Uluslararası Eğitimde Yenilikçi Yaklaşımlar Dergisi
International Journal of Innovative Approaches in Education 2021, Vol. 5 (1), 33-43
https://doi.org/10.29329/ijiape.2021.333.3

Original article

An Investigation Into Problem Solving Skills of Gifted Students' via Multi-Facet Rasch Analysis

Opinions of Turkish Language and Literature Teacher Candidates On Teaching Practice Course

Şirin Çetin [a, *], İrem Kar [b] & Burcu Durmaz [c]

[a] Department of Biostatistics, Faculty of Medicine, Tokat Gaziosmanpaşa University, Tokat, Turkey
[b] Department of Biostatistics, Faculty of Medicine, Ankara University, Ankara, Turkey
[c] Department of Mathematics and Science Education, Faculty of Education, Süleyman Demirel University, Isparta, Turkey

Abstract

Problem solving, which is an important skill for mathematics, has long been recognized as the focus of the curriculum. This study focuses on both the systematic of problem solving and the use of strategies for different solutions, so aim of this study was to examine gifted middle school students' mathematical problem-solving skills via multi-facet Rasch analysis. For this purpose, data were collected via 12 non-routine mathematical open-ended problems from a total of 276 students who attend Science and Art Centers. The data were scored by two independent researchers according to the criteria in the range of 1-5 points per problem. In the study, it was seen that Grader 1 behaved more tolerant than Grader 2 in terms of evaluation for all problems except 6th problem. In addition, the most difficult and the easiest problems were differed between graders. From the findings of the study, it can be said that even if the evaluation criteria were clearly defined the difference between graders was not resolved. It can be suggested that to take necessary precautions for making vital decisions in all grade levels via teacher made tests and to use more than one grader instead of just one grader for evaluation.

Keywords: Gifted, Mathematics Education, Non-Routine Problems.

Received: 20 February 2021 * **Accepted:** 10 March 2021 * **DOI:** https://doi.org/10.29329/ijiape.2021.333.3

* **Corresponding author:**

Şirin Çetin, Department of Biostatistics, Faculty of Medicine, Tokat Gaziosmanpaşa University, Tokat, Turkey.
Email: cetinsirin55@gmail.com

INTRODUCTION

Problem solving, which is an important skill for mathematics, has long been recognized as the focus of the curriculum (NCTM, 2000). Problem solving serves for three different purposes such as teaching about, for and via problem solving in mathematics lessons (Schroeder and Lester, 1989). This study focuses on both the systematic of problem solving and the use of strategies for different solutions, so the skills in teaching for problem solving were examined. In order to examine the systematics of problem solving, non-routine problems are more functional than routine problems which are in textbooks (Altun, 2014) and non-routine problems are more related to mathematical creativity so they're also defined as the problems that are not known which strategies should be used to reach a solution and require the generation of original ideas (Milgram, 2017). According to Polya (1957), who asserts that there are four stages that must be passed in order to see the systematics of the problem solving process, non-routine problems are more convenient than routine problems in order to study on all stages of problem solving. In addition to these, according to their peers, "challenge" is an important component in the education of gifted students who perform at high levels in different fields, can be met by integrating non-routine problems with teaching. Because gifted students have difficulties and uncertainities while solving non-routine problems, they have to rearrange their knowledge to overcome these situations (Dreyfus, 2007; Hershkowitz, Schwarz & Dreyfus, 2001).

When the studies on mathematical problem solving are examined in the literature different topics were studied such as the difficulties experienced by the students in problem solving stages, the change in cognitive and affective characteristics related to problem solving after experimental instruction, the skills and levels of using problem solving strategies (Lee, 1982; Rose, 1991; Hogan, 2004; Gürcan Töre, 2007; Yıldız, 2008; Taşpınar, 2011; Durmaz, 2014).

From the measurement of problem solving skills to the evaluation of mathematical creativity, two basic theories are used in the development of the scales used in the process of making educational decisions: Item Response Theory (IRT) and Classical Test Theory (CTT). IRT is a theory that emerges from the limitations of the CTT. This theory is also called Modern Test Theory. IRT has many advantages over CTT. Items and test parameters were dependent on sampling in CTT, whereas item parameters in IRT were independent of the distribution of individuals' trait levels. The examined property levels of individuals are independent of the items in the measurement tool (Hambleton, Swaminathan and Rogers, 1991). In order to obtain talent estimation in IRT analysis rater, item and person surfaces included in the analysis are tried to be placed on a common axis called "logit". While processing directly on raw scores in the CTT, the measurements for each surface are converted to an equally spaced logit scale and estimations are made about the ability levels of the students considering the components of all surfaces. In other words, talent estimation of individuals is calculated on the basis of all points given by all raters to all items. It can be said that in IRT analysis, where three surfaces are in the form of individual, rater and item, ability estimations are obtained with the help of the function obtained from the three surfaces processed from the analysis (Linacre, 1989; Lunz & Wright, 1997).

The Multi-Facet Rasch Model (MFRM) is an extension of the single parameter model IRT. For MFRM, the formula of i^{th} person's probability of receiving x points from n^{th} rater for j^{th} item can be given as (Linacre, 1989):

$$P(X = x_{nijx}) = \frac{\exp\left[x(\theta_n - \beta_i - C_j) - \sum_{k=0}^{x} F_k\right]}{\sum_{r=o}^{m} \exp\left[r(\theta_n - \beta_i - C_j) - \sum_{k=0}^{r} F_k\right]}$$

In the formula: x = 0, 1, …,m

θn : the level of the n^{th} person's monitoring characteristic

βi : i^{th} item's difficulty level

Cj : j^{th} rater's attitude

Fk : The difficulty of observing k^{th} category according to the $(k-1)^{th}$ category

There is no limitation in the difference between the raters in the MFRM analysis; these differences are kept under control and statistics on individuals and items are calculated according to the formula (Abu Kassim, 2007). It is seen that MFRM is used in many researches about education for different purposes such as peer assessment process, the reliability of a statistics exam consisting of open-ended questions at university level, evaluation of scientific research assignments, evaluation of secondary school mathematics curriculum, evaluation of activities in science course, micro teaching practices of preservice teachers who are taking special teaching methods and examining presentation skills (Baştürk, 2010; Semerci, 2011; Batdı, 2014; Güler, 2014; Yüzüak, Yüzüak & Kaptan, 2015; Şahin, Taşdelen Teker & Güler, 2016; Kaya Uyanık, Güler, Taşdelen Teker and Demir, 2019; Köse, Usta & Yandı, 2016).

While teachers make some educational decisions about their students, they mostly use teacher-made tests and prefer open-ended questions instead of multiple-choice questions for ease of preparation (Başol, 2013). Open-ended questions; reducing the measurement error (Turgut and Baykul, 2012), eliminating the chance factor that appears to be a disadvantage in multiple choice tests (Braun, Bennett, Frye and Soloway, 1990) and to ensure the emergence of skills that cannot be measured by multiple choice questions (Bahar, Nartgün, Durmuş & Bıçak, 2010). Despite all its advantages, the biggest disadvantage of open-ended questions is that reliability can be low due to the rater, which is one of the most important features of a measurement tool, because such questions cannot be scored objectively as in multiple choice questions (Romagnano, 2001; Takunyacı, 2016). This situation causes the scores of the individuals taking the tests consisting of open-ended questions to differ according to the rater (Tekin, 2009; Atılgan, Kan & Doğan, 2009). In order to avoid this limitation, it may be suggested that teachers should be given training on scoring open-ended questions. Because, in one study, mathematics teachers improved their stiffness/generosity levels after the training and teachers started to make more consistent scores with each other (Takunyacı, 2016). Similarly, İnceçam, Demir & Demir (2018) examined the competence of middle school teachers to prepare open-ended items frequently used in written surveys in classroom assessment and evaluation, and concluded that teachers were very inadequate.

Based on the literature, it can be said that open-ended questions, which are used as a tool for making important educational decisions, there are some cases that threaten the reliability of measurement results both during the preparation and evaluation processes. The assessment and evaluation process, in which it is needs to be carried out impartially, plays a critical role for each country to direct its human resources to the appropriate places. Based on these reasons, in this study, it is examined whether different raters' attitudes for open-ended questions used in the assessment of the ability of gifted students mathematical problem solving skills in the evaluation process may cause rater bias or not. In the literature, it is considered that the study will contribute to both practice and further researches since there are no studies examining the evaluation of non-routine problem solving skills with MFRM and evaluating the use of open-ended questions in the assessment of academic skills of gifted students in terms of rater attitudes.

METHODOLOGY

Research Design

In this study, we examined an existing situation so the case study model is used (Karasar, 2008).

Research Sample

The study group consisted of 276 gifted middle school students attending the Science and Art Center. The criterion sampling, which is one of the purposeful sampling methods because it deals with the gifted students is used and also easy sampling method is taken into account because the students are preferred who are attending Science and Art Centers where are close to the researchers'. In the criterion sampling, some criteria are determined for the study group and the units that meet these criteria are included in the study (Yıldırım & Şimşek, 2008). The criterion in this study is that identification by the official institutions as gifted.

Research Instruments and Procedures

The data collection tool of the research is a problem solving skill test consisting of 12 open-ended questions. Non-routine problem types such as make an organized list and write an equation etc. are included in this test. Researchers seeked for other studies in the literature while preparing the questions. Pilot study has been done at different grade levels and the test has been finalized according to expert opinions.

Data Analysis

The data of the study were first examined by two independent researchers in accordance with the scoring criteria determined by the researchers before rating process. Then the researchers independently scored the students' responses to the problem solving test in the range of 1-5 points. Accordingly (1 point) solution or chosen strategy is completely wrong, (2 points) one of them -solution or chosen strategy- is wrong, (3 points) solution or selected strategy is partially correct, (4 points) most of the solution is correct and chosen strategy is suitable for the solution of the problem, (5 points) solution is completely correct and selected strategy is classified as appropriate for the solution. The Spearman-Brown correlation coefficient was calculated for non-normally distributed scores and the reliability between the raters was 0.954. The reliability coefficient obtained as a result of Rasch analysis is similar to the Cronbach Alpha and KR-20 reliability coefficients (Linacre, 1997). The data which were evaluated by the researchers is given in Table 1.

Table 1. Data records students assessed by two raters with 12 item basis responses

		Items						
Students	Raters	1	2	3	…	7	…	12
1	1	2	5	5	…	3	…	4
1	2	1	5	5	…	2	…	3
2	1	1	5	5	…	4	…	1
2	2	1	4	4	…	2	…	1
…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…
276	1	4	5	5	…	5	…	5
276	2	2	5	4	…	4	…	4

Table 1 shows that each student’s (276 students in total) answers to each test item (12 items in total) were rated by two raters. The facets included in the many facet Rasch analysis were arranged as examinee, item and rater. For instance, the first line of the data set shows the points given by the first rater to the answers by the student 1 (rater 1 gave 2 points to the answer by student 1 to item 1). The last line of the data set shows the points given by the second rater to the answers by student 276. RUMM 2020 program was used to analyze the data set. The variables included in the data set are raters and 12 items that are answered for problem solving assessment. The responses of 276 students in the data set for 12 items were scored by 2 raters between 1-5. The weighted maximum likelihood method was used as the parameter estimation method. After testing the assumptions of Rasch analysis, paired sample t-test was conducted to determine whether there was a difference between the person ability estimations for interactive and additive modes.

FINDINGS

In this section, findings are discussed. Figure 1 shows facet item map created by Multi Facet Rasch analysis.

```
        FACET 1                                         FACET 2
------------------------------------------------------------
  1,0 |                                           1,0 |
      |                                               |
      |                                               |
      | I0012        I0009                            | 2
      | I0008        I0010        I0004               |
  0,0 | I0005        I0001                        0,0 |
      | I0007        I0011        I0006               |
      |                                               |
      | I0003                                         | 1
      |                                               |
 -1,0 |                                          -1,0 |
      |                                               |
      | I0002                                         |
      |                                               |
      |                                               |
 -2,0 |                                          -2,0 |
------------------------------------------------------------
```

Figure 1. Facet item map created by multi facet Rasch analysis

In the Facet 1 column of Figure 1, the items were ranked by their difficulty levels. The item difficulty increases from the bottom to the top of this column. Accordingly, the most difficult question was item 12, and the easiest question was item 2. The raters are shown in the Facet 2 column of Figure 2. The rater severity increased from the bottom to the top of the Facet 2 column. Therefore, it was determined that rater 2 was more severe rating than grader 1. According to the results of paired sample t-test, there was a significant difference between the means of interactive and additive modes' ability estimations ($p<0.001$). Hence, it can be concluded that there was a significant rater effect, thus, the item difficulty levels should be evaluated by rater. In Table 2, item summaries for raters are given.

Table 2. Item summaries for raters

Rater 1						Rater 2					
Item	Location	SE	Residual	χ^2	p	Item	Location	SE	Residual	χ^2	p
I2	-1,785	0,078	0,713	13,415	0,004	I2	-0,912	0,079	0,248	2,842	0,417
I3	-0,919	0,054	-0,443	13,108	0,004	I6	-0,335	0,045	-0,988	7,263	0,064
I1	-0,704	0,066	-0,780	2,103	0,551	I3	-0,069	0,057	-1,106	7,729	0,052
I8	-0,678	0,047	0,690	6,124	0,106	I7	0,317	0,054	-2,905	19,776	0,000
I11	-0,676	0,048	2,859	21,846	0,000	I11	0,394	0,058	0,833	2,888	0,409
I7	-0,449	0,048	0,006	10,205	0,017	I5	0,461	0,060	0,857	2,647	0,449
I12	-0,386	0,047	-0,004	6,348	0,096	I4	0,556	0,057	1,067	2,626	0,453
I5	-0,329	0,051	-0,836	4,624	0,202	I1	0,765	0,074	0,840	16,556	0,001
I10	-0,139	0,044	-0,688	1,125	0,771	I10	0,908	0,053	-2,189	6,218	0,101
I9	-0,099	0,045	-0,524	4,740	0,192	I9	1,121	0,055	-1,889	8,360	0,039
I6	-0,055	0,048	-1,391	7,171	0,067	I8	1,465	0,054	0,468	5,505	0,138
I4	0,025	0,055	-1,231	5,383	0,146	I12	1,523	0,058	-2,123	10,033	0,018

SE:standard error, Residual: fitted residuals, χ^2:chi-square test statistics, p:p value

Table 2 shows that ascending sort of item difficulties for raters. According to Table 2, the difficulty level of the items for each raters ranged between -1.785 logit and 0.025 logit , -0.912 logit and 1.523 respectively. In all items except 6th, Rater 1 is more lenient in scoring than Rater 2. While item 2 is the easiest item for both raters, item 4 is the most difficult item for Rater 1 and item 12 is the most difficult item for Rater 2. The greatest difference between raters was observed in 6th and 8th items. For Rater 1, item 6 is the second most difficult, item 8 is the fourth easiest, while for rater 2, item 6 is the second easiest and item 8 is the second most difficult item. The problems in items 6 and 8 can be solved by looking for a pattern and working backwards problem solving strategies. In item 2, there is a problem requiring an unusual division process, students are expected to evaluate the result in a real-life manner in a division process, so this problem can be considered as the easiest problem for both raters. In item 4, students are expected to find the missing data that is necessary for solving a problem but which is not given to them. The reason why students have difficulty in solving these kinds of problems can be explained by not working with such problems, although emphasis is given to the questions that should be asked in the problem solving stages in textbooks and curriculums. In the 12th item, an authentic problem situation, which is not familiar with the mathematics courses of the students but which is important with Programme for International Student Assessment International (PISA), is used.

DISCUSSION and CONCLUSION

In this study, the scores of gifted middle school students from a 12-question problem solving test related to non-routine mathematical open-ended problems were examined with the Multi Facet Rasch Model in terms of two different raters. As a result of the study, although scoring criteria were determined before starting the scoring process, it was seen that there were differences in the scoring process between the raters. It was found that rater 1 was more tolerant than rater 2 except for one problem in the test. However, according to the raters, the most difficult and easiest questions in the problem solving test also vary. It is known that some of the students' school achievement scores are used in placement process especially in high school and higher education examinations. On the other hand, the school achievement scores vary according to the scoring and opinion of the teachers who are in the same school and differentiate in scoring even if they apply the same test. In such cases where rater bias cannot be prevented, taking necessary cautions plays a vital role for individuals to be placed in appropriate education programs and institutions and to receive training in line with their abilities. However, in this way, qualified human resources of the countries will be evaluated correctly, individuals will be trained in line with their interests and abilities and will be taken into business life. Thus, neither the society nor the individuals will be harmed as a result of the decisions that have been mixed. In order to prevent this risk and to increase the reliability of exams, trainings can be given to teachers to rate measurement instruments consisting of open-ended questions and the score obtained from multiple raters can be used for evaluation (Takunyacı, 2016).

When the findings of the other studies do not overlap with this study, it is concluded that there is no absolute agreement between CTT and IRT based talent estimations (İlhan, 2016). In addition, it was observed that the evaluation based on the MFRM and CTT models did not make any difference in terms of item difficulties of open-ended questions (İlhan & Güler, 2018). While Rasch Analysis cannot reveal the reasons for bias in rating, it shows the sources of bias in terms of raters (Semerci, Semerci & Duman, 2013). Consequently, measures can be taken to prevent rater bias by conducting studies that will reveal the reasons of bias in the future researches. In addition, since it provides the opportunity to examine the measurement results in many respects, it can be recommended to use the assessment tools in the development of measurement tools and in the analysis of examinations in which important decisions are made (Güler, 2014).

As in every research, this study has some limitations, too. One of them is that the study was conducted on only two raters. In further studies, scoring behaviors of more raters can be examined. Similar studies can also be carried out for different educational levels and courses. Based on the practice, it may be suggested to take the necessary measures in the process of using teacher-made tests for important decisions to be taken at every stage of the training, and to evaluate more than one rater instead of one rater for more reliable results.

REFERENCES

Abu Kassim, N. L. (2007, June). Exploring rater judging behaviour using the many-facet Rasch model. Paper Presented in the Second Biennial International Conference on Teaching and Learning of English in Asia: Exploring New Frontiers (TELiA2), Holiday Villa Beach & Spa

Resort, Langkawi. Faculty of Communication and Modern Languages, University Utara Malaysia. Available online at: http://repo.uum.edu.my/3212/1/Noor1.pdf

Altun, M. (2014). *Ortaokullarda (5-8. Sınıf) matematik öğretimi*, Aktüel Yayıncılık, 10. Baskı.

Atılgan, H., Kan, A., & Doğan, N. (2009). *Eğitimde ölçme ve değerlendirme.* Anı Yayıncılık.

Bahar, M., Nartgün, Z., Durmuş, S., & Bıçak, B. (2010). *Geleneksel-tamamlayıcı ölçme ve değerlendirme teknikleri.* Ankara: Pegem Akademi Yayıncılık.

Başol, G. (2013). *Eğitimde ölçme ve değerlendirme.* Ankara: Pegem Akademi Yayıncılık.

Baştürk, R. (2010). Evaluating of research assignments with many facets Rasch measurement model. *Measurement and Evaluation in Education and Psychology*. 1(1), 51-57.

Batdı, V. (2014). Ortaöğretim matematik öğretim programı içeriğinin Rasch ölçme modeli ve Nvivo ile analizi, Turkish Studies, 9(11), s.93-109.

Braun, H.I., Bennett, R.E., Frye, D., & Soloway, E. (1990). Scoring constructed responses using expert systems. *Journal of Educational Measurement*, 27(2), 93-108.

Dreyfus, T. (2007). Processes of abstraction in context the nested epistemic actions model. Available online: http://escalate.org.il/construction_knowledge/papers/dreyfus.pdf

Durmaz, B. (2014). Üstün yetenekli ilköğretim öğrencilerinin problem çözme stratejilerini öğrenme düzeyleri. (*Doktora tezi)*, Uludağ Üniversitesi.

Güler, N. (2014). Analysis of open-ended statistics questions with many facet Rasch model. *Eurasian Journal of Educational Research*, 55, 73-90. http://dx.doi.org/ 10.14689/ejer.2014.55.5

Gürcan Töre, C. (2007). İlköğretim 6. sınıf öğrencilerinin problem çözme sürecini bilme ve uygulama düzeylerinin araştırılması. (*Yüksek lisans tezi),* Eskişehir Osmangazi Üniversitesi.

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

Hershkowitz, R., Schwarz, B. B. & Dreyfus, T. (2001). Abstraction in Contexts: Epistemic Actions, *Journal for Research in Mathematics Education*, 32 (2), 195-222.

Hogan, B. (2004). Singapore math: A problem-solving approach. *Principal (Reston, Va.),84*(2), 22-3.

İlhan, M. (2016). Açık uçlu sorularla yapılan ölçmelerde klasik test kuramı ve çok yüzeyli Rasch modeline göre hesaplanan yetenek kestirimlerinin karşılaştırılması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, *31*(2), 346-368. http://dx.doi.org/10.16986/HUJE.2016015182

İlhan, M., & Güler, N. (2018). A comparison of difficulty indices calculated for open-ended items according to classical test theory and many facet rasch model, *Eurasian Journal of Educational Research*, 75(2018), p.99-114.

İnceçam, B., Demir, E.Ü., & Demir, E. (2018). Ortaokul öğretmenlerinin sınıf içi ölçme ve değerlendirmelerde yazılı yoklamalarda kullandıkları açık uçlu maddeleri hazırlama yeterlikleri, *İlköğretim Online*, 17(4), s. 1912-1927.

Karasar, N. (2008). *Bilimsel Araştırma Yöntemi.* Ankara: Nobel Yayıncılık.

Kaya Uyanık, G., Güler, N., Taşdelen Teker, G. ve Demir, s. (2017). Türkiye'de eğitim alanında yayımlanan ölçek geliştirme çalışmalarının uygunluğunun çok yüzeyli rasch modeli ile incelenmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, *8*(2), 183-199.

Kaya Uyanık, G., Güler, N., Taşdelen Teker, D., & Demir, S. (2019). Fen bilimleri dersi etkinliklerinin çok yüzeyli Rasch modeliyle analizi. *Kastamonu Education Journal, 27*(1), 139-150.

Köse, İ. A., Usta, H. G., & Yandı, A. (2016). Sunum yapma becerilerinin çok yüzeyli rasch analizi ile değerlendirilmesi. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, *16*(4).

Lee, S. K. (1982). Fourth graders' heuristic problem-solving behavior, *Journal for Research in Mathematics Education,* 13 (2).

Linacre, J. M. (1989). *Many-facet Rasch measurement.* Chicago: MESA Press.

Linacre, J. M. (1997). KR-20 or Rasch reliability: Which tells the "truth". *Rasch Measurement Transactions, 11*(3), 580-581.

Lunz, M.E., & Wright, B. D. (1997). *Latent trait models for performance examinations.* In J. Rost & R. Langeheine (Eds.), Applications of latent trait and latent class models in the social sciences (pp. 80-88). Münster, Germany: Waxmann.

Milgram, R. J. (2007). *What is mathematical proficiency?* In A. H. Schoenfeld (Ed.), Assessing mathematical proficiency (pp. 31-58). New York: Cambridge University Press.

NCTM (2000). *Principles and Standards for School Mathematics.* National Council of Teachers of Mathematics, Reston/VA.

Polya, G. 1957. *How to solve it*. New York: Doubleday.

Romagnano, L. (2001). The myth of objectivity in mathematics assessment. *Mathematics Teacher*, 94(1), 31-37.

Rose, T. D. (1991). *Strategies and skills used by middle school students during the solving of nonroutine mathematics problems.* Unpublished EdD. University of Tennessee.

Semerci, Ç. (2011). Mikro öğretim uygulamalarının çok yüzeyli rasch ölçme modeli ile analizi, *Eğitim ve Bilim,* 36(161), s. 14-25.

Semerci, Ç., Semerci, N., & Duman, B. (2013). Yüksek lisans öğrencilerinin seminer sunu performanslarının çok yüzeyli rasch modeli ile analizi, *Sakarya Üniversitesi Eğitim Fakültesi Dergisi*, Sayı 25, 7-22.

Schroeder, T. L. & Lester, F. K. Jr. (1989). *Developing understanding in mathematics via problem solving*. In P. R. Trafton (Ed.), New directions for elementary school mathematics. (pp. 31-42). Reston, VA: NCTM

Şahin, M. G., Taşdelen Teker, G., & Güler, N. (2016). An analysis of peer assessment through many facet rasch model, *Journal of Education and Practice*, Vol 7, No 32.

Takunyacı, M. (2016). *Çoktan seçmeli sorulara dayalı olmayan bir kitle matematik sınavı sürecinin değerlendirilmesi: Grup uyumu değerlendirme modeli*, Doktora Tezi, Marmara Üniversitesi.

Taşpınar, Z. (2011). *İlköğretim 8. sınıf öğrencilerinin matematik dersinde kullandıkları problem çözme stratejilerinin belirlenmesi.* Yüksek Lisans Tezi, Gazi Üniversitesi, Ankara.

Tekin, H. (2009). *Eğitimde ölçme ve değerlendirme.* Ankara: Yargı Yayınevi.

Turgut, M. F., & Baykul, Y. (2012). *Eğitimde ölçme ve değerlendirme,* Ankara: Pegem Akademi Yayıncılık.

Yıldırım, A., & Şimşek, H. (2008). *Nitel araştırma yöntemleri.* Ankara: Seçkin Yayıncılık.

Yıldız, V. (2008*). Polya'nın problem çözme adımlarına dayalı matematik öğretiminden sonra altıncı sınıf öğrencilerinin problem çözme becerileri, problem çözmeye karşı tutumları ve matematiğe karşı tutumlarındaki değişimin incelenmesi.* Yüksek Lisans Tezi, Orta Doğu Teknik Üniversitesi, Ankara.

Yüzüak, A., Yüzüak, B., & Kaptan, F. (2015). Performans görevinin akran gruplar ve öğretmen yaklaşımları doğrultusunda çok-yüzeyli rasch ölçme modeli ile analizi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, *6*(1), 1-11.