International Journal of Assessment Tools in Education
2024, Vol. 11, No. 3, 567–588
https://doi.org/10.21449/ijate.1386295
journal homepage: https://dergipark.org.tr/en/pub/ijate
Research Article

Algebraic knowledge for teaching test: An adaptation study

Ali Bozkurt[1*], **Begüm Özmusul**[2]

[1]Gaziantep University, Gaziantep Faculty of Education, Department of Educational Sciences, Gaziantep, Türkiye
[2]Gaziantep University, Nizip Faculty of Education, Department of Educational Sciences, Gaziantep, Türkiye

ARTICLE HISTORY
Received: Nov. 05, 2023
Accepted: July 15, 2024

Keywords:
Algebra teaching,
Algebraic knowledge,
Adaptation study,
Preservice Teachers,
MKT-PFA.

Abstract: In this study, the Mathematical Knowledge for Teaching-Elementary Patterns Functions and Algebra-Content Knowledge (MKT-PFA) test, originally developed in English as part of the "Learning Mathematics for Teaching Project" at Michigan University, was adapted into Turkish. The test comprises two equivalent forms, A and B, each translated into Turkish and culturally adapted through consultations with two mathematics education academics and five secondary school math teachers pursuing doctoral studies. A total of 328 pre-service teachers at a Turkish public university's elementary school mathematics teaching department were administered form A (14 questions, 29 items) and form B (12 questions, 27 items) at a one-week interval. Psychometric analyses revealed high reliability (KR-20: A=0.712, B=0.735; Lord reliability: A=0.733, B=0.756), and strong correlations (r_{pbi}) with the original English forms, indicating suitable adaptation. Item difficulties analyzed using a one-parameter Item Response Theory model showed a normal distribution, affirming the tests' validity for assessing pre-service teachers' algebra teaching knowledge in Türkiye.

1. INTRODUCTION

Mathematics education is a field that requires interaction between teachers and students in classrooms, professional knowledge, and reasoning to invite students to the learning process of mathematics (Ball et al., 2008). This teaching process, which consists of interactions between teachers and students, helps students act as critical thinkers and develop their reasoning (Cohen, 2011). The teacher's interactions with students in the classroom begin and are developed through the "teaching job". This “teaching job” allows students to reason, interpret, criticize textbook practices on specific topics, use representations correctly, and create examples of mathematical concepts, algorithms, or proofs (Hill et al., 2005). Therefore, teachers should possess certain competencies, such as mathematical knowledge for teaching, interactions with students, technology integration, and understanding of student diversity, to structure their mathematics instruction effectively (Ball et al., 2005; Ma, 1999). These competencies can equip mathematics teachers with the essential skills and knowledge required to enhance their students' achievement and foster positive attitudes toward mathematics. This study focuses on

*CONTACT: Ali BOZKURT ✉ alibozkurt@gantep.edu.tr Gaziantep University, Gaziantep Faculty of Education, Department of Educational Sciences, Gaziantep, Türkiye

e-ISSN: 2148-7456

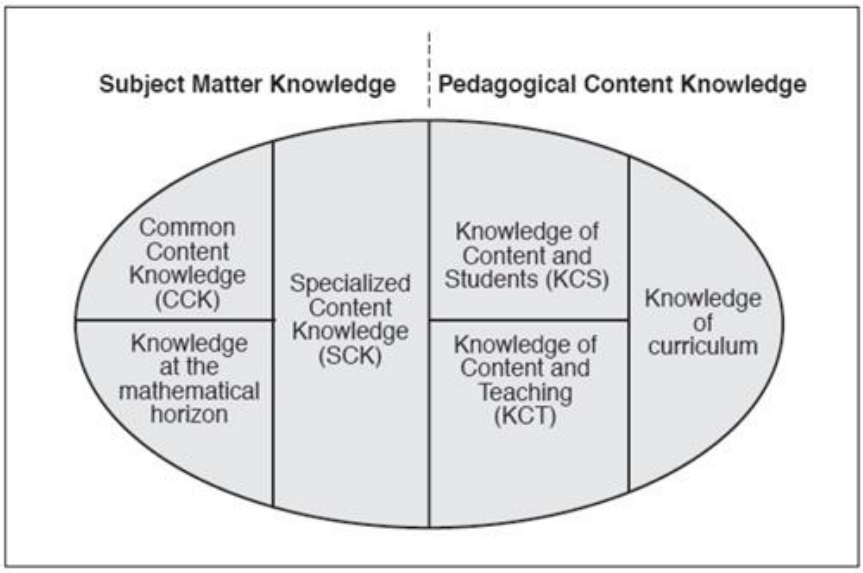
Mathematical Knowledge for Teaching (MKT), one of the competencies identified in the literature, for teachers and preservice teachers to structure the ‘teaching job'.

The theoretical basis of mathematical knowledge for teaching (MKT) in instruction is grounded in the idea that what teachers need to know is determined by what teachers do in their teaching practice (Ball, 1990). Studies in the literature (An et al., 2004; Ma, 1999) draw attention to the quantity and quality of teachers' mathematical knowledge for teaching. Additionally, these studies have identified significant differences in mathematics teaching across different countries. This situation is crucial in understanding how teachers acquire mathematical knowledge and apply it in teaching mathematics in varying countries (Ball & Hill, 2008; Hill et al., 2005). MKT assists us in comparing teaching and learning processes across countries (Knipping, 2003). This study presents an adaptation study of the algebra teaching knowledge test for measuring the algebra knowledge for teaching future teachers.

1.1. Mathematical Knowledge for Teaching (MKT)

The knowledge that a teacher should have in the teaching process can be classified under two main headings: Pedagogical content knowledge and subject matter knowledge (Shulman, 1986). The former is concerned with presenting the knowledge in the relevant field to the student by transforming it into a teachable structure while the latter is the knowledge about the basic principles, concepts, laws, and theories stipulated by the curriculum of the field. The components of these types of information are given in Figure 1.

Figure 1. *Domain of MKT (Ball et al., 2008, p.403; Shulman, 1986).*



In Figure 1, there are three sub-fields under the title of pedagogical content knowledge; namely, "knowledge of content and teaching, knowledge of content and students, and knowledge of content and curriculum" (Ball et al., 2008). Under the title of subject area information, there are three sub-fields: "common content knowledge, specialized content knowledge, and horizon content knowledge". Although this division can be used to analyze subfields, it is intertwined with teaching practices (Kim, 2016; Koellner et al., 2007).

The details of the teaching information vary according to the course. Mathematical knowledge for teaching comes to the fore for the mathematics course (Hill et al., 2005). This information is essential for realizing mathematics teaching (National Mathematics Advisory Panel, 2008). However, the required pedagogical information can vary according to the learning objectives of the mathematics course. One of the learning areas in mathematics is algebra, a field of mathematics that involves developing rules to represent functional expressions and relations, expressing these rules with symbols, writing and solving equations, and making generalizations

from calculations with numbers (Lew, 2004; Welder & Simonsen, 2011). Some researchers (Kieran et al., 2018; Lew, 2004) focus on the abstract features that distinguish it from arithmetic in the definition of algebra and define algebraic thinking as "the ability to operate on an unknown quantity as if the quantity is known, as opposed to arithmetic reasoning involving operations on known quantities" (Langrall & Swafford, 1997, p. 2). Some others (Driscoll, 1999; Zazkis & Liljedahl, 2002) have noted the critical importance of functions that play in algebra, which is characterized as the capacity to represent quantitative situations in algebraic thinking. In both cases, they are part of algebraic thinking which is aimed to be improved in algebra teaching. In the algebra teaching process, students are expected to be taught algebra and gain algebraic thinking skills (Schmittau, 2005). Charalambous (2008) concluded that there is a potent relationship between teacher knowledge and teaching performance. Therefore, the fact that teachers need to know how to teach the basic concepts of algebra may cause students to have difficulty learning algebra. For this reason, Hill and Ball (2009) developed the 'patterns, functions, and algebra' test to measure teachers' and preservice teachers' algebra knowledge for teaching. Within the scope of this study, the adaptation of the test developed by Hill and Ball (2009) was carried out in Turkish. Thus, with the adaptation of this achievement test, the algebra knowledge for teaching preservice teachers can be assessed. Consequently, based on the levels of algebra knowledge for teaching among preservice teachers, mathematics educators can enrich the scope of the algebra teaching course as specified by the Council of Higher Education (CoHE (YÖK: Yükseköğretim Kurumu), 2018). Teachers go through the candidacy process to gain competence in the professional context. Morris et al. (2009) mentioned that preservice teachers could define mathematical concepts but could not spontaneously apply planning or assessment of teaching and learning in line with their learning objectives. Huang and Kulm (2012) indicated that preservice teachers need more knowledge about the place of the term of function in the curriculum, its teaching, and content knowledge. In particular, the study concluded that the flexibility of the preservice teachers in the use of different representations and the weakness in the selection of function perspectives. He also recommended that the teacher training program should provide content areas that are consistent with the curriculum. Strand and Mills (2014) stated that preservice teachers used the "guess and check" strategy while using variables to represent unknown numbers in algebra problems and then writing numbers instead of variables while solving. Thus, preservice teachers are in different thinking processes to confirm their ideas.

1.2. Mathematics Teacher Education Program in Türkiye

The General Competencies for the Teaching Profession, which outline the knowledge, skills, and attitudes necessary for effectively and efficiently fulfilling the teaching profession, were updated in 2017. In addition, the Teacher Strategy Paper was published in 2017. In the mentioned documents, new goals and expectations, as well as new competencies related to teaching, are included. In addition, some official documents such as the 10th Development Plan (2014-2018), the Strategic Plan of the Ministry of National Education (2015-2019), Türkiye's higher education qualifications framework, educational sciences field qualifications, and teacher training have been published over time (CoHE (YÖK), 2018). Considering the developments required in teacher training, as well as the structural changes in the Turkish education system, societal demands, and social needs, the necessity of updating teacher education undergraduate programs has emerged. (CoHE (YÖK), 2018). In this direction, the elementary education mathematics teacher undergraduate program was changed in 2018. While field courses such as algebra, differential equations, and elementary number theory were intense in the curriculum before 2018, since 2018, mathematics education, such as teaching algebra, geometry and measurement, numbers, statistics, and probability has begun to be given more place.

To determine the mathematics teaching knowledge of teachers and preservice teachers, tests are developed specifically for various learning areas. However, using these tests directly to assess the situation in different countries may not yield reliable results. In this context, it is

necessary to adapt the developed measurement tool for each country to be applied, so much so that many studies have revealed that the characteristic features of the teaching systems they examine are influenced by culturally located teaching practices (Delaney et al., 2008; Knipping, 2003; Ma, 1999; Stiegler & Hiebert, 1999; Wilson et al., 2001).

In the field of mathematics education, the Learning Mathematics for Teaching (LMT) project has developed MKT tests to measure teachers' knowledge of mathematics teaching (Ball & Hill, 2008; Hill & Ball, 2004; Hill & Ball, 2009). Adaptation studies of MKT tests developed in the USA to different languages and cultures were carried out. Some of these are the following:

- Delaney et al. (2008) adapted the forms developed for the learning domains of numbers and operations, algebra, and geometry from MKT tests for use in Ireland. They found that some Irish teachers were unsure of the meaning of certain terms and suggested changes to the general cultural context for adaptation.
- Mosvold and Fauskanger (2009) determined that there was a need for significant changes in the process of adapting the form developed for the geometry learning domain from the MKT scales to Norway. For example, it has been observed that some concepts in the scale are not found in the Norwegian curriculum. Some changes have been made to make it more usable, valid, and reliable for Norwegian teachers.
- Ng et al. (2012) found some contextual problems and differences in teaching practices and representations in the process of adapting the form developed for the geometry learning field, one of the MKT scales, to Indonesia.
- Cole (2012) found cultural incompatibility between America and Ghana in the questions in the form developed for the learning domain of numbers and operations from MKT scales.
- Kim (2020) conducted a study on adapting the form developed for the algebra learning field from the MKT scales. In the study, it was determined that Korean teachers had a high rate of answering the MKT test correctly, but the relationship between teaching methods and algebraic reasoning was low.
- Esendemir and Bindak (2019) adapted the geometry teaching knowledge scale, which is a learning area of mathematics, of secondary school mathematics teachers into Turkish.

When the studies on the adaptation of MKT in the literature above are examined, it is seen that there are fewer adaptation studies of forms measuring algebraic knowledge for teaching. Regarding the field of algebra learning, Delaney et al. (2008) observed that Irish teachers examined their algebraic knowledge for teaching and adapted it to their own culture. Similarly, Kim (2020) observed that Korean teachers adapt their algebraic knowledge for teaching to their own culture to measure it. Within the scope of this study, the Mathematical Knowledge for Teaching-Elementary Patterns Functions and Algebra Content Knowledge (MKT-PFA) forms given in Hill and Ball (2009) were adapted to measure the algebra teaching knowledge of pre-service teachers. When the literature is examined, there is no Turkish adaptation of an algebra teaching tool used internationally. In this respect, these achievement tests measuring algebraic knowledge for teaching preservice mathematics teachers in Türkiye are expected to contribute to the literature.

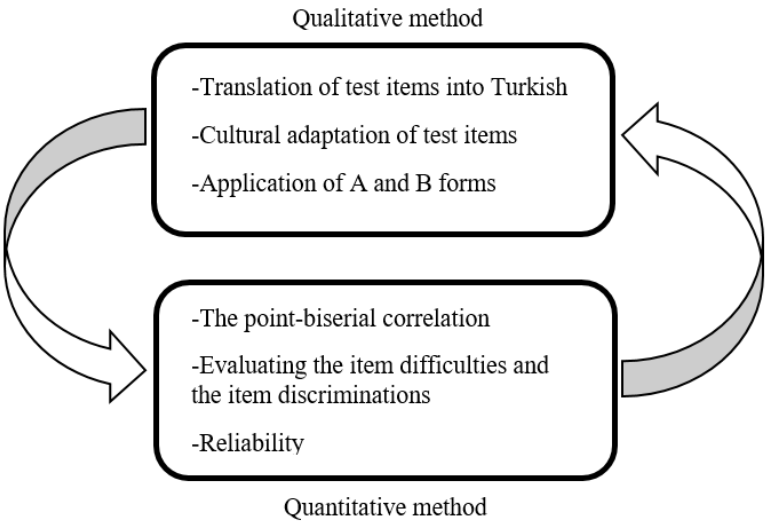
2. METHOD

The study is aimed to adapt the MKT-PFA test which was developed in English by Hill and Ball (2009) within the scope of the "Learning Mathematics for Teaching Project" carried out at the Michigan University, into Turkish. Both qualitative and quantitative methods were used in the adaptation process of the test (see Figure 2).

Qualitatively, to adapt the test items to Turkish, analysis was conducted as a result of interviews with field experts in line with the adaptation framework in the Delaney et al. (2008) study. The adaptation framework proposed by Delaney et al. (2008), which was used in the present study,

was also used in adaptation studies conducted to measure mathematical knowledge for teaching of teachers and preservice teachers in different countries (Esendemir & Bindak, 2019; Kim, 2020; Mosvold & Fauskanger, 2009; Ng, 2012; Ng et al., 2012). In this respect, this framework was considered to be sufficient. Quantitative data were analyzed using psychometric methods.

Figure 2. *The adaptation process of MKT-PFA.*



2.1. Algebraic Knowledge for Teaching Test

With the A and B equivalent forms of the MKT-PFA test, it was aimed to measure the algebraic teaching knowledge of teachers and preservice teachers in the United States. The tests were developed to examine the structure of teachers' and preservice teachers' teaching knowledge (Hill et al., 2004), how teachers learn to teach mathematical knowledge (Hill & Ball, 2004), and how teacher knowledge is related to achievements in students' mathematics achievement (Ball et al., 2005). There are 14 questions in Form A of the MKT-PFA test. In this form, participants are asked to evaluate the options of the 4th, 9th, 14th, 20th, and 22nd questions within the scope of "yes, no and I'm not sure" options. Form B has 12 questions. In this form, participants are asked to evaluate the options of the 6th, 13th, 16th, 19th, and 25th questions within the scope of the "yes, no and I'm not sure" options.

2.2. Adaptation Process of Algebraic Knowledge for Teaching Test

A review of the related literature reveals that, for the adaptation studies of the scales, (i) the test items should be translated into the language to be adapted, (ii) the items should be culturally adapted, (iii) the test should be applied to the relevant sample group, and (iv) validity and reliability studies should be done (Delaney et al., 2008). In this study, these stages were followed within the scope of adapting the items in the A and B forms of the MKT-PFA test to Turkish to determine the algebra teaching knowledge levels of preservice elementary school mathematics teachers in Türkiye.

2.2.1. *Translation of test items into Turkish*

In the first stage of the adaptation studies, the items in the A and B forms of the MKT-PFA test were translated from English to Turkish. An English education expert was consulted during the translation process of the test items. The cultural conformity of the terms has not been taken into account when translating the texts. For this reason, it was assumed that the test items did not undergo any changes in this process. Thus, without changing the mathematical substance of the test items, a one-to-one translation was made from English to Turkish.

2.2.2. *Cultural adaptation of test items*

In adaptation studies, intercalarily to the translation process into another language, the available test should also be culturally adapted. Materials devoid of cultural components may cause participants to focus on another thing (Hambleton, 1994). This distraction may negatively influence the success of the attendees regarding the items (Yen, 1993). In order to determine whether it reflects the situations that would arise in the classrooms in Türkiye, interviews were conducted with a group of 7 participants. Two of the participants were mathematics educators and five were elementary school mathematics teachers with doctoral degrees in mathematics education, and taught algebra and algebra teaching. During the interviews, the items were adapted according to the following four criteria: (i) changes in the general cultural context, (ii) changes in the context of the school culture, (iii) changes in the mathematical structure, and (iv) changes in the language structure (Delaney et. al., 2008). The group discussed the changes to be made to make each element suitable for Turkish culture. Eventually, a final judgment was made for each change.

A critical question appears regarding the adaptability of a test developed in one country to another: To what extent does the test match the algebra knowledge of elementary preservice mathematics teachers in Türkiye, where the test will be adapted? It is thought that the best mathematics educators and experienced mathematics teachers can answer this question. Therefore, at the end of the interviews, this question was asked to the participants as it is, and all participants agreed that each item in the forms was consistent with the content in Türkiye.

2.2.3. *Application of A and B forms of the test to elementary mathematics preservice teachers*

The sample sizes most frequently used in previous IRT studies were reviewed while deciding on the sample sizes to apply the Algebra Teaching Knowledge test within the scope of the study. Kline (1994) recommends a sample size of one-tenth (ten times as many participants as the number of items). On the other hand, research in the literature (Pekmezci & Avşar, 2021; Şahin & Anıl, 2017; Yang, 2007) states that at least 150 samples can be created in tests with a single parameter and the number of items between 20-30. Additionally, Sheng (2013) stated that as the sample size increases, there is no significant change in model-data fit values under the unidimensional theory. Additionally, AIC is commonly used as an information criterion for statistical model selection (Burnham & Anderson, 2002). Moreover, AIC tends to perform better with smaller sample groups (Boykin et al., 2023). Similarly, it was observed that there was no significant change in the model-data fit values of the adapted test after 300 samples (Pekmezci & Avsar, 2021). For this reason, the sample of the test to be adapted consists of preservice mathematics teachers studying at the faculty of education of a state university in Türkiye. It was applied to a total of 328 3rd and 4th-grade preservice mathematics teachers, 217 of whom were female and 111 of whom were male, taking the algebra teaching course.

2.3. Situation of Satisfying Item Response Theory (IRT) Assumptions

In Item Response Theory (IRT), the ability parameter that defines a respondent is not dependent on a group of test items (Holmes & Brian, 2019). Another feature that is valid for all models of IRT is that they must meet certain assumptions of IRT. The necessity of meeting these assumptions varies according to IRT's models (Reyhanlıoğlu & Doğan, 2020). One-dimensional IRT has two commonly accepted assumptions: unidimensionality and local independence (Baker & Kim, 2017; Edelen & Reeve, 2007).

Unidimensionality recognizes that the achievement test has a single latent ability (Reyhanlıoğlu & Doğan, 2020). What is sufficient and necessary for this assumption to be met is that there is a dominant component or factor that is measured by the test items and affects test performance. This dominant constituent or factor (element) is called the ability measured by the test (Crocker & Algina, 1986; Hambleton & Swaminathan, 1985). In addition, when a one-dimensional test is applied to all populations, the conditional distributions obtained from the test results are

expected to be similar (Hambleton & Swaminathan, 1985). Researchers (Aryadoust et al., 2021; Chou & Wang, 2010; Hambleton et al., 1991; Han, 2022) cited many analyses to show that the test is one-dimensional. The main analytical technique is Exploratory Factor Analysis (EFA). Before performing EFA, the KMO (Kaiser-Meyer-Olkin) statistic value was examined. For form A, the KMO value was determined as 0.722 and the Bartlett's sphericity test statistical value was determined as KMO and Bartlett's Test 1655.537 (sd = 406, p<0.05). For form B, the statistical value of Bartlett's sphericity test was determined as 0.716 and KMO and Bartlett's Test was determined as 1597.755 (sd = 378, p<0.05). If the KMO value is greater than 0.60 and the Bartlett test results show a statistically significant difference, it means that the data is suitable for factor analysis (Tabachnick & Fidell, 2012). Considering the KMO value and Bartlett statistics, it can be said that the sample size is suitable for factorization. When the eigenvalues of the factors for form A were examined, 3 factors were seen above 1. However, while the eigenvalue for the first factor (3.020) is almost 3 times the eigenvalue for the second factor (1.099), the eigenvalue for the second factor (1.099) is twice the eigenvalue for the third factor (1.049). When the eigenvalues of the factors for form B were examined, 3 factors above 1 were observed. However, while the eigenvalue for the first factor (3.970) is almost 3 times the eigenvalue for the second factor (1.155), the eigenvalue for the second factor (1.155) is more than the eigenvalue for the third factor (1.132). Lord (1980) states that a single-factor structure may exist in cases where the eigenvalue of the first factor is significantly greater than the second factor and the eigenvalues of the second factor and the third factor are close to each other. Furthermore, when EFA was conducted on both forms, it was observed that the item loadings of the items in the forms were greater than .30. Upon reviewing studies in the literature (Tabachnick & Fidell, 2012), it is seen that this is considered sufficient. It is seen that PCAR (principal component analysis of residuals), one of these analyses, is used by the test developers. One of these analyses, PCAR (Principal Component Analysis of Residuals), appears to be used by test developers. For this reason, PCAR analysis was performed to show that the adaptation of the MKT-PFA test is one-dimensional. PCAR of the adapted test was obtained as 1.2. According to Smith and Miao (1994), since this value is less than 1.4, it indicates that the adapted test may have one-dimensionality. For this reason, the adapted test is one-dimensional. In addition, AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) are frequently used information criteria in statistical model selection (Boykin et. al., 2023). Both AIC and BIC help in model selection by considering the fit and complexity of a model. For a one-parameter model, BIC is lower than AIC because BIC expresses the complexity of the model. In this framework, AIC and BIC values were calculated in both forms. It was observed that the BIC values of both Form A (AIC: 7704.9 BIC: 7484.905) and Form B (AIC: 6442.714 BIC: 6237.891) were lower than the AIC value. Therefore, it can be said that the tests are uniparametric. In addition, the developers of the MKT-PFA test stated that the test is only aimed at the algebra knowledge for teaching of teachers and preservice teachers.

Items are regressed according to the latent variable in the Rasch measurement; for this reason, the fact that unexplained variances in the items are not related to each other is explained by concept of local independence (Borsboom, 2005). Local independence is when individuals' responses to different items of a test are statistically independent or unrelated to each other (Fan & Bond, 2019; Hambleton & Swaminathan, 1985; Wright, 1996; Yen, 1993). However, for the responses to the items to be statistically independent of each other, the ability measured by the test items must be kept constant (Lord & Novick, 1968). Statistics such as Yen's (1993) 3rd quarter are available to provide local independence assumptions. To ensure the assumptions, the answer to a question in the test must not be a clue for the answer to the other question (Borsboom, 2005; Hambleton & Swaminathan, 1985). Reyhanlıoğlu and Doğan (2020) stated that it is sufficient for the measured structure to be one-dimensional to ensure the local

independence assumption. Accordingly, it can be said that the adapted test meets the local independence assumption because it meets the unidimensionality assumption of MKT-PFA.

2.4. Validity and Reliability

Three psychometric analyses are performed to examine the validity and reliability of the adapted version of a test: Comparing the r_{pbi} (rpbi are the correlation coefficients of the items themselves), evaluating the item difficulties and item discrimination, and calculating the reliability of the form (Delaney et al., 2008; Ng, 2012). In this context, the validity and reliability of the adapted test need to be evaluated, entailing a comparison of the r_{pbi} between the USA and Türkiye measurements, evaluation of the item difficulties and item discriminations of the items in the A and B forms of the MKT-PFA test using a one-parameter IRT model, and the evaluation of the MKT-PFA test. The reliability of the A and B forms was calculated. The KR-20 value was calculated within the scope of the reliability of the test.

Point biserial correlation is used to examine how one item relates to all other items (de Ayala, 2013). The higher the point biserial correlation of an item, the stronger the relationship between that item and the measured construct. In other words, the higher the r_{pbi} of an item, the better it can distinguish individuals whose quality under investigation is closer to each other (Delaney et al., 2008; Ng, 2012; Marcinek et al., 2022). In the context of this study, r_{pbi} with high scores indicates that the items can distinguish teachers with closer algebra knowledge.

Negative r_{pbi} values of an item indicate that teachers with higher mathematics teaching knowledge would probably answer this item incorrectly, and the item may not measure the intended construct. Researchers analyzing LMT item properties evaluated all items with negative r_{pbi} values as poorly functioning (Delaney et al., 2008; Esendemir & Bindak, 2019; Fauskanger et al., 2012; Kim, 2020; Kwon et al., 2012; Marcinek & Partová, 2016; Marcinek et al., 2022; Ng, 2012; Ng et al., 2012). In addition, some studies showed scatterplots (Kim, 2020; Kwon et al., 2012), performed a Fisher Z transform on r_{pbi} values to place them on the interval scale (Delaney et al., 2008; Marcinek et al., 2022), identified outliers (Ng, 2012; Ng et al., 2012) and expressed correlations between the r_{pbi} values of items in the USA and those used in their own countries (Delaney et al., 2008; Esendemir & Bindak, 2019; Ng, 2012).

Items with r_{pbi} value of around zero show no relationship between how respondents answered the item and their general mathematics teaching knowledge level. In other words, when we remove such an item from the test, it cannot be said whether the teacher who gave the correct answer was generally more successful than the teacher who gave the wrong answer (Hambleton et al., 1991). Therefore, r_{pbi} predictive was able to examine the difficulty levels and the overall reliability of the items in the context of the relationship between countries. If there is a difference between these items, it means that these items perform differently between cultures (Cronbach & Shavelson, 2004; Delaney et al., 2008).

A one-parameter IRT model was used to calculate the item difficulty values of the test. Depending on the sample size of the data obtained from the pilot study, researchers can use one- or two-parameter IRT models (Delaney et al., 2008; Ng, 2012; Esendemir & Bindak, 2019). When looking at item difficulty, 0 is considered to represent average teacher skill. Items with a difficulty value of less than 0 are considered easier, and items higher than 0 are considered more difficult (Ng, 2012). In addition, the test information curve maximum was generated for each form to examine how useful the measures were. The test information curve provides information on whether the measures were more difficult or less difficult for the average preservice teachers, i.e., whether the measures can discriminate among preservice teachers of different level of abilities (Baker & Kim, 2017).

Finally, after the final version of the PFA test was provided, the reliability of measurements for Form A and Form B, which calculates how consistent respondents' scores are across multiple items or tests, was computed. Test reliability measures the consistency of test takers' scores on

more than one item (Delaney et al., 2008). A widely used reliability measure from classical test theory is the KR-20. KR-20 is reported in the reliability of achievement tests evaluated as 0-1 (Cronbach & Shavelson, 2004). For the reliability of the test, the KR-20 value is expected to be above .70 (Cronbach & Shavelson, 2004). In addition, Lord reliability is included in the reliability of tests graded as 1 and 0 in IRT (Çelen, 2008; Frary, 1989; Özdemir, 2004). Lord reliability is typically calculated based on item parameters obtained from IRT and individuals' responses (Sireci et al., 1991). This measurement is used to assess the internal consistency of a test and indicates the repeatability of an individual's performance on the test. In the literature, it is also expressed as estimated reliability or reliability coefficient (Embretson & Reise, 2013). Estimated reliability is a measure reflecting how accurately a test measures individuals' true abilities. This reliability measure reflects the internal consistency of the test and indicates that the test items measure consistently with each other (Embretson & Reise, 2013). For this reason, Lord reliability, one of the reliabilities of the test's IRT, is also included.

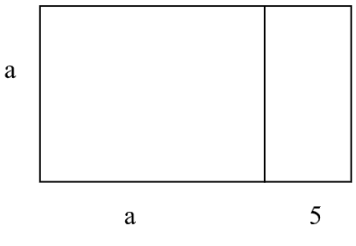
3. RESULTS

In this part of the study, the findings obtained from the cultural adaptation process and psychometric analyses are included to examine the validity and reliability of the Turkish versions of the A and B forms of the MKT-PFA test.

3.1. Cultural Adaptation of Test Items

In the first stage of the adaptation process, the A and B forms of the MKT-PFA test were translated from English to Turkish. In this process, attention was paid to the direct translation of the expressions in the original form. An example of item is given Table 1 in form A of the MKT-PFA test and its translation into Turkish.

Table 1. *An item in form A of the MKT-PFA test and its translation into Turkish.*

Ms. Whitley was surprised when her students wrote many different expressions to represent the area of the figure below. She wanted to make sure that she did not mark as incorrect any that were actually right. For each of the following expressions, decide whether the expression correctly represents or does not correctly represent the area of the figure. (Mark REPRESENT, DOES NOT REPRESENT, or I'M NOT SURE for each.)	Zeynep öğretmenin öğrencilerinin aşağıdaki şeklin alanını temsil etmek için birçok farklı ifadeyi gördüğünde şaşırdı. Zeynep öğretmen gerçekte doğru olanları yanlış olarak işaretlemediğinden emin olmak istedi. Aşağıdaki he bir ifade için verilen şeklin alanının doğru temsil edilip edilmediğine karar verin. (Her bir seçenek için TEMSİL EDER, TEMSİL ETMEZ veya EMİN DEĞİLİM şıklarından birini işaretleyin.)
	

	Correctly represents	Does not correctly represent	I'm not sure		Temsil Eder	Temsil Etmez	Emin Değilim
a) a^2+5	1	2	3	a) a^2+5	1	2	3
b) $(a+5)^2$	1	2	3	b) $(a+5)^2$	1	2	3
c) a^2+5a	1	2	3	c) a^2+5a	1	2	3
d) (a+5)a	1	2	3	d) (a+5)a	1	2	3
e) 2a+5	1	2	3	e) 2a+5	1	2	3
f) 4a+10	1	2	3	f) 4a+10	1	2	3

The sample statements regarding the changes in the general cultural context and the number of changes are given in Table 2 from the interviews with a group of experts in the field, which helps to determine whether the items reflect the situations that will arise in the classrooms in Türkiye after the item selection is translated.

Table 2. *Exemplars of general contextual changes to items and frequency of changes.*

Type of change	Example from original U.S. form	Example from adapted Turkish form	Number of items changed	
			Form A	Form B
People's names	Ms. Ashton	Asya öğretmen	18	20
	Ms. Diaz	Deniz öğretmen		
	Leah	Leyla		
	Earl	Enes		
Non-mathematical language	Baseball cards	Oyuncu kartı	5	6
	Mix contains	Çerez		

In Table 2, it is seen that the changes in the context of general culture are evaluated in the sub-themes of "people's names and non-mathematical language". 18 changes were made in Form A, and 20 changes were made in Form B, which was developed in the context of people's names. There were 5 changes in form A and 6 changes in form B, which was developed in the context of non-mathematical language. Therefore, in this context, a total of 49 changes were made in the context of general culture, 23 changes in A form and 26 changes in B form.

The second stage in the cultural adaptation process includes changes in the context of school culture. Sample statements regarding the changes in this framework and the number of changes are presented in Table 3.

Table 3. *Examples of school contextual changes and frequency of changes in items*

Changes' type	Original form	Adapted Turkish form	frequency of items changed	
			Form A	Form B
School language	Textbook	Ders kitabı	6	8
	Brainstorm	Beyin fırtınası		
	Ms. Hamid's class	Hatice öğretmenin öğrencileri		
Structure of education system	Kyle's method	Zeki'nin çözüm yolu	3	3
	Byron's approach to the problem	Burak'ın problem yaklaşımı		
	Task	Görev		

In Table 3, it is seen that the changes in the context of school culture are evaluated in the sub-themes of "school language and structures of the education system". There were 6 changes in form A and 8 changes in form B, which was developed in the context of school language. In the context structures of the education system, 3 changes were made in the A form and 3 changes in the B form. Therefore, a total of 20 changes were made in the context of school culture, including 9 changes in Form A and 11 changes in Form B.

The third type of change in the adaptation process involves changes in the mathematical structure. Since the changes in this category do not disturb the mathematical structure of the items, the probability of affecting the mathematical knowledge of the test takers is very low. According to this, sample statements about the changes and the number of changes are presented in Table 4.

Table 4. *Exemplars of mathematical changes to items and frequency of changes.*

Type of change	Example from original U.S. form	Example from adapted Turkish form	Number of items changed	
Symbolic notations	50 percent	%50	3	1
	n^{th}	n.		
Mathematical language	Hexagon	Altıgen	29	36
	Doubling its length	Boyunun 2 katı		
	Area	Alan		
Units of measurement	A half	Yarım	5	7
	1 ounce	10 gr		

In Table 4, it is seen that the changes in the mathematical structure are evaluated in the sub-themes of "symbolic notations, mathematical language, and unit of measurement". 3 changes were made to the questions in the A form, and 1 change in the B form, which was developed in the context of the symbolic notations. There were 29 changes in the questions in the A form developed in the context of mathematical language and 36 changes in the B form. 5 changes were made to the questions in the A form developed in the context unit of measurement, and 7 changes were made to the B form. Therefore, a total of 74 changes were made in the context of the mathematical structure, 37 changes in the A form and 47 changes in the B form developed within this scope. After the changes in the measurement units of the developed test, the measurement units were converted to the metric units used in Türkiye. Thus, the adapted test has been changed to suit the mathematics culture of Türkiye as a result of the changes in the A Form and the B Form.

The fourth change in the adaptation process includes changes in the language structure. According to this, sample statements about the changes and the number of changes are presented in Table 5.

Table 5. *Exemplars of language structure changes to items and frequency of changes.*

Type of change	Example from original U.S. form	Example from adapted Turkish form	Number of items changed	
			Form A	Form B
Language structure	For each item	Her bir madde için	2	2
	Circle ONE answer	Sadece bir seçeneği işaretleyiniz		

It is essential to ensure the intelligibility of the items due to the changes that may occur in the language structure during the translation of the sentences or words in the test items into a different culture. The expression "For each item" in the original test is translated into Turkish as "her bir madde için". However, since the options in a multiple-choice test are expressed as "şık" in Turkish, the sentence is arranged as "her bir şık için".

3.2. Validity and Reliability

In order to examine the validity and reliability of the adapted version of a test, point biserial correlations were compared within the framework of psychometric analysis, item difficulty values, and reliability values of the forms were calculated.

3.2.1. *Point-biserial correlation results (r_{pbi})*

In Classical Test Theory, r_{pbi} was used to differentiate an item between respondents with higher mathematics teaching knowledge and those with low mathematics teaching knowledge. r_{pbi} for each item of the measurements of the A form of the PFA test in the Turkish context compared with the sample from the US teachers is given in Table 6.

Table 6. r_{pbi} *for patterns MKT-PFA test (Form A and Form B) items ordered by estimates on Turkish Algebraic Teaching for Knowledge test (Form A and Form B).*

Turkish r_{pbi} (Form A)	U.S. r_{pbi} (Form A)	Turkish r_{pbi} (Form B)	U.S. r_{pbi} (Form B)
0.321	0.440	0.185	0.420
0.512	0.543	0.288	0.523
0.425	0.443	0.192	0.231
0.451	0.493	0.345	0.350
0.504	0.753	0.491	0.491
0.164	0.220	0.410	0.478
0.337	0.632	0.313	0.534
0.345	0.442	0.405	0.540
0.441	0.745	0.336	0.346
0.444	0.341	0.368	0.423
0.375	0.598	0.336	0.560
0.258	0.285	0.431	0.567
0.301	0.333	0.426	0.506
0.423	0.498	0.500	0.625
0.416	0.499	0381	0.602
0.450	0.696	0.479	0.747
0.384	0.575	0.492	0.762
0.264	0.489	0.414	0.755
0.404	0.433	0.368	0.848
0.402	0.775	0.275	0.286
0.358	0.694	0.434	0.634
0.394	0.700	0.513	0.659
0.321	0.670	0.453	0.595
0.362	0.428	0.323	0.328
0.418	0.554	0.356	0.709
0.414	0.595	0.242	0.379
0.338	0.513	0.427	0.588
0.134	0.131		
0.343	0.342		

For each item in form A of the PFA test, r_{pbi} was compared with the measurements in the Turkish and US contexts. According to Hambleton et al. (1991) criteria the correlation between r_{pbi} is high (r=0.635; t=4,275; p<0.001). With this result, the test can measure the intended characteristics of teachers and preservice teachers, as there is a high correlation between the measurements in the USA and the measurements in Türkiye. The following additions were made to the Results section: The correlation values of the 17th and 26th items in the adapted A form and the correlation values of the 1st, 3rd, 21st, and 27th items in the B form were found to be < .3. It is also noted that the correlation values obtained for the 26th item in the A form and the 3rd and 21st items in the B form, when compared to the US version, were also < .3.

For each item in form B of the PFA test, r_{pbi} was compared with the measurements in the Turkish and US contexts. According to Hambleton et al. (1991) criteria the correlation value between r_{pbi} is high (r=0.6381; t=4.1438; p<0.001). With this result, we can say that the test can measure the intended characteristics of teachers and preservice teachers, as there is a high correlation between the measurements in the USA and in Türkiye.

3.2.2. *One-parameter IRT results*

In the study, item difficulty and item discrimination values for each item in the forms are given in Table 8 by using a one-parameter IRT model to obtain the item difficulty of the items in the A and B forms of the MKT-PFA test.

Table 7. *Item difficulties and discriminations of the items in the A and B forms of the MKT-PFA test.*

Form A			Form B		
Item	Item difficulty	Item discrimination	Item	Item difficulty	Item discrimination
3	-2.474	.913	1	-2.603	.903
4a	-1.033	.782	2	2.970	.927
4b	-0.767	.676	6a	-0.954	.842
4c	-1.326	.816	6b	-2.016	.930
4d	-1.888	.927	6c	-2.064	.933
7	0.792	.636	6d	-1.006	.869
8	-2.760	.959	8	-3.713	.960
9a	-0.413	.572	13a	-0.588	.796
9b	-2.034	.945	13b	0.851	.830
9c	-0.578	.728	13c	0.603	.903
9d	-2.054	.937	15	0.954	.842
11	4.451	.409	16a	2.063	.854
14a	2.222	.461	16b	1.565	.793
14b	1.788	.779	16c	2.461	.892
14c	1.085	.682	16d	3.325	.945
14d	-2.762	.966	19a	-1.851	.830
15	2.328	.837	19b	-3.641	.958
17	3.102	.899	19c	-3.325	.945
18	1.612	.757	19d	-3.139	.972
20a	5.133	.974	21	1.150	.272
20b	6.042	.986	25a	4.242	.974
20c	6.639	.991	25b	4.476	.979
20d	5.813	.984	25c	4.043	.970
22a	-4.126	.948	25d	1.288	.751
22b	-2.558	.858	25e	3.325	.945
22c	-1.084	.828	27	1.071	.715
22d	-5.042	.985	28	-2.132	.908
26	-0.277	.549			
27	1.261	.708			

It is seen that the item discrimination indexes of the items are greater than .40. According to Brennan and NCME (2006), it can be said that the discrimination of all items in forms A and B is good. Item difficulty parameters reflect the differentiation states of the participants in the item process (Baker, 2001; de Ayala, 2013). For this reason, the labels used to define the discrimination of the substances in the MKT-PFA test can be associated with the value ranges of the parameters, as indicated in Table 8:

Table 8. *Item difficulty distribution of the items in the A and B forms of the MKT-PFA test.*

Level of difficulty	Form A	Form B
Very easy ($-4 \geq x$)	4	4
Easy ($-2 \geq x > -4$)	7	5
Moderate ($2 \geq x \geq -2$)	7	7
Hard ($4 > x > 2$)	6	6
Very hard ($x \geq 4$)	5	5

When the item difficulties are examined, it is seen that the items in the A and B forms show a normal distribution. The forms adapted to this distribution can distinguish those with high mathematical knowledge in the sample from those with low mathematical knowledge.

3.2.3. *Reliability results*

The reliability of the A form and B forms of the MKT-PFA test, that is, the KR-20 values of how consistent the scores of the respondents are over more than one item or multiple tests, are given in Table 9.

Table 9. *Reliability of MKT-PFA test.*

Form	Number of the items	KR-20 values	Lord reliability
PFA-A form	N=29	.712	.733
PFA-B form	N=27	.735	.756

The KR-20 value of the A form of the MKT-PFA test data obtained as a result of the application to the preservice teachers was calculated as .712, and the KR-20 value of the B form as .735. The Lord reliability of the A form of the MKT-PFA test of the data obtained as a result of the application to the preservice teachers was calculated as .733, and the Lord reliability of the B form as .756. The measurements obtained from the test are reliable with this value obtained. When Table 9 is examined, there is a difference between the reliabilities of Form A and Form B. The reason for this is that the number of items in Form A is more than the number of items in Form B.

3.2.4. *Test Information Curve of A and B Forms of MKT-PFA Test*

The test information curve expresses the level of knowledge at which the achievement test best measures individuals. Figure 3 shows the test data curves for Form A and Form B. The x-axis in the graphs is the scale score of the preservice teachers; 0 generally corresponds to the average preservice teacher in the population studied; Negative scores indicate less knowledgeable preservice teachers, and positive scores indicate more knowledgeable preservice teachers.

Figure 3. *Test information curve of A and B forms of MKT-PFA test.*

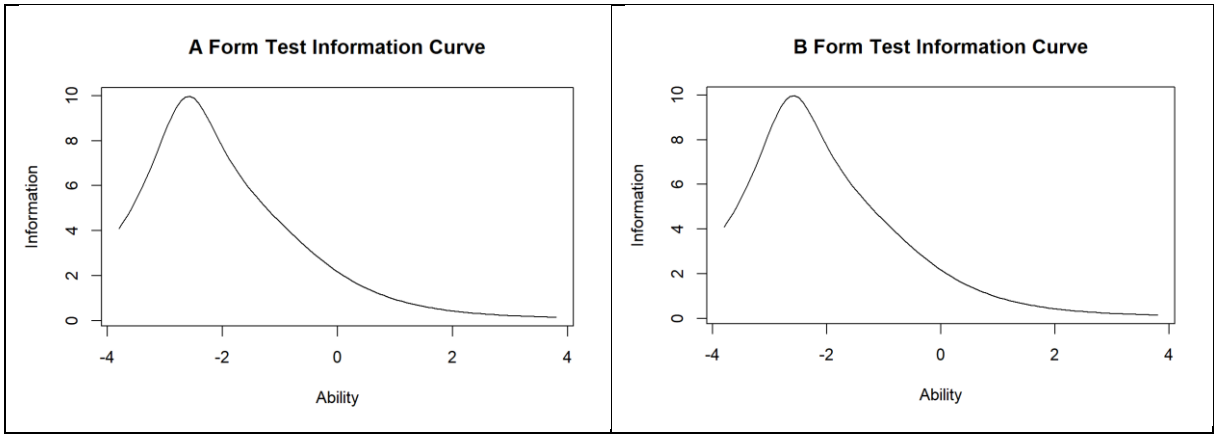


Figure 3 presents that the adapted form A and form B provide less information for preservice teachers who are 2.5 standard deviations above the mean and 1.5 standard deviations below the mean. Therefore, it means that form A and form B of the adapted test better distinguish preservice teachers with higher algebra knowledge for teaching from average or less algebra knowledge for teaching.

4. DISCUSSION and CONCLUSION

This section discusses the results of the analysis with qualitative and quantitative approaches in the adaptation process of the MKT-PFA test in the study. In this context, the cultural adaptation in the qualitative approach of the adaptation of the MKT-PFA test was analyzed in 4 categories.

The findings are discussed in each category. In the quantitative approach, the results obtained within the scope of the r_{pbi}, evaluating the item difficulties and reliability are discussed. Within the scope of this study, the Turkish adaptation of the items in the A and B forms of the MKT-PFA test was done by Delaney et al. (2008) following the steps given. Based on this research and similar studies, it can be said that the process of adapting such tests to a culture different from the one developed is arduous (Esendemir & Bindak, 2019; Marcinek et al., 2022; Ng, 2012; Ng et al., 2012).

4.1. Discussion of Studies in the Context of Cultural Adaptation of Test Items

The cultural adaptation process of the test items was carried out in four stages.

Changes in the context of general culture: The subject of general culture includes changing the non-mathematical but using daily language names and words in the test in a way that is suitable for the culture in question (Delaney et al., 2008). The use of food names in the question roots or options of the adapted test in mathematical problems serves to adapt to the cultural context. However, when adapting to a different country, the names of dishes or games in context may mean something different for the participants to whom the test will be applied (Ng, 2012). For this reason, cultural changes were made in this context in the study. While making changes, the names of similar foods were found without changing the mathematical situation in the problem. In addition, it was seen that changing the expression "baseball cards" to "playing cards" in the adapted culture in the adapted test did not make a semantic difference. For this reason, it did not create a change in the mathematical situation. Similarly, Ng (2012) adapted the word "pie" to their own culture as a cake or cake in their study.

Changes in the context of school culture: While adapting the items in the test, it was seen that the way of addressing the teachers differed between cultures. In the Turkish context, the term "teacher" is added next to the teacher's name, while in the Norwegian context, teachers are generally addressed by their first names. In addition, the expression of students in classes or groups is another matter of difference for different countries. For example, in Norway, there is an official statement that “classroom” should not be used when referring to student groups (Mosvold & Fauskanger, 2009). In addition to such differences in the context of school culture, it is seen that there are significant differences in teaching practices between cultures (Stiegler & Hiebert, 1999). In particular, the use of tangible materials as tools or models for representing mathematical ideas is different from the US in that many developing country teaching environments may not include physical manipulatives. For any of these manipulatives, for example, "Pattern Blocks" in the current study are clarified by providing either a description or a picture or both. While the context of the school culture is an important factor in determining the mathematical knowledge of the instructors, other factors also affect the mathematical knowledge required by the instructors.

Changes in the context of mathematical structure: Changes in the context of the mathematical structure of the items in the adapted test have the potential to lead to changes in the difficulty of the test (Delaney et al., 2008). Most symbolic expressions used in mathematics are universally acceptable. However, there may be differences between cultures regarding definitions or terminologies. Although technical terms such as “domino stones” or “mosaic” in the mathematical language context of the MKT-PFA test are available in Türkiye and the USA, these terms are not used at the primary level. Instead, a more general term, such as “pattern”, is used. Such changes do not affect the integrity of the test in measuring their mathematical knowledge, as they are not the terms that teachers use in their teaching. For this reason, terms that measure teachers' familiarity with certain technical words may be preferred instead.

Similarly, Ng (2012) changed the term “polygon” to a more familiar term for Indonesian teachers, thus replacing it with “bangun datar segibanyak” meaning “multilateral flat shape”. For Indonesian tutorials, this is a more descriptive term. Therefore, test items can be more understandable and easily adapted when evaluating teachers’ knowledge of polygon definition.

While these differences in mathematical language do not pose a problem in the Irish context (Delaney et al., 2008), there are substantial variations within the context of mathematical language in test items in the examples of Türkiye, Korea, Indonesia, and Norway. There are also changes in the units of measurement. For example, while the unit of weight in the adapted A form was "ounces" in the developed context, it was changed to "package number" in the context of Türkiye.

Correspondingly, in the context of Indonesia and Norway, they expressed the measure of butter in the MKT-G test as "sticks" or "number of cups". The situation for items, including money, is as follows: the fact that the difference between the currency in Türkiye and US has created a problem in terms of mathematical situations in the context of the items. For this reason, using equivalent values of money does not make mathematical sense. These two contexts are nearly impossible to translate into any of these languages without changing the entire context. This incomparable context problem poses a serious threat to the equivalence of the adapted data collection tool. Delaney et al. (2008) stated that there are relative similarities between the Irish and US forms of MKT forms, but there are differences in mathematical language, representation of concepts, measurement units, content and student knowledge. Although such differences can be ignored as they are mathematically insignificant, they indicate that differences in teachers' performance on some items are sensitive to seemingly minor changes in items. For this reason, Delaney et al. (2008), Ng (2012), and Marcinek et al. (2022) stated that many changes can be made in the items of tests adapted to different cultures since the methods of teaching mathematics in cultures with different languages are significantly different.

Changes in language structure context: Delaney et al. (2008) stated minor language problems in the process of adapting the MKT test to Norway. They stated that these changes would not change the validity of the test items. However, additional explanations should be created to avoid confusion that may make the explanations at the root of the question or item in the test long and complex. As a result, all these factors should be considered when determining the mathematical knowledge of teachers and preservice teachers in different countries (Delaney et al., 2008).

4.2. Discussion of Findings Obtained from Psychometric Tests

After the cultural adaptation process for the items in the A and B forms of the MKT-PFA test was completed, point biserial correlation was obtained for each item in the A and B forms. It was concluded that r_{pbi} of the data was highly correlated between Türkiye and the United States. It was observed that there was a high level of correlation between the test adapted to Turkish and the test developed in the USA. It is seen that the correlations of some questions in the test adapted to Turkish are <.3. This situation also appears to be the case in the original form of MKT-PFA. In addition, these correlations in the study may be higher when working with larger sample groups. In addition, when we determine the subgroups in the low-correlation questions in the adapted form A and form B and look at the relationship at the class level, it is seen that the correlation value among the 4th grade preservice teachers is >.3. It was observed that there was a high correlation between the test adapted to Turkish and the test developed in the USA. In addition, using a one-parameter IRT model, it was seen that the distribution of the item difficulty values obtained for each item in the forms could distinguish between those with high mathematical knowledge in the sample and those with low mathematical knowledge in the adapted forms. Additionally, when the discrimination values of the items were examined, it was concluded that they were >.40. It can be said that form A and form B of the adapted tests can distinguish between preservice teachers who have good algebra knowledge for teaching and preservice teachers whose algebra knowledge for teaching is average or less. The test is reliable with these values obtained according to the KR-20 and Lord reliability values obtained for both forms of the MKT-PFA test. Finally, when the A form and B form Test Information Curve of the adapted test are examined, it is seen that the form better distinguishes the preservice teachers with higher algebra teaching knowledge from the preservice teachers with average or less

algebra teaching knowledge. In the pilot study for US teachers who participated in California's Mathematical Professional Development Institute, the MKT-PFA test Form A, Form B, and Form C provided maximum information for less knowledgeable teachers whose abilities are one-half standard deviation below the mean (Hill, 2007). The adapted Form A and Form B provide less information for preservice teachers who are 2.5 standard deviations above the mean and 1.5 standard deviations below the mean. Therefore, it means that Form A and Form B of the adapted test better distinguish preservice teachers with higher algebra knowledge for teaching from average or less algebra knowledge for teaching.

In this study, the MKT-PFA test was adapted to examine the mathematical knowledge of teachers and preservice teachers in Türkiye. In the adaptation process, the results of adaptation were included when translating test items from one language to another and for use in a different environment than intended. Delaney et al. (2008) suggest that international comparisons of teachers' mathematical knowledge should be evaluated in light of the differences that may exist in teachers' mathematical knowledge used in each country. Thus, clear guidelines should be developed to adapt the mathematical teaching information items. In addition, the differences in the mathematical knowledge of teachers or preservice teachers between countries can be explained by the differences in the mathematical knowledge used by teachers or preservice teachers in the relevant countries (Mosvold & Fauskanger, 2009; Ng, 2012). When comparing the knowledge of teachers between countries, it is insufficient to adapt the items from one country alone. For this reason, our study will shed light on the studies comparing different cultures with the Turkish context for the MKT-PFA test. Such research may lead to further development of the theoretical structure of MKT and possible cultural differences related to this structure. Additionally, Algebraic Knowledge for Teaching focuses on the knowledge and skills required for teachers or preservice teachers to improve their ability to explain and teach algebraic concepts to students. Ball et al (2008) discuss this special knowledge that teachers should have and how they can guide students' understanding of algebraic concepts. It aims to provide information to mathematics educators about the algebra teaching knowledge of preservice teachers with the adapted algebra knowledge for teaching tests. In this way, they can build "Algebra Teaching" courses aimed at the algebra teaching knowledge levels of preservice teachers. In addition, different instructional designs can be applied to better understand the relationship and interaction between mathematics teaching and MKT.

When we look at the results of the psychometric tests, the adaptation of the A and B forms of the MKT-PFA test is generally appropriate based on the psychometric analyses. In other words, a test developed to measure the mathematics knowledge of secondary mathematics teachers working in schools in the U.S. was successfully adapted to the Turkish context.

Acknowledgments

These results are a part of the University of Gaziantep doctoral thesis research project numbered EF.DT.22.04.

Declaration of Conflicting Interests and Ethics

The authors declare no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the authors. **Ethics Committee Number:** Gaziantep University, 21.03.2022-162213.

Contribution of Authors

Ali Bozkurt: Investigation, adaptation of the instruments and receiving experts' opinions. **Begüm Özmusul**: Adaptation of the instruments, data analysis, resources and visualization.

Orcid

Ali Bozkurt https://orcid.org/0000-0002-0176-4497
Begüm Özmusul https://orcid.org/0000-0003-0163-5406

REFERENCES

An, S., Kulm, G., & Wu, Z. (2004). The pedagogical content knowledge of middle school, mathematics teachers in China and the U.S. *Journal of Mathematics Teacher Education, 7*(2), 145-172.

Aryadoust, V., Ng, L.Y., & Sayama, H. (2021). A comprehensive review of Rasch measurement in language assessment: Recommendations and guidelines for research. *Language Testing, 38*(1), 6-40. https://doi.org/10.1177/0265532220927487

Baker, F.B. (2001). *The basics of item response theory* (2nd ed). (ED458219). https://eric.ed.gov/?id=ED458219

Baker, F.B., & Kim, S.H. (2017). *The basics of item response theory using R* (Vol. 969). Springer.

Ball, D.L. (1990). Prospective elementary and secondary teachers' understanding of division. *Journal for Research in Mathematics Education, 21*(2), 132-144.

Ball, D.L., & Hill, H.C. (2008). Measuring teacher quality in practice. In D. H. Gitomer (Ed.), *Measurement Issues and Assessment for Teaching Quality*, pp. 80-98. SAGE.

Ball, D.L., Hill, H.C., & Bass, H. (2005). Knowing mathematics for teaching: Who knows mathematics well enough to teach third grade, and how can we decide? *American Educator (Fall 2005)*, 14-46.

Ball, D.L., Thames, M.H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special. *Journal of Teacher Education, 59*(5), 389-407. https://doi.org/10.1177/0022487108324554

Brennan, R.L., & National Council on Measurement in Education (NCME). (2006). *Educational measurement.* Praeger.

Borsboom, D. (2005). *Measuring the mind: Conceptual issues in contemporary psychometrics*. Cambridge University.

Boykin, A.A., Ezike, N.C., & Mysers, A.J. (2023). Model-data fit evaluation: Posterior checks and Bayesian model selection. *International Encyclopedia of Education (4th Edition)*, 279-289.

Burnham, K.P., & Anderson, D.R. (2002). *Model selection and multimodal inference: a practical information-theoretic approach*. Springer.

Charalambous, C.Y. (2008). *Prospective teachers' mathematical knowledge for teaching and their performance in selected teaching practices: Exploring a complex relationship*. (Doctoral dissertation) University of Michigan.

Chou, Y.T., & Wang, W.C. (2010). Checking dimensionality in item response models with principal component analysis on standardized residuals. *Educational and Psychological Measurement, 70*(5), 717-731. https://doi.org/10.1177/0013164410379322

Cohen, I. (2011). Teacher-student interaction in classrooms of students with specific learning disabilities learning English as a foreign language. *Journal of Interactional research in communication disorders, 2*(2), 271-292. https://doi.org/10.1558/jircd.v2i2.271

Council of Higher Education, (CoHE (Yükseköğretim Kurulu), 2018). *New teacher training programs, reasons for updating the programs, innovations and implementation principles* [In Turkish]. https://www.yok.gov.tr/Documents/Kurumsal/egitim_ogretim_dairesi/Yeni-Ogretmen-Yetistirme-Lisans_Programlari/AA_Sunus_%20Onsoz_Uygulama_Yonergesi.pdf [In Turkish]

Cole, Y. (2012). Assessing elemental validity: The transfer and use of mathematical knowledge for teaching measures in Ghana. *ZDM, 44*(3), 415-426. https://doi.org/10.1007/s11858-012-0380-7

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. (ED312281). ERIC. https://eric.ed.gov/?id=ED312281

Cronbach, L.J., & Shavelson, R.J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement, 64*(3), 391-418.

Çelen, Ü. (2008). Comparison of validity and reliability of two tests developed by classical test theory and item response theory. *Elementary Education Online, 7*(3), 758-768.

de Ayala, R.J. (2013). *The theory and practice of item response theory.* Guilford.

Delaney, S., Ball, D.L., Hill, H.C., Schilling, S.G., & Zopf, D. (2008). Mathematical knowledge for teaching: Adapting US measures for use in Ireland. *Journal of Mathematics Teacher Education, 11*(3), 171-197. https://doi.org/10.1007/s10857-008-9072-1

Driscoll, M. (1999). *Fostering algebraic thinking: a guide for teachers grades 6-10.* NH: Heinemann.

Edelen, M.O., & Reeve, B.B. (2007). Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Quality of life research, 16*, 5-18. https://doi.org/10.1007/s11136-007-9198-0

Embretson, S.E., & Reise, S.P. (2013). *Item response theory.* Psychology.

Esendemir, O., & Bindak, R. (2019). Adaptation of the test developed to measure mathematical knowledge of teaching geometry in Turkey. *International Journal of Educational Methodology, 5*(4), 547-565. https://doi.org/10.12973/ijem.5.4.547

Fan, J., & Bond T. (2019). Applying Rasch measurement in language assessment: Unidimensionality and local independence. In V. Aryadoust & M. Raquel (Eds.), *Quantitative data analysis for language assessment, Vol. I: Fundamental techniques* (pp. 83–102). Routledge. https://doi.org/10.4324/9781315187815

Fauskanger, J., Jakobsen, A., Mosvold, R., & Bjuland, R. (2012). Analysis of psychometric properties as part of an iterative adaptation process of MKT items for use in other countries. *ZDM, 44*, 387–399. https://doi.org/10.1007/s11858-012-0403-4

Frary, R.B. (1989). Partial credit scoring methods for multiple choice Tests. *Applied Measurement in Education, 2*(1), 79-96.

Hambleton, R.K, Swaminathan, H., & Rogers, H.J. (1991). *Fundamentals of item response theory*. SAGE Publications.

Hambleton, R.K., & Swaminathan, H. (1985). *Item response theory: principles and applicatio ns.* Academic Publishers Group.

Hambleton, R.K. (1994). Guidelines for adapting educational and psychological tests: A progress report. *European Journal of Psychological Assessment, 10*, 229–240.

Han, H. (2022). The effectiveness of weighted least squares means and variance adjusted based fit indices in assessing local dependence of the rasch model: Comparison with principal component analysis of residuals. *PloS ONE, 17*(9). https://doi.org/10.1371/journal.pone.0271992

Hill, H.C. (2007). Mathematical knowledge of middle school teachers: Implications for the no child left behind policy initiative. *Educational Evaluation and Policy Analysis, 29*(2), 95–114. https://doi.org/10.3102/0162373707301711

Hill, H.C., & Ball, D.L. (2004). Learning mathematics for teaching: Results from California's mathematics professional development institutes. *Journal for Research in Mathematics Education, 35*(5), 330-351.

Hill, H.C., Rowan, B., & Ball, D.L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research journal, 42*(2), 371-406. https://doi.org/3699380

Hill, H.C., Schilling, S.G., & Ball, D.L. (2004). Developing measures of teachers' mathematics knowledge for teaching. *The Elementary School Journal, 105*(1), 11-30. https://doi.org/10.1086/428763

Hill, H., & Ball, D.L. (2009). The curious and crucial case of mathematical knowledge for teaching. *Phi Delta Kappan, 91*(2), 68-71. https://doi.org/10.1177/00317217090910021

Holmes, F. & Brian F.F. (2019). A comparison of estimation techniques for IRT models with small samples, *Applied Measurement in Education, 32*(2), 77-96. https://doi.org/10.1080/08957347.2019.1577243

Huang, R., & Kulm, G. (2012). Preservice middle grade mathematics teachers' knowledge of algebra for teaching. *The Journal of Mathematical Behavior, 31*(4), 417-430. https://doi.org/10.1016/j.jmathb.2012.06.001

Kieran, C., Kieran, C., & Ohmer. (2018). *Teaching and learning algebraic thinking with 5-to 12-year-olds* (pp. 79-105). Springer.

Kim, Y. (2016). Interview prompts to uncover mathematical knowledge for teaching: focus on providing written feedback. *The Mathematics Enthusiast, 13*(1), 71-92. https://doi.org/10.54870/1551-3440.1366

Kim, Y. (2020). Korean teachers' mathematical knowledge for teaching in algebraic reasoning. *Journal of Educational Research in Mathematics, (Special Issue)*, 185-198. https://doi.org/10.29275/jerm.2020.08.sp.1.185

Knipping, C. (2003). Learning from comparing. *Zentralblatt für Didaktik der Mathematik, 35*(6), 282-293.

Kline, P. (1994) *An Easy Guide to Factor Analysis.* Routledge.

Koellner, K., Jacobs, J., Borko, H., Schneider, C., Pittman, M.E., Eiteljorg, E., & Frykholm, J. (2007). The problem-solving cycle: A model to support the development of teachers' professional knowledge. *Mathematical Thinking and Learning, 9*(3), 273-303. https://doi.org/10.1080/10986060701360944

Kwon, M., Thames, M.H., & Pang, J. (2012). To change or not to change: Adapting mathematical knowledge for teaching (MKT) measures for use in Korea. *ZDM, 44*, 371–385. https://doi.org/10.1007/s11858-012-0397-y

Langrall, C.W. & Swafford J.O. (1997). Grade six students' use of equations to describe and represent problem situation. *Paper presented at the American Educational Research Association*, Chicago, IL.

Lew, H.C. (2004). Developing algebraic thinking in early grades: Case study of Korean elementary school mathematics. *The Mathematics Educator, 8*(1), 88-106.

Lord, F.M. & Novick, M.R. (1968). *Statistical theories of mental test scores*. Addison-Wesley Pub. Co.

Ma, L. (1999). *Knowing and teaching elementary mathematics: Teachers' understanding of fundamental mathematics in China and the United States.* Lawrence Erlbaum Associates, Inc.

Marcinek, T., & Partová, E. (2016). Exploring cultural aspects of knowledge for teaching through adaptation of U.S.-developed measures: Case of Slovakia. Paper presented at *the 13th International Congress on Mathematical Education*. Hamburg, Germany.

Marcinek, T., Jakobsen, A., & Partová, E. (2022). Using MKT measures for cross-national comparisons of teacher knowledge: case of Slovakia and Norway. *Journal of Mathematics Teacher Education*, 1-31. https://doi.org/10.1007/s10857-021-09530-3

Morris, A.K., Hiebert, J., & Spitzer, S.M. (2009). Mathematical knowledge for teaching in planning and evaluating instruction: What can preservice teachers learn?. *Journal for research in mathematics education, 40*(5), 491-529. https://doi.org/40539354

Mosvold, R., & Fauskanger, J. (2009). *Challenges of translating and adapting the MKT measures for Norway.* Paper presented at the American Educational Research Annual Meeting in San Diego, CA.

National Mathematics Advisory Panel (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel.* U.S. Department of Education.

Ng, D. (2012). Using the MKT measures to reveal Indonesian teachers' mathematical knowledge: Challenges and potentials. *ZDM, 44*(3), 401-413. https://doi.org/10.1007/s11858-011-0375-9

Ng, D., Mosvold, R., & Fauskanger, J. (2012). Translating and adapting the mathematical knowledge for teaching (MKT) measures: The cases of Indonesia and Norway. *The Mathematics Enthusiast, 9*(1), 149-178. https://doi.org/10.54870/1551-3440.1238

Özdemir, D. (2004). A comparison of psychometric characteristics of multiple choice tests based on the binaries and weighted scoring in respect to classical test and latent trait theory. *Hacettepe University Journal of Education, 26*, 117-123.

Pekmezci, F.B., & Avşar, A.Ş. (2021). A guide for more accurate and precise estimations in Simulative Unidimensional IRT Models. *International Journal of Assessment Tools in Education, 8*(2), 423-453. https://doi.org/10.21449/ijate.790289

Reyhanlıoğlu, Ç., & Doğan, N. (2020). An analysis of parameter invariance according to different sample sizes and dimensions in parametric and nonparametric item response theory. *Journal of Measurement and Evaluation in Education and Psychology*, *11*(2), 98-112. https://doi.org/10.21031/epod.584977

Şahin, A., & Anıl, D. (2017). The effects of test length and sample size on item parameters in item response theory. *Educational Sciences: Theory & Practice, 17*(1), 321-33. https://doi.org/10.12738/estp.2017.1.0270

Schmittau, J. (2005). The development of algebraic thinking. *Zentralblatt für Didaktik der Mathematik, 37*(1), 16-22.

Sheng, Y. (2013). An empirical investigation of Bayesian hierarchical modeling with unidimensional IRT models. *Behaviormetrika*, *40*(1), 19-40. https://doi.org/10.2333/bhmk.40.19

Shulman, L.S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, *15*(2), 4-14.

Sireci, S.G., Thissen, D., & Wainer, H. (1991). On the reliability of test let-based tests. *Journal of Educational Measurement, 28*(4), 237-247.

Smith, R.M., & Miao, C.Y. (1994). Assessing unidimensionality for Rasch measurement. *Objective Measurement: Theory into Practice, 2,* 316-327.

Stiegler, J.W., & Hiebert, J. (1999). *The Teaching Gap. Best ideas from the world's teachers for improving education in the classroom*. The Free.

Strand, K., & Mills, B. (2014). Mathematical content knowledge for teaching elementary mathematics: A focus on algebra. *The Mathematics Enthusiast, 11*(2), 385-432. https://doi.org/10.54870/1551-3440.1307

Tabachnick, B., & Fidell, L. (2012). *Using multivariate statistics.* Pearson.

Welder, R.M., & Simonsen, L.M. (2011). Elementary Teachers' Mathematical Knowledge for Teaching Prerequisite Algebra Concepts. *Issues in the Undergraduate Mathematics Preparation of School Teachers, 1*.

Wilson, L., Andrew, C., & Sourikova, S. (2001). Shape and structure in primary mathematics lessons: A comparative study in the North-east of England and St Petersburg, Russia-some implications for the daily mathematics lesson. *British Educational Research Journal, 27*(1), 29-58.

Wright, B.D. (1996). Local dependency, correlations and principal components. Rasch *Measurement Transactions, 10*(3), 509–511. https://www.rasch.org/rmt/rmt103b.htm

Yang, S. (2007). *A comparison of unidimensional and multidimensional RASCH models using parameter estimates and fit indices when assumption of unidimensionality is violated* [Doctoral dissertation]. The Ohio State University.

Yen, W.M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement, 30*(3), 187–213.

Zazkis, R., & Liljedahl, P. (2002). Generalization of patterns: The tension between algebraic thinking and algebraic notation. *Educational Studies in Mathematics*, *49*(3), 379-402.

APPENDIX

Some released items from MKT - test

1. Zeliha ve öğretmeni Zeliha'nın doğum gününde beraber kurdukları aşağıdaki problemi sınıf arkadaşlarına sormuşlardır:

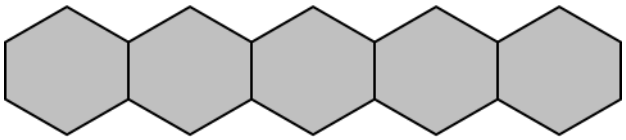
Zeliha'nın şimdiki yaşı erkek kardeşinin yaşının iki katıdır. Kaç yıl sonra Zeliha'nın yaşı kardeşinin yaşının yine iki katı olur?

Arkadaşları aşağıdaki cevapları vermiştir. Bu cevaplardan hangisini doğru olarak kabul edersiniz? (Sadece bir şıkkı işaretleyiniz.)

A) Her 2 yılda bir olur.
B) Zeliha'nın yaşına bağlıdır.
C) Zeliha'nın yaşı, şimdiki yaşının 2 katı olduğunda olur.
D) Bir daha asla olmaz.

2. Jale öğretmen dersinde kullanmak için şu problemi hazırlamıştır:

Aşağıdaki gibi bir satıra yan yana 100 düzgün altıgeni dizerseniz oluşan şeklin çevresi ne olur?



Jale öğretmen öğrencilerinden gelebilecek farklı çözümleri görmek istemiştir. Aşağıda verilmiş muhtemel öğrenci çözümlerinden hangileri doğru cevaba götürür? (Her bir şık için evet, hayır veya emin değilim seçeneklerinden birini işaretleyiniz.)

		Evet	Hayır	Emin Değilim
A)	4 x 100 + 2	1	2	3
B)	(6 x 100) – 2 x 99	1	2	3
C)	4 x 98 + 2 x 5	1	2	3
D)	6 x 100	1	2	3