OPEN ACCESS
EURASIA Journal of Mathematics Science and Technology Education
ISSN: 1305-8223 (online) 1305-8215 (print)
2017 13(6):2423-2439
DOI 10.12973/eurasia.2017.01233a

Generalizability Theory Research on Developing a Scoring Rubric to Assess Primary School Students' Problem Posing Skills

Osman Cankoy
Atatürk Öğretmen Akademisi, N. CYPRUS
Hasan Özder
Atatürk Öğretmen Akademisi, N. CYPRUS

Received 23 July 2016 ▪ Revised 29 August 2016 ▪ Accepted 4 September 2016

ABSTRACT
The aim of this study is to develop a scoring rubric to assess primary school students' problem posing skills. The rubric including five dimensions namely solvability, reasonability, mathematical structure, context and language was used. The raters scored the students' problem posing skills both with and without the scoring rubric to test the reliability of the rubric. The study used generalizability theory to test the reliability of the scores obtained with and without the use of the rubric. More reliable scoring was obtained using the scoring rubric. The G and phi coefficients rose somewhat after increasing the number of items and raters in both scoring methods. However, increasing the number of items affected these coefficients slightly more than increasing the number of the raters.
Keywords: generalizability theory, assessing problem posing skills, problem posing, scoring rubric

INTRODUCTION

Today, individuals are supposed to get more prepared to cope with the demands of the century since information is increasing rapidly everyday, and it is inevitable for people to include technological, technical, and procedural instruments into the daily life routine. One of the essential elements of being well educated is to be capable of solving problems successfully. In this context, many studies and research findings emphasize the importance of problem solving (Beyazit, 2013; Jonassen, 2000; NCTM, 1989, 1991, 2000, 2004; Polya, 1957, 1973; Schoenfeld, 1985, 1992; Xin, 2007). It is highly probable that problem solving and the importance attributed to it having a large part in the relevant literature has influenced education specialists, researchers and particularly teachers and so led to the inclusion of many problem solving activities in learning environments. Likewise, it would make a greater contribution to learning and reasoning to tackle problem solving activities in a more creative way, especially in the contexts which are non- routine (Sakshaug & Wohlhuter, 2010). The attainments obtained from the routine or non-routine problem solving activities in the school

Correspondence: Hasan Özder, *Atatürk Öğretmen Akademisi, N. CYPRUS.*
hasan.ozder@aoa.edu.tr

State of the literature

- Successful problem posing means successful problem solving. So, problem solving and problem posing are considered as integrated skills that support each other.
- Although many researchers emphasize problem posing, there is a lack of classroom activities in line with problem posing.
- Previous studies have mainly focused on the quantitative aspects of the problems created by students.

Contribution of this paper to the literature

- Problem posing can be assessed more effectively by using an analytical scoring rubric.
- The scoring rubric suggested can especially enable researchers and teachers to assess the qualitative aspects of problems posed by students.
- The generalizability theory is suggested in ensuring the reliability of similar scoring rubrics.

environment can also be used to solve daily life problems (Singer & Voica, 2013). However, these problems (particularly the problems individuals face in daily life) may not be structured neatly as in textbooks. Real life problems are usually complicated, and they generally require description and reformulation. As Kilpatrick (1987) states, most real life problems are created and then solved by individuals. In other words, the members of the dynamic societies today should be able to adapt to new situations that cannot be anticipated, change their jobs and residences and continue their careers in different professions. Being capable of describing and formulating mathematical problems greatly help this type of individual to make good decisions (Singer & Voica, 2013). In other words, there is a need for individuals who can successfully pose problems in addition to those who can successfully solve them. For this reason, many researchers handle problem solving and problem posing as integrated skills that support each other (Cifarelli & Sevim, 2015; Kilpatrick, 1987; Silver, 1993, 1994). "Under certain conditions, problem posing includes students' posing (writing) problems as well as changing the current problem and creating new problems based on it" (Cankoy & Darbaz, 2010, pages 11-12).

Although education specialists and researchers attribute more importance to problem posing today, it has a smaller share in classroom activities than the traditional problem solving practices (Chang, 2007; English, 1998; Lowrie, 1999, 2002a, 2002b; NCTM, 1989, 2000; Silver, 1994, 1995; Stoyanova, 2003). Many researchers argue that problem posing activities make a contribution to students' problem solving skills and mathematical thinking (Silver, 1994; Stoyanova, 2003), which makes it essential for problem posing to have a larger role in classroom learning activities (Akay, 2006; Van Harpen & Presmeg, 2013; Knott, 2010). Although problem posing is becoming more and more important, there are not many studies that specifically focus on the quality of problems posed by students and the ways of reasoning they use when posing problems. For this reason, many education specialists and researchers stress that there is a need for studies analyzing the problem structures created by students (English, 1997, 2003; Kilpatrick, 1987; Singer, Ellerton & Cai, 2013). The current study is

important since it examines the reliability of a scale developed by the researchers to measure especially the qualitative aspects of the problems posed by students.

RELATED LITERATURE

The Importance of Problem Posing

Problem posing is an activity that involves cognitive practices focused on reasoning and that challenges the individual (Cai & Hwang, 2002). Silver (1994) and Stoyanova (2000) claim that problem posing can be tackled in five different ways: (1) posing a problem in a free way, (2) posing a problem with an answer already given to it, (3) posing a problem based on certain conditions, (4) creating questions about a problem situation, and (5) posing a problem based on the given mathematical operations in it.

Posing a problem under certain conditions or reorganizing an existing problem requires a considerable degree of cognitive effort and positively affects mathematical development (Cai et al., 2013; Doyle, 1983). Problem posing is not only a learning activity but it also improves students' conceptual understanding, enhances their mathematical communication skills, interests them in mathematics and their environment and gives them the opportunity to use creativity (NCTM, 1991). From another perspective, problem posing can be described as the effort to analyze and solve an existing problem by tackling it in sub-problems (Polya, 1957). In this context, successful problem posing also means successful problem solving (Cai & Hwang, 2002).

Assessing Problem Posing

Problem posing is very important in mathematics and in real life, which highlights the need to assess this skill effectively. A review of the relevant literature on problem posing skills reveals that the most obvious question is whether or not a posed problem is solvable (English, 1997; Ergün, Gürel and Çorlu, 2011; Kilpatrick, 1987; Leung & Silver, 1997; Silver & Cai, 1996). Leung and Silver (1997) conducted a study which stresses the solvability of a problem that is posed based on the third dimension ("impossible to solve") and the fourth dimension ("insufficient information") of a five-dimension scale which was created to classify the problems posed by students. Although a posed problem may look solvable using the given information, the information given and the answer may be unreasonable and have a structure that cannot be used in real life. For this reason, many researchers indicate that the reasonability of a posed problem should also be measured (Cai et al., 2013; English, 1998; Mestre, 2002; Silver, 1994). Mathematical structure, or the algebraic format, of a problem may make it more difficult to solve (Koedinger & Nathan, 2004). In this context, the mathematical structure of a problem can be used as an important measure for the quality of that problem (Işık & Kar, 2015; Koedinger & Nathan, 2004; Riley & Greeno, 1998; Van Harpen & Presmeg, 2013). Relevant studies indicate that posed problems can also be assessed by their context (Singer, Ellerton & Cai, 2013; Van Harpen & Presmeg, 2013) and the language used in the problem (English, 1998; Van Harpen & Presmeg, 2013) in addition to their quantitative and solution-based aspects.

Briefly, a posed problem can be assessed for its (1) solvability, (2) reasonability, (3) mathematical structure, (4) context and (5) language. The current study aims to determine the reliability of a scale that uses this rubric to assess posed problems in qualitative terms.

The relevant studies found that using scoring rubric reliably assess the peculiarities of human being such as problem solving skill (Büyükkıdık, 2012; Hızarcıoğlu, 2013; Sefer, 2006), writing composition skill (Beyreli and Arı, 2009; Çetin and Kelecioğlu, 2004; Ömür and Erkuş, 2013; Novak, Herman and Gearhart, 1996), achievement on written examinations administered in the Turkish Language and Literature course (Kan, 2005), open-ended mathematical problems (Güler and Gelbal, 2010), proportional reasoning skill (Akkuş and Paksu, 2006) graphic reading skill (Atmaz, 2009), achievement in the Science course (Aytaç, 2006; Eser and Gelbal, 2013; Marzano, 2002), English writing skill (Kayapınar, 2014), achievement on grammar and reading comprehension (Kan, 2007) and cello playing performance (Birel, 2014).

There are few studies focused on the assessment of problem posing skills. Işık and Kar (2015) analyzed the semi-structured problems created by sixth graders, and in a similar way to the current study, they examined the main mathematical structure and language dimensions of the problems. Van Harpen and Presmeg (2013) created a rubric to assess the problem posing skills of American and Chinese high school students, and similarly, tried to describe routine and non-routine problem contexts.

Cai et al. (2013) used a scoring rubric to assess students' problem posing skills with the purpose of revealing the effect of the middle school mathematics curriculum, and the two dimensions of that rubric are similar to the current study's solvability and context dimensions. Ergün, Gürel and Çorlu (2011) analyzed the understandability of problems, problems' suitability with physics principles, problems' structures, the number of the questions asked, types of problems and their solvability dimensions and also developed a rubric. They used classical test theory to test their scoring rubric's reliability. The current study determined the problem posing dimensions in a new and a different way than the way Ergün, Gürel and Çorlu (2011) utilized, and used generalizability theory as it anticipated reliability better. G studies are conducted to classify measurement errors by specific sources of variation (Brennan, 2001; Crocker & Algina, 1986; Güler, Uyanık & Teker, 2012; Shavelson & Webb, 1991).

> *Estimating different sources of error variance allows researchers to better understand how the features of their measurement system contribute to the deviation in observed scores from the true score. They may then use information about the sources of error variance to make decisions about how to decrease the amount of error associated with different measurement facets in future studies (e.g., increasing the number of observers or number of sessions used to compute scores) (Bottema-Beutel, Lloyd, Carter & Asmus, 2014).*

Generalizability theory is usually performed in two stages while conducting research.The first stage is Generalizability (G) Study. At this stage reserachers work on to

what extend the observations represent or desciribe the population. Second stage is the Decision (D) Study where the observations are used to get the most reliable measures and to decide on the best measurement desing (Easton,1989). "In the G-Theory framework, the object of measurement can be crossed with different facets" (Naumenko, 2015: 5). In the current study, a two facet individual x item x rater (in x i x r) desing was used. In the current study since the problems posed by each student are scored by each rater at this measurement process all variance sources are crossed with each other. In such desings the variance components in line with scorer variance sources are expected to be close to zero. This reveals the consistency amongst the raters (Güler, Uyanık ve Teker, 2012).

The current study aimed to determine the changes in the reliability level of primary school students' problem posing skill scores given by same raters with and without the use of the developed scoring rubric. The current study also examined to what extent more raters and items affect reliability.

METHODOLOGY

The current study examined the reliability of assessing primary school students' problem posing skills with and without the use of the developed scoring rubric. Three pre-service teachers scored the problems posed by 25 students (three problems by each student) based on the free problem posing approach without using the scoring rubric and then using the rubric developed by the researchers. Then, the current study compared the reliability of the scores using both approaches. Generalizability theory was used to determine reliability.

Study Sample

The sample of the current study included the fifth graders studying in Çağlayan Cumhuriyet Primary School in Nicosia, Northern Cyprus, in the 2012-2013 academic year. The school is located in a district where the families of the students are at an intermediate socioeconomic level. The three pre-service teachers who scored the students' problem posing were also included in the current study sample.

The current study data were collected using two different methods. Each of the 25 students posed three problems using the free problem posing approach (in other words, any way they wanted), and the posed problems were scored by the same raters, first without the scoring rubric and then with the scoring rubric. The researchers gave a briefing to the raters on how to use the scoring rubric they created to score the problems. The researchers took being easily accessible and having high motivation into consideration when selecting the pre-service teachers as raters.

Preparation and Development of the Scoring Rubric

In the current study, 25 primary school students were asked to pose three questions any way they desired. The problems posed by the students were scored by three pre-service teachers without the scoring rubric (as shown in **Table 1**) and then using the developed rubric.

Table 1. Scoring Rubric for Problem Posing Skill

CATEGORY	SUB-CATEGORY	EXPLANATION	SCORE
Solvability	Solvable	The information given in the problem is sufficient to solve the problem and find the solution.	1
	Unsolvable	The information given in the problem is not sufficient to solve the problem and find the solution.	0
Reasonability	Reasonable	The information given in the problem and the solution is reasonable and applicable in real life.	1
	Unreasonable	The information given in the problem and the solution is not reasonable and applicable in real life.	0
Mathematical Structure	Result unknown model	The unknown element of the problem is at the end. (arithmetic)	0
	Start unknown model	The unknown element of the problem is at the beginning (algebra)	1
Context	Routine	The subject handled by the problem is in a form frequently used by teachers in classrooms, and in a structure frequently seen in textbooks.	0
	Non-routine	The subject handled by the problem is distant from the forms used by teachers in classrooms, and in a unique structure that is rarely seen in textbooks.	1
Language	Clarity-Understandability	The language used in the problem is very clear, understandable and fluent.	1
		The language used in the problem is not clear, understandable and fluent.	0
	Obeying grammar rules	The problem completely obeyed the grammar rules to express the question.	1
		The problem partly obeyed or did not obey grammar rules at all to express the question.	0

The five-dimensional scoring rubric for problem posing skills is based on the three-dimensional scale developed by Cankoy (2014) with the purpose of measuring the problem posing skills of the students in the sample. In the development of the scoring rubric, the researchers evaluated the steps suggested by Beng (2008), Taggart, Phifer, Nixon and Wood (1998), Goodrich (2000) and Nitko (2009), and followed these stages:

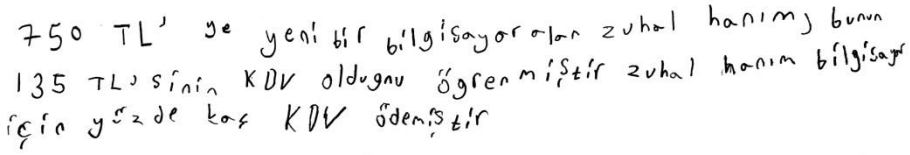
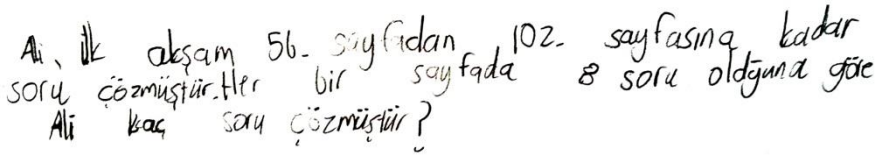
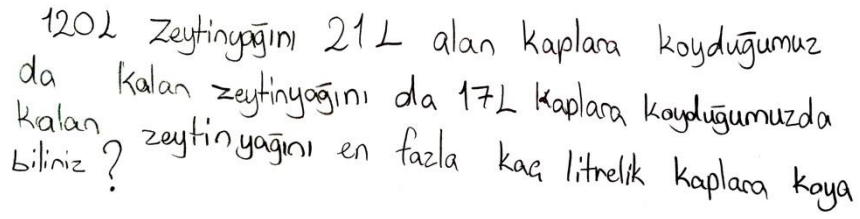
1. The problem posing skills expected from the students for the problem posing skill were determined based on a review of the relevant literature. In this context, the current study asserted (1) solvability, (2) reasonability, (3) mathematical structure, (4) context and (5) language dimensions.

2. After determining the dimensions, the researchers also identified the sub-dimensions of each dimension and decided that the scoring could be done using one and zero points.
3. For the current study's scoring rubric, the researchers consulted the opinions of mathematics teaching experts and assessment and evaluation experts.
4. The pre-service teachers conducted a pilot study with the scoring rubric and then revised it.

Process

1. Scoring without the rubric: Three pre-service teachers independently scored the problems posed by 25 fifth graders using a holistic approach, where the pre-service teachers scored the students' problem posing skills without using a rubric. In this practice, the three problems posed by each student were scored separately. The pre-service teachers were asked to score each question from 0 to 6, and the total score of each student was calculated by adding these points.
2. Scoring with the rubric: Three pre-service teachers independently scored the problem posing skills of 25 fifth graders using the scoring rubric developed by the researchers. A review of the relevant literature on problem posing indicated that a posed problem should be evaluated for its: (1) solvability, (2) reasonability, (3) mathematical structure, (4) context and (5) language. Accordingly, the scoring rubric was developed as shown in **Table 1**. Thus, scores for any problem posed could be ranged from a minimum score of 0 to a maximum score of 6. The researchers added the points given by the raters for the three problems posed by each student and calculated the means. See **Table 2** for sample problems posed by the students. **Figure 1** shows a sample problem scored by a rater using the rubric. Expression in the boxes are English translations of the hand writings.

Table 2. Sample Problems Posed by Students

Sample Problem	Sub-Categories
 *Mrs. Zuhal, bought a new computer with a price of 750 TL. She noticed that 135 TL of the price was VAT. What percent of the price is VAT?*	Solvable Reasonable Start-Unknown Non-Routine Clear Grammatically Good
 *In the afternoon, Ali solved the questions included in the pages from 56th to 102nd. If there were 8 questions in each page, how many questions did Ali solved?*	Solvable Unreasonable Result-Unknown Routine Clear Grammatically Poor
 *If we put 120 liter oliveoil into the cups each with a capacity of 21 liters and then put the rest to the cups each with a capacity of 17 liters, then at most how many cups are needed to put the rest of the oliveoil?*	Unsolvable Unreasonable Result-Unknown Non-Routine Not Clear Grammatically Poor

The sentences in italics are the translated versions of the problems posed.

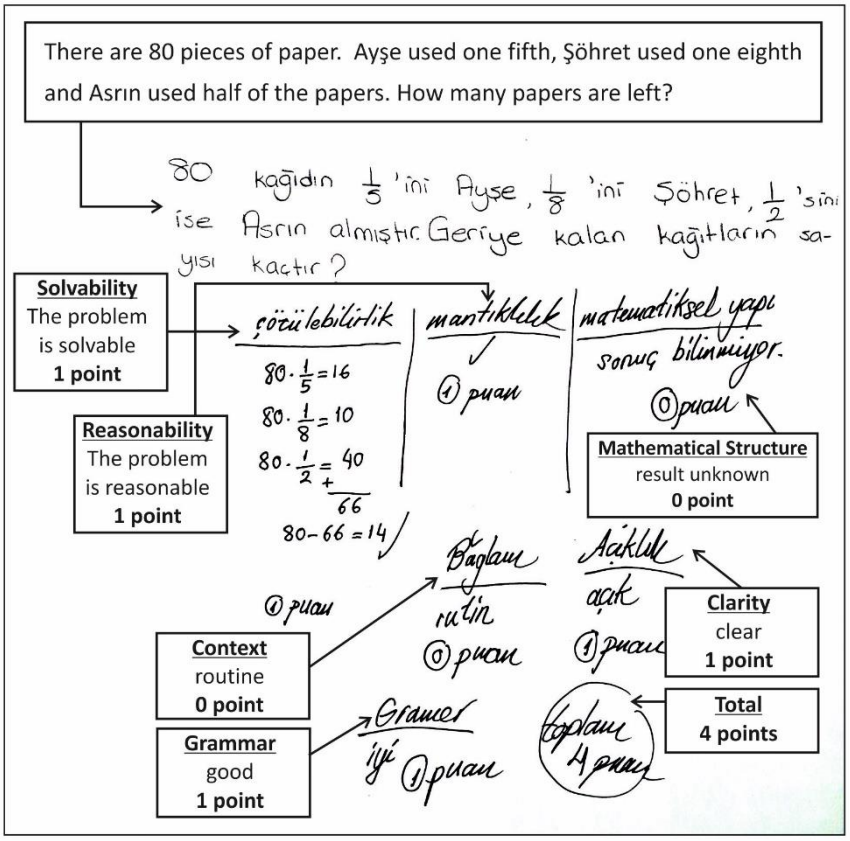


Figure 1. A sample problem scored by a rater using the rubric

Data Analysis

The data were analyzed using the Edu-G program in conjunction with generalizability theory using the variance components found by the current study. Using the generalizability theory, which is mainly based on ANOVA, it is possible to find the percentage of the total variance by the sources of variation described in the study collectively and individually. The current study calculated the generalizability and reliability coefficients using completely crossed patterns with two variables (individual x item x rater). The current study also did a K-study based on generalizability theory addressed to different alternatives, and calculated the G and Φ (phi) coefficients for the reliability of the scoring.

FINDINGS AND DISCUSSION

Scoring Without the Scoring Rubric

Table 3 shows the variance components and the percentages of explaining total variance of the results of scoring 25 students' problem posing skills without using a rubric by three pre-service teachers. It is found that the highest variance was 24.4% by the raters, followed by 16.9% by individuals and concluded with 0.0% by the items.

Table 3. The Variance Components Anticipated by A (in x i x r) Pattern G-Study of the Scores Given Without Using Rubric and Their Percentages of Explaining the Total Variance

	Dimensions				
Source	**SS**	**df**	**MS**	**Corrected**	**%**
IN (Indiviual)	141.084	24	5.88	0.35	16.9
I (Item)	0.44	2	0.22	-0.02	0.0
R (Rater)	78.25	2	39.12	0.51	24.4
IN-I	100.01	48	2.08	0.53	25.4
IN-R	52.86	48	1.10	0.20	9.7
IR	0.98	4	0.24	-0.01	0.0
IN-I-R	47.24	96	0.49	0.49	23.5

Table 4. The Variance Components Anticipated by a (in x i x r) Pattern G Study of the Scores Given Using the Rubric and Their Percentages of Explaining the Total Variance

	Dimensions				
Source	**SS**	**df**	**MS**	**Corrected**	**%**
IN (Indiviual)	140.56	24	5.86	0.44	31.6
I (Item)	10.11	2	5.05	0.04	3.4
R (Rater)	1.15	2	0.57	0.00	0.1
IN-I	76.56	48	1.59	0.38	26.9
IN-R	33.52	48	0.69	0.08	5.8
IR	1.07	4	0.26	-0.00	0.0
IN-I-R	43.60	96	0.45	0.45	32.2

Individual x item (in x i) common effect explained 24.4% of the total variance. This finding shows that raters give different scores to different items when they do not use a scoring rubric. Individual x rater (in x r) common effect explained 9.7% of the total variance. Thus, raters did not give very different scores to different individuals. Item x rater common (i x r) effect explained 0.0% of the total variance. This shows that raters did not give different scores to items, but gave scores close to each other. Individual x item x rater (in x i x r) common effect explained 23.5% of the total variance.

The reliability coefficient, which was calculated according to the generalizability theory, was found to be 0.43. This finding shows that scoring without a rubric has low reliability in comparison to scoring with the developed rubric.

The Reliability of Scoring with the Rubric

Table 4 shows the variance components and the percentages of explaining total variance of the results of scoring 25 students' problem posing skills using a rubric by three pre-service teachers. It is found that the highest variance was 31.6% by the students, items had 3.4% and raters explained the least variance with 0.1%.

The variance component of the students explains the total variance at a high rate, which shows that students differ in their problem posing skills. This is consistent with the findings of the studies by Büyükkıdık (2012) and Kan (2007). However, it differs from the current study

findings of Eser and Gelbal (2013). Eser and Gelbal (2013) found a higher variance for the item variable. The variance component anticipated for the item variable had a low percentage of explaining the total variance, which means that the three problems posed by each student were at the same level and did not differ. The variance component anticipated for the raters had a very low percentage of explaining the total variance, which means that the consistency among the raters was very strong. This finding is similar to the findings of Büyükkıdık (2012) and Güler and Gelbal (2010).

Individual x item common effect explained 26.9% of the total variance. This shows that the interaction between the student and the question is an indicator of the change in students' performances in each question. Therefore, students' performances differed by question. This finding is not consistent with the findings by Büyükkıdık (2012), Eser and Gelbal (2013), and Kan (2007). Individual x rater common effect explained 5.8% of the total variance. This shows that raters did not give very different scores to the problems. In other words, the raters' scores did not differ by student. This shows the difference between the individuals regarding their performance. Thus, individual differences can be determined using the scoring rubric. Item x rater common effect explained 0.0% of the total variance, which shows that raters did not score the items differently, but gave similar scores. This finding is similar to the research findings of Kan (2007). Individual x item x rater collectively explained 32.2% of the total variance. This rate may be an indicator of the fact that individual x item x rater effect and/or coincidental errors may be on a large scale. This finding is consistent with the research findings of Eser and Gelbal (2013) and Kan (2007). It is desirable for the variance of the residue component to be as low as possible (Güler, Uyanık and Teker, 2012: 76). The reliability coefficient calculated according to the generalizability theory was found to be 0.67. This shows that scoring with a rubric is more reliable. This finding is consistent with other research findings (Novak, Herman and Gearhart, 1996).

Reliability and Phi Coefficients Anticipated by Alternative Decision Studies

In the (in x i x r) pattern used in the current study, three raters scored each of 25 students using a rubric based on three items, and it was anticipated that the G coefficient was 0.69 and the Phi coefficient was 0.67. **Table 5** shows that increasing and reducing the number of items affected the G and Phi coefficients more than increasing and reducing the number of raters. When the number of items is stable, and one item is added (p=3, m=4), the G coefficient is 0.74, an increase of 0.05. When the item number remains the same, and one rater is added (m=3, p=4), the G coefficient is 0.71, an increase of 0.02. This shows that increasing the number of items leads to an increase in reliability. For instance, when the number of raters is stable (p=3) and the number of items is increased to 7, the G coefficient reaches 0.81.

Table 5. G and Phi Coefficients Anticipated by Alternative K Study

Number of raters	Number of items	Alternative K studies			
		With the rubric		Without the rubric	
		G	Phi	G	Phi
1	3	0.55	0.54	0.39	0.25
2	3	0.65	0.63	0.50	0.37
3	3	0.69	0.66	0.54	0.42
3	4	0.73	0.72	0.59	0.46
3	5	0.76	0.75	0.63	0.48
3	6	0.79	0.78	0.65	0.50
3	7	0.81	0.80	0.68	0.51
4	3	0.70	0.68	0.56	0.47
5	3	0.72	0.70	0.58	0.50

Scoring without a rubric gives the same result. When the number of raters is stable, and one item is added (p=3, m=4), the G coefficient increases from 0.54 to 0.59. When the number of items is stable, and one rater is added (m=3, p=4), the G coefficient is 0.56. In this case, increasing the number of items also leads to an increase in reliability. For instance, when the number of raters is stable (p=3), and the number of items is increased to 7, the G coefficient reaches 0.68. When a rubric is used for measurement, raters explain 0.1 of total variance, and the item dimension is 3.4. Therefore, increasing the number of items increases reliability more effectively. These findings are also consistent with the findings of Güler and Gelbal (2010). However, they contrast with the findings of Büyükkıdık (2012). According to Büyükkıdık (2012), increasing or reducing the number of raters by 2 had a greater effect on G and Phi coefficients than increasing or reducing the number of tasks.

CONCLUSIONS AND SUGGESTIONS

Conclusions

The current study aimed to determine the reliability of scoring fifth graders' problem posing skills with and without the use of scoring rubric, and it was found that scoring with a rubric was more reliable. Additionally, it is possible to say that the use of a scoring rubric increases inter-rater reliability as well as revealing the differences amongst individuals (students). In both scoring methods, the G and Phi coefficients partially increased when the number of items and raters were increased. However, increasing the number of items increases the coefficients slightly more effectively than increasing the raters. Since the items written by the students were similar, it was concluded that increasing the number of items does not increase reliability.

Suggestions

Using the scoring rubrics at performance evaluation is particularly important in terms of reliability. Scoring rubrics are necessary especially in measuring students' performance based on skills such as problem posing. The scoring rubric developed by the researchers of the current study can be used to assess problem posing skills. When measuring students' problem posing skills, students should be asked to pose more than seven problems to increase the reliability of the measurement items. Future studies can be conducted considering other problem posing models like semi-structured and structured problem posing. Future studies should investigate the effectiveness of using the rubric developed in the current study in teaching problem posing.

REFERENCES

Akay, H. (2006). Problem kurma yaklaşımı ile yapılan matematik öğretiminin öğrencilerin akademik başarısı, problem çözme becerisi ve yaratıcılıkları üzerindeki etkisinin incelenmesi. *Ankara: Gazi Üniversitesi, Eğitim Bilimleri Enstitüsü*, (yayınlanmamış doktora tezi).

Akkuş, O., & Duatepe-Paksu, A. (2006). Orantısal akıl yürütme becerisi testi ve teste yönelik dereceli puanlama anahtarı geliştirilmesi. *Eğitim Araştırmaları, 6*, 25, 1-10.

Atmaz, G. (2009). Puanlama yönergesi (rubric) kullanılması durumunda puanlayıcı güvenirliğinin incelenmesi. *Mersin Üniversitesi Sosyal Bilimler Enstitüsü*, yayınlanmamış yüksek lisans tezi.

Aytaç, N. N. (2006). Üniversite öğrencilerinin Newton'un hareket yasalarını anlamalarının değerlendirilmesinde dereceli puanlama anahtarı geliştirilmesi ve kullanımı. Balıkesir: *Balıkesir Üniversitesi, Fen Bilimleri Enstitüsü*, Yayınlanmamış Yüksek Lisans tezi.

Beng, C. S. (2012). Rubrics: beyond scoring, an enabler of deeper learning. *Assessing Student Learning, 15*(3), 1-4.

Beyazit, İ. (2013). An investigation of problem solving approaches, strategies, and models used by the 7th and 8th grade students when solving real-world problems. *Educational Sciences: Theory & Practice, 13*(3), 1920-1927.

Beyreli, L. & Arı, G. (2009). Yazma performansını değerlendirmede çözümleyici puanlama yönergesi kullanımı -değerlendirmeciler arası uyum araştırması-. *Kuram ve Uygulamada Eğitim Bilimleri, 9*(1), 85-125.

Birel, S. A. (2014). Viyolonsel öğretimde performansı değerlendirmeye yönelik hazırlanan puanlama anahtarının (Scoring Rubric) sınanması ve değerlendirilmesi. Ankara: *Gazi Üniversitesi, Eğitim Bilimleri Enstitüsü*, Yayınlanmamış Doktora Tezi

Bottema-Beutel, K., Lloyd, B., Carter, E. W., & Asmus, J. M. (2014). Generalizability and decision studies to inform observational and experimental research in classroom settings. *American Journal on Intellectual and Development Disabilitie*, *119*(6), 589-605.

Brennan, R. L. (2001). *Generalizability theory*. New York: Springer-Verlag.

Brookhart, S. M. (2013). *How to create and use rubrics for formative assessment and grading*. ASCD. Alexandria,VA.

Büyükkıdık, S. (2012). Problem çözme becerisinin değerlendirilmesinde puanlayıcılar arası güvenirliğin klasik test kuramı ve genellenebilirlik kuramına göre karşılaştırılması. (Yayinlanmamış Yüksek Lisans Tezi). Hacettepe Üniversitesi. Ankara

Işık, C., & Kar, T. (2015). Altıncı sınıf öğrencilerinin kesirlerle ilgili açık-uçlu sözel hikayeye yönelik kurdukları problemlerin incelenmesi. *Turkish Journal of Computer and Mathematics Education, 6*(2), 230-249.

Cai, J., & Hwang, S. (2002). Generalized and generative thinking in U.S. and Chinese students'mathematical problem solving and problem posing. *Journal of Mathematical Behavior*, *21*(4), 401-421.

Cai, J., Moyer, J. C., Wang, N., Hwang, S., Nie, B., & Garber, T. (2013). Mathematical problem posing as a measure of curricular effect on students' learning. *Educational Studies in Mathematics*, *83*(1), 57-69.

Cankoy, O., & Darbaz, S. (2010). Problem kurma temelli problem çözme öğretiminin problemi anlama başarısına etkisi. *Hacettepe University Education Faculty Journal, 38*, 11-24.

Cankoy, O. (2014). Interlocked Problem Posing and Children's Problem Posing Performance in Free Structured Situations. *International Journal of Science and Mathematics Education, 12,* 219-238.

Chang, N. (2007). Responsibilities of a teacher in a harmonic cycle of problem solving and problem posing. *Early Childhood Education Journal, 34*(4), 265-271.

Cifarelli, V. V. & Sevim, V. (2015). Problem posing as reformulation and sense-making within problem solving, in F. M. Singer, N. F. Ellerton & J. Cai (Eds), Mathematical Problem Posing: From Research to Effective Practice, New York, NY: Springer, pp. 177-194.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Harcourt Brace Javanovich College Publishers, USA.

Çetin, B., & Kelecioğlu, H. (2004). Kompozisyon tipi sınavlarda kompozisyonun biçimsel özelliklerinden kestirilen puanların anahtarla ve genel izlenimle elde edilen puanlarla ilişkisi. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, *26*, 19-26.

Doyle, W. (1983). Academic work. *Review of Educational Research*, *53*, 159-199.

Eason, S. H. (1989). Why generalizability theory yields better results than classical test theory. *Mid- South Educational Research Association Annual Meeting*: 8-10 November 1989- Little Rock, AR.

English, L. D. (1997). Promoting a problem-posing classroom. *Teaching Children Mathematics, 3*, 172-179.

English, L. D. (1998). Children's problem posing within formal and informal contexts. *Journal for Research in Mathematics Education, 29*(1), 83-106.

Ergün, H., Gürel, Z., & Çorlu, M. A. (2011). Problem tasarlama performansının değerlendirilmesinde kullanılacak bir rubriğin geliştirilmesine ilişkin bir araştırma. *Milli Eğitim*, *191*, 39-55.

Eser, Ç. D., & Gelbal, S. (2013). Genellenebilirlik kuramı ve lojistik regresyona dayalı hesaplanan puanlayıcılar arası tutarlılığın karşılaştırılması. *Kastamonu Üniversitesi Eğitim Fakültesi Dergisi, 21*(2), 421-438).

Güler, N. & Gelbal, S. (2010). Açık uçlu matematik sorularının güvenirliğinin klasik test kuramı ve genellenebilirlik kuramına göre incelenmesi. *Kuram ve Uygulamada Eğitim Bilimleri, 10*(2), 989-1019.

Güler, N., Uyanık, G. K. & Teker, G. T. (2012). *Genellenebilirlik kuramı*. Ankara: PegemA yayınları.

Goodrich, Andrade, H. (2000). Using rubrics to promopte thinking and learning. *Educational Leadership*, *57*(5), 13-18.

Hızarcıoğlu, Ö. B. (2013). Problem çözme sürecinde dereceli puanlama anahtarı (rubrik) kullanımında puanlayıcı uyumunun incelenmesi. Bolu: *Abant İzzet Baysal Üniversitesi, Eğitim Bilimleri Enstitüsü*, Yayınlanmamış Yüksek lisans Tezi.

Jonassen, D. H. (2000). Toward a design theory of problem solving. *Educational Technology: Research & Development*, *48*(4), 63-85.

Kan, A. (2005). Yazılı yoklamaların puanlanmasında puanlama cetveli ve yanıt anahtarı kullanımının (aynı) puanlayıcı güvenirliğine etkisi. *Eğitim Araştırmaları Degisi*, *20*, 166-177.

Kan, A. (2007). Effects of using a scoring guide on essay scores: generalizability theory. *Perceptual and Motor skills*, 105, 891-905.

Kayapınar, U. (2014). Measuring essay assessment: Intra-rater and inter-rater reliability. *Eurasian Journal of Educational Research*, 57, 113-136

Kilpatrick, J. (1987). Problem formulating: where do good problems come from? In A. H. Schoenfeld (Ed.), *Cognitive Science and Mathematics Education,* NJ: Lawrence Erlbaum Associates, pp. 123–147.

Knott, L. (2010). Problem posing from the foundations of mathematics. *The Montana Mathematics Enthusiast, 7*(2–3), 413–432.

Leung S. K., & Silver, E. A. (1997). The role of task format, mathematics knowledge, and creative thinking on the arithmetic problem posing of prospective elementary school teachers. *Mathematics Education Research Journal, 9*(1), 5-24.

Lowrie, T. (1999). Free Problem Posing: Year 3/4 students constructing problems for friends to solve, in J. Truran& K. Truran (Eds.), *Making a Difference*, 328-335. Panorama, South Australia: Mathematics Education Research Group of Australasia.

Lowrie, T. (2002a). Designing a framework for problem posing: young children generating open-ended tasks. *Contemporary Issues in Early Childhood, 3*(3), 354-364.

Lowrie, T. (2002b). Young children posing problems: The influence of teacher intervention on the type of problems children pose. *Mathematics Education Research Journal, 14*(2), 87-98.

Marzano, R. J. (2002). A comparison of selected methods of scoring classroom assessment. *Applied Measurement in Education*, *15*(3), 249-267.

Mestre, P. J. (2002). Probing adults' conceptual understanding and transfer of learning via problem posing. *Applied Developmental Psychology, 23*, 9-50.

Mertler, C. A. (2001). Designing scoring rubrics for your classroom. Practical Assessment, Research & Evaluation, 7(25). (2014, June). Retrieved from http://PAREonline.net/getvn.asp?v=7&n=25.

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics,* Reston, VA: Author.

National Council of Teachers of Mathematics. (1991). *Professional standards for teaching mathematics*, Reston, VA: Author.

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

National Council of Teachers of Mathematics (NCTM). (2004). *Teaching children mathematics*. (2014, October 16). Retrieved from http://my.nctm.org/eresources/article_summary.asp?URI=TCM2005-04-3a&from=B

Nitko, A. J. (2001). *Educational assessment of students* (3rd ed.). Upper Saddle River, NJ: Merrill.

Naumenko, O. (2015). Improving performance assessment score validation practices: an instructional module on generalizability theory. *Working Papers on Language and Diversity in Education*, I(1), 1-17.

Novak, J. R., Herman, J. L. & Gearhart, M. (1996). Establishing validity for performance-based assessments: an illustration for collections of student writing. *The Journal of Educational Research*, *89*(4), 221-233.

Ömür, S. & Erkuş, A. (2013). Dereceli puanlama anahtarıyla, genel izlenimle ve ikili karşılaştırmalar yöntemiyle yapılan değerlendirmelerin karşılaştırılması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, *28*(2), 308-320.

Polya, G. (1957). *How to solve it?* (2nd ed.). Princeton, N.J.: Princeton University Press.

Polya, G. (1973). *How to solve it: A new aspect of mathematical method* (2nd ed.). Princeton, NJ: Princeton University Press.

Popham, W. J. (2003). *Test better, teach better: the instructional role of assessment*. Alexandria, Virginia: Association for Supervision and Curriculum Development.

Riley, M. S., & Greeno, J. G. (1988). Developmental analysis of understanding language about quantities and of solving problems. *Cognition and Instruction*, *5*(1), 49-101.

Sakshaug, L., & Wholhuter, K. (2010). Journey toward teaching mathematics through problem solving. *School Science and Mathematics*, *110*(8), 397-409.

Schoenfeld, A. H. (1985). *Mathematical problem solving*. Orlando, FL: Academic Press.

Schoenfeld, A. H. (1992). Learning to think mathematically: problem solving, metacognition, and sense making in mathematics. In D. Grouvs (Ed.), *Handbook for research on mathematics teaching and learning* (pp. 334-370). New York: MacMillan.

Sefer, G. D. (2006). Matematik dersinde problem çözme becerilerinin dereceli puanlama anahtarı kullanılarak değerlendirilmesi, *Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü*, yayınlanmamış yüksek lisans tezi.

Silver E. (1993) On mathematical problem posing. In - Hirabayasshi, N. Nohda, K. Shigematsu, & F.L Lin (Ed) proceedings of the seventeenth international conference for the psychology of mathematics education (V. I., PP. 66-85) Tsukuba (Japan): international group for the psychology in mathematics education.

Silver, E. A. (1994). On mathematical problem posing. *For the Learning of Mathematics, 14*(1), 19-28.

Silver, E. A. (1995) The nature and use of open problems in mathematics education: mathematical and pedagogical perspectives. *International Reviews on Mathematical Education*, *27*, 67-72.

Singer, F. M., & Voica, C. (2013). A problem-solving conceptual framework and its implications in designing problem-posing tasks. *Educational Studies in Mathematics*, *83*, 9-26.

Singer, F. M., Ellerton, N., & Cai, J. (2013). Problem posing research in mathematics education: new questions and directions. *Educational Studies in Mathematics*, *83*, 1-7.

Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory*: A primer. Newbury Park, CA: Sage.

Stoyanova, E. (2000). Empowering students' problem solving via problem posing: The art of framing "Good" questions. *Australian Mathematics Teacher,* 56(1), 33-37.

Stoyanova, E. (2003). Extending students' understanding of mathematics via problem posing. *Australian Mathematics Teacher*, *59*(2), 32-40.

Szetela, W., & Nicol, C. (1992). Evaluating problem-solving in mathematics. *Educational Leadership*, 49(8), 42–45.

Taggart, G. L., Phifer, S. J., Nixon, J. A., & Wood, M. (1998). *Rubrics: a handbook for construction and use*. Lancaster, PA: Technomic Publishing Company, Inc.

Xin, P. Y. (2007). Word problem solving tasks in textbooks and their relation to student performance. *The Journal of Educational Research, 100*(6), 347-359.

Van Harpen, X. Y., & Presmeg, N. (2013). An investigation of relationships between students' mathematical problem-posing abilities and their mathematical content knowledge. *Educational Studies in Mathematics, 83* (1), 117–132.

http://iserjournals.com/journals/eurasia