, and to use his energies and abilities effectively to achieve his goals. Accordingly, this will create a sense of self-efficacy and positive appreciation for the individual that helps him to adapt with his psychological and social environment.

The concept of professional identity has attracted the interest of researchers in various medical, legal and educational fields. But the focus of recent studies on the professional identity was higher in the educational field, specifically on the extent of teachers' sense of perception their identity and how to develop it, and on the factors affecting the formation of that identity, such as the work conditions (Schepens, Aelterman, & Vlerick, 2009). Researching on the formation of TPI is very important in understanding how teachers feel towards their changing roles in today's schools; because many educational issues are changing rapidly, forcing teacher to deal and control differently with these changes and how facing them (Beijaard et al., 2004). The development of the professional identity of teachers has received more attention from decision-makers and researchers and based on the fact that reform of the education system set new requirements and standards for quality related to the professional identity of teachers. This means increasing the teacher's awareness of teaching profession in terms of: how the teacher will work?

How the performance could be improved and how the professional growth occurs? (Živković, 2013).

Professional identity is defined as the sense of the teacher in his or her own role through the roles he or she performs inside and outside the school, and through the knowledge, skills and learning competencies, and the commitment to the values and ethics of his profession (Albaqi'i, 2014). While others defined professional identity as the group of attitudes, beliefs and hobbies that individual has in matters and attitudes related to the profession that the individual seeks to achieve (Alzubaidi & Al-Kahali, 2014). According to Goldin, Hannula, Heyd-Metzuyanim, Jansen, Kaasila, Lutovac and Pantziara (2016), professional identity is the way in which the individual is associated with a specific society and participate within it, and it is a dynamic structure changes over time. In general, the TPI is described as the teacher's own vision as an expert in the academic content of the subject, an expert in teaching, and an educational expert. A number of factors influence the forming of the TPI, including the context and environment of the teaching, the teaching experience, and the curriculum vitae. The professional identity of the teacher is a continuous process of integration between subjective and professional aspects.

On the other hand, the subject of the construction of standards occupies an important place in the educational and psychological literature as it has been studied with great attention by psychologists and educators. Despite the rich pedagogical literature in describing the different methods and procedures to build psychological standards according to the models of IRT especially in Rasch model framework, but it did not receive great attention at the level of Arab studies and research (Hamadneh, 2015; Dabous, 2016; Sabah, 2017), the researcher found that the majority of psychological standards used in the Arab studies were built according to the concepts of the Classical Test Theory (Hamadneh, 2009; Ithawi, 2009; Abdulwahab, 2010; Ali, 2012; AL-Zayla'I, 2014). There are many criteria used to select measurement items in general, and trend metrics in particular and the vast majority of these standards stemmed from Classical Test Theory of measurement (AL-Nabhan, 2004).

Method

The aim of this study is to use the modern theory of measurement which is known as the Rasch model framework in forming the standards of the TPI in the Kingdom of Saudi Arabia. This study used adapted version of the questionnaire that designed by Albaqi'i (2014) to measure the Mathematics TPI of the participants of this study. The adapted instrument contains 45 items as showed in Appendix 1. This study aimed at Mathematics teachers in Tabuk educational region, Hail educational region and Sakaka educational region who works in the Saudi Arabia Government Secondary schools in the regions. From each region 200 teachers purposively selected from each region.

The descriptive approach is used in this study, which is a method of scientific research based on the study of reality or phenomenon as it exists in reality and describe it accurately and express it metaphorically in terms of describing the phenomenon and clarifying its characteristics or in quantitative expression in terms of giving a numerical description of the amount or size of this phenomenon and degrees of correlation with other phenomena. Since the present study aims to build and develop the professional identity scale for Mathematics teachers in the Kingdom of Saudi Arabia, the descriptive approach is the appropriate method for the subject of the current study. This study involves two stages, at first stages, the questionnaire is administered to the samples for the psychometric properties' analysis of TPI questionnaire. The questionnaire is considered as the First Version of Mathematics TPI questionnaire. The data obtained from the administration of the first version were analyzed based on the Rasch model psychometric properties such as person and item reliability, dimensionality, fit values, the category structure and DIF analysis. The results of the psychometric properties for the first version of the questionnaire are used as basis for the further improvement of the questionnaire at the second stage of the study. Therefore, the revised version of the questionnaires was administered to the sample of the study at the second stage of questionnaire. The data obtained from the second stage of the study is analyzed using Rasch model technique in order to check whether there are some improvements of its psychometric properties' values such as person and item reliability, dimensionality, fit values, the category structure and DIF analysis.

Data Analysis

Rasch model is used in this study as a framework to assess the psychometric properties of Mathematic TPI questionnaire in Saudi Arabia context. In this study, Rasch analysis is a psychometric technique that was developed to improve the precision with which researcher assess the instrument quality, and compute Mathematics TPI measurement.Rasch analysis also helps to think in more sophisticated ways with respect to the constructs variables that wish to measure. Some life sciences education researchers are already using Rasch techniques (Reeves &Marbach-Ad, 2016), but many continue to use instrument development and validation approaches that rely on Classical Test Theory (CTT).

Rasch techniques have greatly impacted the manner in which social science research makes use of tests and surveys (Panayides, Robinson, &Tymms, 2010). In this study, the Rasch framework offers procedures for constructing and revising Mathematics TPI questionnaire in Arab Saudi context and documenting measurement properties of the questionnaire. Rasch techniques also enable research to make critical corrections when using raw survey data of Mathematics TPI questionnaire. Specifically, Rasch techniques allow nonlinear raw data to be converted to a linear scale, which then can be evaluated through the use of parametric statistical tests (Boone, Staver, & Yale, 2014).

Rasch model framework produces a set of probabilistic models developed for the purpose of describing response patterns of respondents to individual items. This framework provides some advantages over other measurement approaches, such as CTT, in the sense that measures obtained with a Rasch analysis are expressed on an interval scale. Additionally, items' measures are not sample dependent nor vice versa (i.e., persons' measures are not dependent on the items included in an instrument). Also, individual errors for each subject are provided that can be used in practice to further examine a test taker's score. Another concern that emerged with the analysis of the model involved the notion of unidimensionality of the scale The Rasch model requires a scale to consist of one primary construct in order to develop a stable "yardstick" composed of a set of psychometrically sound quality items. The presence of multiple latent constructs in a single scale introduces noise, and ultimately hinders the precision and accuracy with which a single construct can otherwise be measured. Thus, unidimensionality is a desired property for constant measures that are designed for the purpose of creating a research measure.

Results

(i)   The Psychometric Properties for the First Version of Mathematics TPI Questionnaire

The findings of Rasch model analysis showed that the person reliability index of an individual is 0.90. The finding of the analysis show that the individual separation index is 2.95 which indicates that there are approximately three levels of respondent ability in the TPI questionnaire. The analysis result shows that the item reliability index is 0.93. The findings show the item separation index is 3.56 which indicates that there are approximately four levels of item difficulty in the Mathematics TPI questionnaire. The results of Rasch-residual-based Principal Components Analysis (PCA) of the standardized residuals for the Mathematics TPI questionnaire showed that the eigenvalue of the unexplained variance in the first contrast was 5.4 which is larger than the standard value. Thus, the measure can be considered unidimensional. In addition, the PCA of the standardized residuals showed that the variance explained by measures was 52.9%, assuming that unobserved data matches Rasch model predictions which at least 40% (Linacre, 2006).

The findings show that the item's mean squared infit is between 0.86 to 1.25 and the item mean squared outfit is between 0.81 to 1.34. This is aligned with Bond and Fox (2015), for Likert scale items, the item that is compatible with the Rasch model assumption should be in the range of 0.6 to 1.4. The result of PTMEA Corr analysis showed that psychometric properties of Mathematics TPI questionnaire did not have any such items. All the items in the instrument showed high positive PTMEA Corr.

Item No.	*An Inter content jlinaDIF. asked in the original Item*	DIF	Feedback from Arab Math teachers about the content of the original item	Suggestion by the expert for the improvement of the item	Revised item
Q01	I have a comprehensive and profound knowledge of	Qualification	The level of question is too high to be agreed by Math teacher as some of Math teachers	The item should be maintained but the sentence can be rephrased to be more	I have a sufficient knowledge of mathematics

An essential conceptual feature of rating scale design is that increasing amounts of the underlying variable in a respondent correspond to increasing probabilities

of the respondent being observed in higher categories of the rating scale (Linacre, 2002). Rasch analysis provides an effective framework within which to verify, and improve, the functioning of rating scale categorization. Therefore, the category structure is analyzed to confirm whether the patterns fits with Rasch model framework. The findings show the category mean squared infit is between 0.96 to 1.08 and the category mean squared outfit is in the range of 0.95 to 1.05. Thus, in term of fit analysis, the category measures displayed a pattern that fits the Rasch model assumption of sequencing from easy to difficult.

However, an issue arises when interpreting the results of step difficulties advances based on category measures. This can be observed when the category labelled as 1 has a category measure of -2.33 logit, followed by category label of 2 has a measure of -0.81 logit, category label of 3 has a measure of 0.06 logit and category label of 4 has a measure of 0.86 logit, and finally, category label of 5 has a measure of 2.15 logit. The problematic step difficulties advances can be identified from Category 2 to Category 3 (value of a step difficulty advance = 0.87), and Category 3 to Category 4 (value of a step difficult advance = 0.80). The values of step difficulties advances did not surpass 1.0 logits. These results are contradicted with Linacre (2002) statement that for a five-category rating scale, advances of at least 1.0 logits between step calibrations are needed in order for that scale to be equivalent to four dichotomies.

According to Linacre (2002), when the advance is less than 1.0 logits, redefining the categories to have wider

substantive meaning or combining categories may be indicated. Stone and Wright (1994) demonstrate that, in a survey of perceived fear, combining five ordered categories into three in the data increases the test reliability for their sample. Zhu, Updyke, & Lewandowski(1997) report similar findings for a self-efficacy scale. Based on this reasoning, the five-points Likert scale Mathematics TPI questionnaire are rescaled to three points Likert scale Mathematics TPI questionnaire, and the result of category structure of three points Likert scale Mathematics TPI questionnaire is reported at second phase of Rasch model analysis.

The purpose of conducting the analysis on DIF is to determine whether the items in TPI questionnaire are functioning fair or bias when administered to sample based on certain demographic factors. In this study, DIF is done based on teacher's experience (novice and non-novice), gender and qualification. For the teaching experience, the findings showed that there is one items which had t value more than 2 with p-value is less than 0.5. This item is Q12. Therefore, Q12 is revised for the improvement of instrument validity. For gender purpose, the finding showed that there is only one item which had t value more than 2 with p value that is less than 0.5. This item is Q34. Therefore, Q34 has been revised for the improvement of instrument validity

For qualification factor, the finding showed that there are three items which had t value more than 2 with p value that is less than 0.5. Those items are Q01, Q06, Q25 and Q37. Therefore, those four items have been revised for the improvement of the instrument validity.

(ii)   Improvement of DIF Items

	mathematics education		have different level of qualification, not all Math teachers have advanced degree qualification. The word 'comprehensive' is too subjective for Mathematics Education. The Mathematics Education can also be comprising of Educational Technology, which is not really mastered by Math teachers. The Math teachers may have sufficient knowledge in Education Technology to be integrated into Mathematics Education, but not at the level of comprehensive and profound.	moderate. Replace the word of comprehensive and profound with more achievable word such as 'sufficient'.	education.
Q06	I am contributing to build and develop the school mathematics curriculum.	Qualification	The main task of Math teacher is to teach, this is also including to translate the curriculum into teaching and learning practice in the classroom. At the level as Math teachers, they have already overwhelmed jobs to teach Math based on the curriculum, therefore it is not possible for them to have ample time to build and develop the school mathematics curriculum. Not all Math teachers have been selected by the Ministry of Education to involve with process of developing curriculum.	The item is too high for ordinary Math teachers who are the implementer of the curriculum. However, the item is important to be asked to math teachers to indicate teachers' professional identity from the aspect of teaching and learning. Rather than setting a high benchmark for majority of teachers who works at school, but not at the ministry level, therefore, it is fair to rephrase the wordings to represent majority of Math teacher. As an example, teacher give feedback for the improvement of the curriculum.	I give feedback for the improvement of school mathematics curriculum.
Q12	I collaborate with teachers to organize activities suited to the needs of students in the field of	Experience	It is challenging to collaborate with other teachers. As math curriculum is not integrated with other subjects. Teachers does not have strong networking to	It is better to focus on one dimension to be asked in the question. The statement asks two things in one time which are "collaborate" and "organize activities suited the needs of	I organize activities suited to the needs of students in the fields of mathematics.

	mathematics.		collaborate with teachers with different field. In is time consuming to collaborate with other teachers in different field as there are plenty Math content to be covered in the curriculum. As long teacher can ensure the students understand the Math content through activities that suits with the students' level, they feel that is not necessary to collaborate with other teachers.	students". Teacher may confuse to agree or disagree with one of those two things. It is better to ask about how agree they are to organize activities suited to the needs of student. This is because based on the teachers' comments, it is challenging to do collaboration with other teachers due to the reason of time constraints and packed syllabus in the curriculum.	
Q25	I follow the latest scientific and educational developments.	Qualification	The word scientific is not familiar for the use teaching and learning innovation. For math teacher, they perceive 'scientific' is a term used for the research in the Science and Technology field.	This statement deals with teachers' professional identity from the aspect of professional development. Therefore, this item is important to be asked, however, if the term 'scientific' is being discarded from the content of the item, the meaning of the item still represent the professional development aspect. This is due to reason that the item still asks about how agree the teachers follow the latest educational developments.	I follow the latest educational developments.
Q34	I show great flexibility in the educational learning process.	Gender	"Great flexibility" seems to be difficult for female teachers, especially when communicating with students during the learning process. They prefer to have some boundaries to interact with male students which involve more formal conversation between female teachers with male students.	The item should be culturally sensitive.	I show some flexibility in the educational learning process.
Q37	I use sources of information with a good scientific	Qualification	Not all Math teachers familiar with scientific methodology, unless they have taken a	The term good scientific methodology is too complex to be understood by teachers	I use sources of information wisely.

methodology.	Master degree courses or above. For those who only have Bachelor degree qualification, the term scientific methodology is not being exposed sufficiently to them. They aware with scientific methodology which something related with research methodology, but the level of understanding is not enough for them to apply in their daily work as a teacher.	as practitioners, unless if the teachers have been exposed and experienced sufficiently about scientific methodology. The item will be an advantage for teachers with Master or PhD qualification as they have already undergone research process thoroughly. To be fair, it suggested the statement to rephrased by using common terminology but still capture the same meaning about teachers' professional identity.

Those six Mathematics teachers are
  i. One novice Mathematics teacher
  ii. One non novice Mathematics teacher
  iii. One female Mathematics teacher
  iv. One male Mathematics teacher
  v. One Mathematics teacher with Bachelor degree qualification
  vi. One Mathematics teacher with Master degree of qualification

Once those teachers have returned the form to give feedback on the suitability of the items and some issues related with the items, one expert in Teacher Education review those feedbacks before giving suggestions on how those items can be improved (Table 1). The aim of this process to produce the revised version of item for Q01, Q06, Q12, Q25, Q34 and Q37. After the DIF items have been improved based on the feedback of six Mathematics teachers and one expert in the area of Teacher Education, the improved version of the items has been given to 13 Mathematics teachers for the final confirmation that the items are fair to be used to teachers with different background in terms of gender, teaching qualifications and experiences.

Those 13 Mathematics teachers are asked to fill a form to rate whether they agree that the items do not have any bias to be answered by Mathematics teachers with different backgrounds in term of gender, teaching qualifications experiences. Subsequently, the rating data has been analyzed using Fuzzy Delphi Technique. The Mathematics teachers are labelled from 1 to 13 in Table 1 based on the calculation of the returned questionnaires. The acquisition of Arab Mathematics teachers' consensus is analyzed using Fuzzy Delphi method after all the items in the questionnaire are administered to the Mathematics teachers. The Mathematics teachers are required to give feedback in the specific form which contains choices of scales such as Strongly Disagree, Disagree, Less Agree, Agree and Strongly Agree. The items

which are need to be validated are asked to be rated by the Arab Mathematics teachers whether those items are fair to be asked to Arab Mathematics teachers with different background in term of gender, teaching qualifications and experiences.

The result of the analysis indicated that all three requirements are fulfilled where the threshold value (d) is less than 0.2, the percentage of the consensus among the Arab Math Teachers are more than 75%, and the value of defuzzification which is the average of Fuzzy number (A) more than alpha cut off point which is 0.5. Therefore, it can be concluded that from the perspective of face validity, the issue of DIF for Q01, Q06, Q12, Q25, Q34 and Q37 have been resolved. Therefore, the problematic DIF items of Q01, Q06, Q12, Q25, Q34 and Q37 have been replaced by the improved items which have been validated by using the Fuzzy Delphi method. The next step is taken by administering the revised version of Mathematics TPI questionnaire to 407 Arab Mathematics teachers to examine whether DIF still exist in the revised version of Mathematics TPI questionnaire.Although this study focusses on the quantitative findings for the psychometric properties of Mathematics version TPI questionnaire, nevertheless, a validation process has been undertaken at second stage of this study. The Delphi study in the second stage use Mathematics teachers with different background which related with DIF issues. The finding of Delphi study at the second stage of the study indicate better consensus level, which cause only one round of Delphi study has been conducted. Due to the reason that only one round of Delphi study is done at the second stage, a Fuzzy Delphi technique is chosen to examine the consensus among Mathematics teachers about the revised items of Mathematics TPI questionnaire concerning the DIF issues. By using Fuzzy Delphi technique, the consensus scores which are given by the experts using five Likert scale are changed to Fuzzy scale which are in form of the Fuzzy triangular numbering.

Panel	Item					
	Q01	Q06	Q12	Q25	Q34	Q37
1	0.16	0.47	0.19	0.19	0.52	0.45
2	0.16	0.16	0.49	0.19	0.09	0.14
3	0.14	0.14	0.12	0.12	0.09	0.16
4	0.14	0.16	0.12	0.12	0.09	0.16
5	0.16	0.16	0.19	0.12	0.09	0.16
6	0.14	0.14	0.12	0.12	0.09	0.14
7	0.14	0.14	0.12	0.19	0.09	0.16
8	0.47	0.14	0.12	0.19	0.09	0.16
9	0.14	0.14	0.12	0.12	0.09	0.14
10	0.14	0.14	0.12	0.12	0.09	0.16
11	0.16	0.16	0.19	0.19	0.21	0.14
12	0.14	0.14	0.12	0.12	0.21	0.14
13	0.14	0.14	0.12	0.12	0.09	0.16
Threshold value (d)	0.17	0.17	0.16	0.14	0.14	0.18
Percentage of consensus	92%	92%	92%	100%	92%	92%
Average Fuzzy value (A)	0.71	0.71	0.72	0.74	0.74	0.62

Table 2:The Calculation of Defuzzication

Although this study focusses on the quantitative findings for the psychometric properties of Mathematics version TPI questionnaire, nevertheless, a validation process has been undertaken at second stage of this study. The Delphi study in the second stage use Mathematics teachers with different background which related with DIF issues. The finding of Delphi study at the second stage of the study indicate better consensus level, which cause only one round of Delphi study has been conducted. Due to the reason that only one round of Delphi study is done at the second stage, a Fuzzy Delphi technique is chosen to examine the consensus among Mathematics teachers about the revised items of Mathematics TPI questionnaire concerning the DIF issues. By using Fuzzy Delphi technique, the consensus scores which are given by the experts using five Likert scale are changed to Fuzzy scale which are in form of the Fuzzy triangular numbering.

(iii) The Psychometric Properties for the Revised Version of Mathematics TPI Questionnaire

The revised version of Mathematics TPI questionnaire had been administered to Mathematics teachers in the Kingdom of Saudi Arabia for assessing its psychometric questionnaires. 407 Math teachers had answered the questionnaire and the responses were analyzed using Rasch model analysis technique. The assessment of the psychometric properties for the revised version of Mathematics TPI questionnaire covered seven aspects, namely: fit analysis, person reliability, item reliability, dimensionality, local item dependency, category structure analysis, and DIF analysis

The findings showed that the person reliability index of an individual is 0.90. The findings of the analysis show that the individual separation index is 3.20 which indicate that there are approximately three levels of respondent ability in the Mathematics TPI questionnaire. The analysis result shows that the item reliability index is 0.93. The findings show the item separation index is 3.63 which indicates that there are approximately four levels of item difficulty in the Mathematics TPI questionnaire.

The PCA of the standardized residuals shows that the variance explained by measures was 61.2%. This indicated the unidimensionality improved when Mathematics TPI questionnaire has been improved compared to previous result of unidimensionality. Moreover, the unexplained variance of the eigenvalue for the first contrast (size) is less 3.0 and unexplained variance explained by first contrast is less 5% which considered by Linacre as good indicators for unidimensional instrument (Linacre, 2006).

Subsequently, local dependency is examined based on the residual correlation between pairs of items. This step is important to ensure all items in Mathematics TPI questionnaire is local independence. Local independence is "the probability of a test taker responding correctly to a certain item is not dependent on previous responses or the answers given by other individuals to the same item" (Wale, 2013). Local independence is an assumption in Rasch model. It requires that the items in Mathematics TPI questionnaire should be independent of each other. To be specific, local independence means an item's to be agreed or disagreed should not be related to another item's to be agreed or disagreed under effects of any latent trait that is not the intended latent trait to be measured. The items should only be correlated through the latent trait that the test

is measuring (Lord &Novick, 1968). Violation of local independence assumption will bring imprecise results. As a consequence, the instrument will not be able to distinguish between the effects from target variable and which from unintended variable(s). In this study, the results of largest residual correlation did not indicate any value that exceed 0.7 According to Linacre (2012), locally dependency items are not an issue if the largest residual correlation is not exceeding 0.7

The findings of fit analysis show that the item's mean squared infit is between 0.79 to 1.21 and the item mean squared outfit is between 0.79 to 1.31. The result of polarity analysis indicated all the items in the instrument showed high positive PTMEA Corr. Thus, it is concluded that 45 items in the revised version of Mathematics TPI questionnaire fit with Rasch measurement model framework.

As previous research indicated that the steps difficulties advances are not surpass the cut-off point of 1.4 logit, therefore, the previous five points Likert scale has been changed to three points Likert scale as illustrated at Figure 1. The label of three point of Likert scale is based on the continuum of Disagree to Agree as suggested by Sauro and Lewis (2016).
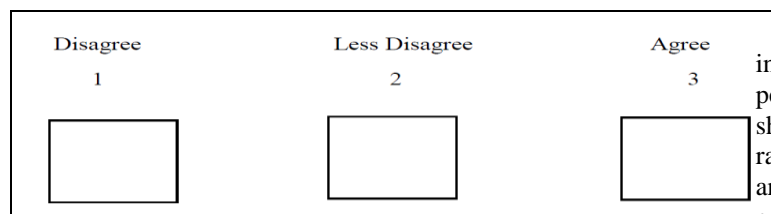

Figure 1: Three Point Likert Scale

The findings show the category mean squared infit is between 0.96 to 1.02 and the category mean squared outfit is in the range of 0.92 to 1.01. The category measures also displayed a pattern that fits the Rasch model assumption of sequencing from easy to difficult. For the Category Measure, the measurement functioning is as expected (i.e. the higher the response scale value, the more positive the category measure value). In this study, based on Linacre (2011), for a three-category scale, the advance must be at least 1.4 logits between step calibrations in order for the scale to be equivalent to two dichotomies. This can be observed based on the step difficulties advances from Category 1 to Category 2 is 2.22, and from Category 2 to Category 3 is 2.22. This showed that the three response scales for each item in the revised version of Mathematics TPI questionnaire are functioning well.

The purpose of conducting the analysis on DIF is to determine whether the items in revised version of Mathematics TPI questionnaire are functioning fair or bias when administered to sample based on the demographic factors. For DIF analysis based on teaching experiences, the findings showed that none of items which had t value more than 2 and p value than is less than 0.5. Specifically, after the modification have been made to the content of Q12 and

Q25, those two items show improvement as there are no evidence of t value more than two and p value less than 0.05.

For DIF analysis based on gender, the findings showed that none of the items had t value more than 2 and p value than is less than 0.5. Previous result showed that Q01, Q06 and Q37 have an issue of differential item functioning. However, after the modification have been made to the content of Q01, Q06 and Q37, those items showed some improvement as there are no evidences of t values more than two and p values less than 0.05.

For DIF analysis based on qualifications, the findings showed that none of the items had t value more than 2 and p value than is less than 0.5. Previous result showed that Q34 has an issue of differential item functioning. However, after the modification have been made to the content of Q34, it showed an improvement as there is no evidence of t value more than two and p value less than 0.05. Therefore, it can be concluded that the items in Mathematics TPI questionnaire has been proven not have any differential item functioning issue-based teaching experiences, gender and qualification.

Discussions

The person reliability index and the item reliability index are indicators of the overall stability of the resulting person and item hierarchies. The findings in this study showed the all the item have infit and outfit values is in range between the value of 0.6 to 1.4. Therefore, the infit and outfit mean squares were within acceptable range for all categories, that is, did not exceed 1.5 as suggested by previous research conducted by Linacre (2012). The guideline from Linacre (2012) was used to assess item fit in the current study to measure in the context of Mathematics in Kingdom of Saudi Arabia. It was suggested by Linacre (2012) that a fit statistic greater than 2.0 indicates off variable noise that is greater than useful information. Such an item would degrade the instrument. Fit statistics between 1.5 and 2.0 indicates noise that neither constructs nor degrades measurement. Linacre (2012) interprets fit statistics between 0.5 and 1.5 as an indicator that the item is productive for measurement. - Linacre (2012) suggested that infit or outfit mean squares below 0.5 indicates overly predictable item. Items that underfit the model might be measuring a construct other than the construct of interest. A very similar research to study fit analysis were conducted by previous researches, such as, Wilczenski (1995), and Chan, Ismail and Sumintono (2014).

The person reliability index can be defined as the extent to which the sample respondents would fall in the same order if given a different measure of recovery. The person reliability index is comparable to CTT reliability coefficient Cronbach's alpha. The findings in this study showed the person reliability of the Mathematics TPI is high. Similar to CTT, in Rasch analysis an analogous reliability coefficient was examined, the person reliability

coefficient. Unlike Cronbach's alpha that is widely used in CTT, the person reliability coefficient in Rasch analysis does not include extreme or perfect scores in the computation. For Mathematics TPI the person reliability of separation statistic was 0.90, based on the acceptability guidelines for reliability in the field as stated in previous research, such as, DeVellis (2016), Engelhard (2013) , George and Mallery (2003), Kline (2000), Kuzniak, Rabbani, Heo, Ruiz-Menjivar, and Grable (2015), Saad, Carter, Rothenberg, and Israelson (1999), the person reliability estimates were considered high. The Rasch item separation statistic for the entire instrument was used to determine the level of distinction possible among items and persons. Separation is the ratio of the square root of the variance explained by the measurement model to that of the unexplained variance or measurement error, including error from model misfit. If person separation is smaller than 2 in a relevant sample and person reliability is smaller than 0.80, the measure may not be sensitive enough to distinguish between high and low performers (Bond & Fox, 2007). Several researchers were fined and explain person reliability in this way. For example, Mutalib, Ghafar, Baharom, and Hamzah (2015) used Rasch model to analysis the implementation effectiveness of final year research project course and found the person reliability of 0.89. They stated that the value of person reliability 0.89 values is characterised as "good" while for item reliability of 0.71 value is ranged as "fair" as suggested in research by Mutalib et al. (2015). Another research, which is Ozbey (2012) reported that the person reliability of the mental health and recovery measure is 0.85.

The item reliability index indicates the extent to which the hierarchy of items would remain stable if the same items were administered to another sample of equal size. When these indices are below acceptable values, common explanations are that either the measure is lacking breadth in item content or that the sample of respondents did not include individuals falling within the full range of recovery from mental illness. Item reliability is unique to Rasch analysis. These Rasch reliability indices are on a scale from 0 to 1 as stated in Bond and Fox (2007). The closer the reliability index is to 1, the greater the stability of the person and item hierarchies. The findings in this study showed the item reliability of the Mathematics TPI in the context of Kingdom of Saudi Arabia showed as high. Item reliability is unique to Rasch analysis. These Rasch reliability indices are on a scale from 0 to 1 (Bond & Fox, 2007). In this study the item reliability was found to be 0.93. This finding is almost similar to previous research like Ozbey (2012). In Ozbey (2012), it studied 1116 individuals with severe and persistent mental illness who attended the introductory session program across Ohio, which showed the item reliability of 0.95.

The findings in this study showed the questionnaires have a good unidimensionality value through the Principal Component Analysis of Rasch Model Analysis result. The unidimensionality requirement is satisfied when the data fit the model. This is often interpreted by practitioners of Rasch measurement to mean that the assumption of unidimensionality is met when the fit values for items and/or persons do not depart significantly from their expected values. The choice between the principal component and other factor analytic procedures for use in this study is more difficult. An almost similar research conducted to measurement tool unidimensionality were done by previous research, such as, Hagell (2014), Smith (1996), Balasubramanian, Li, Bowden, Duncan, Kautz, and Velozo (2016), Yu, Osborn-Popp, DiGangi, and Jannasch-Pennell (2007) and Luthans, Avolio, Avey, and Norman (2007).

Hattie (1985) separated principal component indexes from other factor analysis indexes for several reasons, including the fact that factor analysis requires a hypothesis as to the number of factors. It is for this reason that principal component analysis without rotation was chosen for this study. The findings in this study showed the all the items have high positive point measure correlation. The positive point measure correlation for all the items falls on the satisfactory range where it should be $0.4 < x < 0.8$ as suggested by Said (2016). The point measure correlation indicates how each particular item contributes to the item difficulty. The findings showed that confirmed the well-functioning of items with respect to the directionality of the measure, and strengthened the evidence for unidimensionality for the item data. A similar research researching on principal component indexes were conducted by previous research which in line with Said (2016) suggestion are the Gomes and Nobre, (2012) and the Ghassamia, Asghari, Shaeiri and Safarinejad (2013). A research conducted by Akram, Hussein, Ahmad, Mamat, and Ismail (2015) had showed that all PTMEA Corr were in positive values, where an item was able to distinguish between the ability of respondents and it stated that ahigh PTMEA Corr means that an item is able to distinguish between the ability of respondents.

The findings in this study showed the questionnaires have a good Likert Scale category structure property based on Rasch Model Analysis result. That means the response scales for each item in the questionnaire are functioning well. Response Scales refer to Strongly Disagree, Disagree, Neutral, Strongly Agree and Agree. Finally, the findings in this study showed the all the items in the questionnaire do not have DIF issues. A very similar researches were conducted previously to check the functionality of questionnaire scales, such as, Retief, Potgieter, and Lutz (2013) and Van Zile-Tamsen (2017). In Retief et al. (2013), Rasch model is also used to response to quality issues that are widely recognized as problematic in the refinement of Likert scale questionnaires as the current study. This previous research also used questionnaire under development for the measurement of perceptions of chemistry students on demonstrator effectiveness to illustrate the process of Rasch analysis and instrument refinement. In Van Zile (2017), the use of questionnaires and rating scales to measure important outcomes in higher education is pervasive. This paper outlines a six-step process for using Rasch Analysis to review the psychometric

properties of a rating scale compared to current study with five steps.

DIF were conducted on teaching experiences, gender and qualification among Mathematic Teachers who were involved in this study. It can be concluded that the items in Mathematics TPI questionnaire has been proven not have any DIF issue based on category structure. A similar research was conducted of DIF based on category of experiences which shows the same results were such as Roth, Oliveri, Sandilands, Lyons-Thomas, and Ercikan (2013), Li, Qin, and Lei (2017) and Ramli, Talib, Hassan, and Manaf (2018). From this finding of this study, it also showed that all items in Mathematics TPI questionnaire has been proven do not have any DIF issue based on gender. Previous researches, such as, Fleishman, Spector, and Altman (2002), Smith, and Reise (1998), and Lane (1995)

were conducted to study DIF based on gender issues which had showed a similar outcome. In this study, DIF was also studied based on qualification of the Mathematics teachers who were involved in the study. It has also showed that all items in the Mathematics TPI questionnaire has been proven not have any DIF issue based on qualification too after the improvement were done on the items.

Conclusion

In conclusion, the findings of this study indicated that the Mathematics TPI questionnaire has been proven to be valid and reliable in the Kingdom of Saudi Arabia context from the perspective of Rasch measurement model framework as showed in Table 3.

Table 3
Summary of Results for Rasch Model Analysis

Psychometric Properties based on Rasch Model Analysis	First Version of Mathematics TPI Questionnaire	Revised Version of Mathematics TPI Questionnaire
Person Reliability	High Person Reliability	High Person Reliability
Item Reliability	High Item Reliability	High Item Reliability
Fit Analysis	There are no misfit items	There are no misfit items Additional fit analysis is done which is person fit analysis to elevate the value of dimensionality of the TPI questionnaire, and to enhance the validity of the measurement of Mathematics TPI through profiling based on gender, qualification and experience There are misfit persons, therefore, the misfit persons have been discarded from Rasch model analysis
Category Structure for Likert Scale Category	The findings show the category mean squared infit and outfit displayed a pattern that fits the Rasch model assumption of sequencing from easy to difficult. The values of step difficulties advances did not surpass 1.0 logits which contradicted with Linacre (2011) statement that for a five-category rating scale, advances of at least 1.0 logits between step calibrations are needed in order for that scale to be equivalent to four dichotomies.	The findings show the category mean squared infit and outfit displayed a pattern that fits the Rasch model assumption of sequencing from easy to difficult. The values of step difficulties advances surpass 1.4 logits which contradicted with Linacre (2011) statement that for a three-category rating scale, advances of at least 1.4 logits between step calibrations are needed in order for that scale to be equivalent to two dichotomies.

DIF	DIF analysis is done to 45 items in the first version of TPI questionnaire There are six items which have DIF issues, Those DIF items are Q34, Q12, Q01, Q06, Q25 and Q37. Q34 is a DIF item based on gender. Q12 is a DIF item based on teaching experiences. Q01, Q06, Q25 and Q37 are DIF items-based level of qualification. Other remaining items (39 items) found to be not to have any DIF issue.	Q34, Q12, Q01, Q06, Q25 and Q37 have been revised to resolve to DIF issues in term of gender, teaching experiences and qualifications. The improvement has been made by getting feedback from teachers for the purpose of face validity, and subsequently from the content expert for the content validity. Fuzzy Delphi technique has been done to ensure consensus can be achieved that the items are not bias to certain group. The revised version of Mathematics TPI questionnaire has been administered again. DIF analysis is done on the items. It is found no gender issue for all 45 items in revised version of Mathematics TPI questionnaire.
Dimensionality	Raw variance explained by measures did not surpass 60% cut off value	Raw variance explained by measures surpassed 60% cut off value which indicate a high value of unidimensionality Additional output of dimensionality analysis is Local Dependency output The result showed that all items in Mathematics TPI questionnaire is local independence.

References

[1] Abdulghani, N., & Abulnaeem, M. (2016). Professional identity and its relation with the professional satisfaction and self-esteem: Review paper. Journal of School of Education, Kafrelsheikh University, 2(1), 346-416.

[2] Abdulwahab, M. M. M. (2010). Using item response models in measuring items of some knowledge measurments (Doctor of Philosophy), Minia University, Minia.

[3] Akram, W., Hussein, M. S., Ahmad, S., Mamat, M. N., & Ismail, N. E. (2015). Validation of the knowledge, attitude and perceived practice of asthma instrument among community pharmacists using Rasch analysis. Saudi Pharmaceutical Journal, 23(5), 499-503.

[4] Albaqi'i, N. (2014). Professional identity of a sample of unrwa teachers in jordan. Al-Manara Magazine, 20(2/b), 363-387.

[5] Ali, N. (2012). The effectiveness of using rasch model in constructing a criterion reference test in course of measurement and evaluation in education. (Doctor of Philosophy), Damascus university, Damascus.

[6] AL-Nabhan, M. (2004). Measurement fundamentals in the behavioral sciences. Amman: Dar Alshorouq for Publication and Distribution.

[7] AL-Zayla'I, M. (2014). Constructing a diagnostic criterion-referenced test for the measuring of scientific research skills of the graduate students in the collage of education at al-baha university. (Master), UMM AL- QURA University, Makkah.

[8] Alzubaidi, A. S., & Al-Kahali, S. N. (2014). Gender, grade and anxiety differences in career identity statuses among ninth and tenth grades students in the sultanate of oman. American Arabic Academy of Science and Technology 5(13), 31-44. doi:10.12816/0015407.

[9] Ball, D. L. (2003). What mathematical knowledge is needed for teaching Mathematics. Paper presented at the Secretary's Summit on Mathematics, US Department of Education, Washington, D.C.

[10] Balasubramanian, C. K., Li, C. Y., Bowden, M. G., Duncan, P. W., Kautz, S. A., &Velozo, C. A. (2016). Dimensionality and item-difficulty hierarchy of the lower extremity Fugl-Meyer Assessment in individuals with subacute and chronic stroke. Archives of physical medicine and rehabilitation, 97(4), 582-589.

[11] Beijaard, D., Meijer, P. C., & Verloop, N. (2004). Reconsidering research on TPI. Teaching and Teacher Education, 20(2), 107-128.

[12] Bolívar, A., & Domingo, J. (2006). The professional identity of secondary school teachers in spain: Crisis and reconstruction. Theory and

Research in Education, 4(3), 339-355. doi:https://doi.org/10.1177/1477878506069105.

[13] Bond, T. G., & Fox, C. M. (2007). Applying the rasch model: Fundamental measurement in the human sciences. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

[14] Bond, T. G., & Fox, C. M. (2015). Applying the rasch model fundamental measurement in the human sciences (3rd ed.). New York: Routledge.

[15] Boone, W. J., Staver, J. R., & Yale, M. S. (2014). Rasch Analysis in the Human Sciences. Dordrecht, Netherlands: Springer.

[16] Chan, S. W., Ismail, Z., &Sumintono, B. (2014). A Rasch model analysis on secondary students' statistical reasoning ability in descriptive statistics. Procedia-Social and Behavioral Sciences, 129, 133-139.

[17] Coldron, J., & Smith, R. (1999). Active location in teachers' construction of their professional identities. Journal of Curriculum Studies, 31(6), 711-726.

[18] Dabous, M. (2016). Using item response theory in constructing the item pool in criterion-referenced testing with dichotomous and polytomous items according to two – parameter logistic model. Journal of An-Najah National University' Human Researches, 30(7), 1453-1480.

[19] DeVellis, R. F. (2016). Scale development: Theory and applications (Vol. 26). Sage publications.

[20] Engelhard Jr, G. (2013). Invariant measurement: Using Rasch models in the social, behavioral, and health sciences. Routledge.

[21] Erikson, E. (1968). Identity: Youth and crisis. New York: Norton.

[22] Fleishman, J. A., Spector, W. D., & Altman, B. M. (2002). Impact of differential item functioning on age and gender differences in functional disability. The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 57(5), S275-S284.

[23] George, D., &Mallery, M. (2003). Using SPSS for Windows step by step: a simple guide and reference.

[24] Ghassamia, M., Asghari, A., Shaeiri, M. R., &Safarinejad, M. R. (2013). Validation of psychometric properties of the Persian version of the Female Sexual Function Index. Urology Journal, 10(2), 878-885.

[25] Goldin, G. A., Hannula, M. S., Heyd-Metzuyanim, E., Jansen, A., Kaasila, R., Lutovac, S., . . . Pantziara, M. (2016). Attitudes, beliefs, motivation and identity in Mathematics education: An overview of the field and future directions. Hamburg: Springer Open.

[26] Gomes, A. L. Q., &Nobre, P. (2012). The International Index of Erectile Function (IIEF-15):

Psychometric properties of the Portuguese version. The Journal of Sexual Medicine, 9(1), 180-187.

[27] Graven, M. (2004). Investigating Mathematics teacher learning within an in-service community of practice: The centrality of confidence. Educational Studies in Mathematics, 57(2), 177-211.

[28] Graven, M., & Lerman, S. (2014). Mathematics teacher identity. In S. Lerman (Ed.), Encyclopedia of Mathematics education (pp. 434-438). London: Springer, Dordrecht.

[29] Hagell, P. (2014). Testing rating scale unidimensionality using the principal component analysis (PCA)/t-test protocol with the Rasch model: the primacy of theory over statistics. Open Journal of Statistics, 4(6), 456-465.

[30] Hamadneh, I. M. (2015). The impact of escape alternative position change in multiple-choice test on the psychometric properties of a test and its items parameters. Journal of Education and Practice, 6(9), 23-33.

[31] Hattie, J. (1985). Methodology review: assessing unidimensionality of tests and ltenls. Applied psychological measurement, 9(2), 139-164.

[32] Hussein, H. B. (2017). Investigating the professional identity levels of elementary school Mathematics teachers. Journal of Huaman and Educational Sciences, Babylon University, 32, 84-109.

[33] Ithawi, M. M. S. (2009). Using rasch model according to latent traits theory in building achievement test in a course of psychology of individual differences. (Master), College of Education-Ibn Rushd, Baghdad.

[34] Kelchtermans, G. (2009). Who i am in how i teach is the message: Self-understanding, vulnerability and reflection. Teachers and Teaching: Theory and Practice, 15(2), 257-272.

[35] Kline, P. (2000). The Handbook of Psychological Testing. London: Routlege. Knoef, M., de Vos, K. (in press) The representativeness of LISS, an online probability panel.

[36] Kuzniak, S., Rabbani, A., Heo, W., Ruiz-Menjivar, J., & Grable, J. E. (2015). The Grable and Lytton risk-tolerance scale: A 15-year retrospective. Financial Services Review, 24(2).

[37] Lane, S. (1995). Gender-Related Differential Item Functioning on a Middle-School Mathematics Performance Assessment.

[38] Li, H., Qin, Q., & Lei, P. W. (2017). An examination of the instructional sensitivity of the TIMSS math items: A hierarchical differential item functioning approach. Educational Assessment, 22(1), 1-17.

[39]   Linacre, J. M. (2002). Optimizing Rating Scale Category Effectiveness. Journal of Applied Measurement, 3 (1), 85-106

[40]   Linacre, J. M. (2006). Data variance explained by measures. Rasch Measurement Transactions, 20, 1045–1047.

[41]   Linacre, J. M. (2011). Winsteps® Rasch measurement computer program user's guide. Beaverton, Oregon: Winsteps.com

[42]   Linacre, J. M. (2012). A user's guide to WINSTEPS: Rasch Model Computer Programs. Chicago: MESA Press.

[43]   Lord, F. M. &Novick, M. R. (1968). Statistical theories of mental test scores. Reading, Mass: Addison-Wesley.

[44]   Luthans, F., Avolio, B. J., Avey, J. B., & Norman, S. M. (2007). Positive psychological capital: Measurement and relationship with performance and satisfaction. Personnel psychology, 60(3), 541-572.

[45]   Mutalib, A. A., Ghafar, H. A., Baharom, S., & Hamzah, N. (2015). Rasch model analysis of implementation effectiveness for final year research project course in civil & structural engineering department, ukm. Journal of Engineering Science and Technology, 10, 11-22.

[46]   Ozbey, G. T. (2012). A rasch analysis of the mental health and recovery measure: Reliability and validity. (Doctoral Thesis), University of Toledo, USA.

[47]   Panayides P, Robinson C, Tymms P (2010). The assessment revolution that has passed England by: Rasch measurement. British Educational Research Journal, 36, 611–626.

[48]   Ramli, N. F., Talib, O., Hassan, S. A., &Manaf, U. K. A. (2018). Rasch Analysis and Differential Item Functioning of STEM Teachers' Instructional Preparedness Instrument for Urban and Rural Teachers. International Journal of Academic Research in Progressive Education and Development, 7(4), 211-222.

[49]   Reeves, T. D., &Marbach-Ad, G. (2016). Contemporary test validity in theory and practice: A primer for discipline-based education researchers. CBE—Life Sciences Education, 15(1), rm1.

[50]   Retief, L., Potgieter, M., & Lutz, M. (2013). The usefulness of the rasch model for the refinement of likert scale questionnaires. African Journal of Research in Mathematics, Science and Technology Education, 17(1-2), 126-138.

[51]   Roth, W. M., Oliveri, M. E., Sandilands, D. D., Lyons-Thomas, J., &Ercikan, K. (2013). Investigating linguistic sources of differential item functioning using expert think-aloud protocols in science achievement tests. International Journal of Science Education, 35(4), 546-576.

[52]   Saad, S., Carter, G. W., Rothenberg, M., &Israelson, E. (1999). Testing and Assessment: An Employer's Guide to Good Practices.

[53]   Sabah, A. W. (2017). Using item response theory for constructing a criterion- referenced test in english language (1) for al-quds open university students according to rasch model. Journal of Al-Quds Open University' Humanity and Educational Researches 6(20), 133-151.

[54]   Said, R. F. M. (2016). Application of Rasch measurement model in evaluating student performance for Foundation of Computing II. In 7th International Conference on University Learning and Teaching (InCULT 2014) Proceedings (pp. 251-259). Springer, Singapore.

[55]   Sauro, J. & Lewis, J. R (2016). Quantifying the user experience: Practical Statistics for User Research (2nd ed.). Cambridge: Morgan Kaufmann.

[56]   Schepens, A., Aelterman, A., & Vlerick, P. (2009). Student TPI formation: Between being born as a teacher and becoming one. Educational Studies, 35(4), 361-378.

[57]   Smith, R. M. (1996). A comparison of methods for determining dimensionality in Rasch measurement. Structural Equation Modeling: A Multidisciplinary Journal, 3(1), 25-40.

[58]   Smith, L. L., &Reise, S. P. (1998). Gender differences on negative affectivity: An IRT study of differential item functioning on the Multidimensional Personality Questionnaire Stress Reaction scale. Journal of Personality and Social Psychology, 75(5), 1350.

[59]   Stone M. H. & Wright B.D. (1994) Maximizing rating scale information. Rasch Measurement Transactions, 8, 3, 386.

[60]   Van Zile-Tamsen, C. (2017). Using Rasch analysis to inform rating scale development. Research in Higher Education, 58(8), 922-933.

[61]   Wale, C. M. (2013). Evaluation of the effect of a digital Mathematics game on academic achievement (Doctoral dissertation), University of Northern Colorado.

[62]   .

[63]   Zhu, W., Updyke, W.F. & Lewandowski C. (1997). Post-Hoc Rasch analysis of optimal categorization of an ordered response scale. Journal of Outcome Measurement, 1(4), 286-304.

[64]   Živković, P. (2013). Professional development and teachers professional identity: Self-assessment in republic of serbia. Journal of Educational and Instructional Studies in the World, 3(1), 150-158.

Appendix 1

No	Revised Item
Q1	I have a sufficient knowledge of Mathematics education
Q2	I am working to satisfy the needs of different students in Mathematics.
Q3	I am helping colleagues and teachers in what they need in the technical and administrative aspects of Mathematics.
Q4	I am strengthening the cooperative relations between the staff of teaching Mathematics in school.
Q5	I am reviewing my performance continuously to improve my proficiency in teaching Mathematics.
Q6	I give feedback for the improvement of the school Mathematics curriculum.
Q7	I am considering the needs of students when planning to implement various Mathematics educational activities.
Q8	I am strengthening the school philosophy, vision and educational mission.
Q9	I work on the in-purpose and precise coordination between the different committees in the school.
Q10	I respect the profession of Mathematics education and strive to highlight it as one of the most important professions.
Q11	I provide students with an encouraging and stimulating learning environment for learning Mathematics.
Q12	I organize activities suited to the needs of students in the field of Mathematics.
Q13	I am adopting a perspective that focuses on improving the performance of teachers and students.
Q14	I positively influence my colleagues.
Q15	I accept constructive criticism from others in the aspects of the teaching process.
Q16	I am directing students to apply what they learned, in their real life.
Q17	I am trying to deepest mutual trust between teachers and students.
Q18	I understand the objectives of the school, and I achieve the school general objective
Q19	I have deep understanding about the educational policies, and I follow everything new in that.
Q20	I seek to attend courses and training workshops
Q21	I encourage students to reach the highest levels of learning.
Q22	I master the skills of teamwork that are associated with school goals and philosophy
Q23	I stick to school goals by preparing lessons on a daily basis
Q24	I have understanding about the wrong policies in schoolwork, and I work to address them.
Q25	I follow the latest educational developments.
Q26	I use assessment tools that suit learning outcomes.
Q27	I have a comprehensive knowledge of family factors that influence the student learning process for Mathematics.
Q28	I play the role of the real citizenship and I take responsibility for community issues.
Q29	I follow the success stories of teachers in Mathematics education.
Q30	I link learning topics purposefully.
Q31	I involve parents in making decisions that contribute to the school development process.
Q32	I have a respected professional status in society.
Q33	I am constantly trying to improve my performance.
Q34	I show some flexibility in the educational learning process
Q35	I collaborate and work in team spirit within the school.
Q36	I respect diversity within and outside the school.
Q37	I use sources of information wisely.
Q38	I employ tools and assessment strategies to improve the learning process.
Q39	I devote most of my time and effort to school work during my work time.
Q40	I contribute to solving the problems of colleagues and students that occur within the school.
Q41	I seek to strengthen professional relations with management, colleagues and stakeholders in the field of education to contribute to improving my performance.
Q42	I am constantly trying to find the best teaching methods and strategies for Mathematics.
Q43	I take the opportunity to invite the principal to assess my performance in teaching.
Q44	I see myself as always being renewed.
Q45	I put to myself an ambition for advancement and excellence in the profession of teaching Mathematics.