International Online Journal of Educational Sciences, 2016, 8 (1), 8-25

iojes
www.iojes.net

International Online Journal of Educational Sciences

ISSN: 1309-2707

Developing the Rubric for Evaluating Problem Posing (REPP)*

Yasemin Kaba[1], **and Sare Şengül**[2]

[1]*Kocaeli University, Turkey;* [2]*Marmara University, Turkey*

ARTICLE INFO

Article History:
Received 08.04.2015
Received in revised form 01.09.2015
Accepted 11.09.2015
Available online 20.12.2015

ABSTRACT

Problem posing means either posing new problem situations in terms of given conditions or re-formulating an already written problem. It is also equally important to evaluate the posed problems and rubrics can be used during these evaluations. Rubrics are typical graded scoring keys used for assessing the studies of students. In this sense, the purpose of this study is to develop a rubric which can be used to evaluate the problems posed by students. In accordance with this purpose, the study was carried out with 29, 7th grade students who are studying in a middle school. The data collection tools are; worksheets, discussions during lessons and related literature. The data were descriptively analyzed. As a result of the analysis, an analytical rubric composed of four dimensions was developed. The dimensions of the rubric are as in the following; the text of the problem (language and expression), the compatibility of the problem with the mathematical principles, the type/structure of the problem and the solvability of the problem. For the reliability analysis of the study while the inter-rater agreement index was found as 0,720; the Kappa value 0,700; Pearson correlation coefficient r=0,873; Spearman correlation coefficient r=0,865 and Cronbach alfa coefficient 0,932; the intra-rater agreement index was found as 0,830; Kappa value 0,817; Pearson correlation coefficient r=0,921; Spearman correlation coefficient r=0,912 and Cronbach alfa coefficient 0,959.

© 2016 IOJES. All rights reserved

Keywords:
Problem posing, evaluation, rubric, validity, reliability

Introduction

Studies related to problem posing research in mathematics education have been on the rise, in the recent years. As a consequence, this topic was considered as an important area of exploration for researchers. Problem posing, among the issues that educators have been studying, in the recent years, includes either posing new problems in terms of a given situation or re-formulating the problem. Problem posing, which is cognitively more difficult than solving problems, takes place in the field of teaching mathematics as a recently discovered instrument for analyzing cognitive processes (Mestre, 2002).

Problem posing which provides analytical thinking (Akay & Boz, 2010). Students are faced with complex situations in problem posing processes and they feel responsible themselves from these situations (Gür & Korkmaz, 2003). It helps their individual learning (Işık, Kar, Yalçın & Zehir, 2011). Besides, in addition to improving their problem solving skills, it also improves their skills for recognizing key factors given in a problem (Chang, Wu, Weng & Sung, 2012). The importance of problem posing was stated by Lavy and Bershadsky (2002) as in the following; *'problem posing is more important than problem solving which is a part of mathematical or experimental skills. It is the sing of producing new questions, new probabilities, creative imagination and the actual development in science'.*

When the literature was analyzed, it was seen that there are various problem posing methods (Grundmeier, 2003; Abu-Elwan, 2007). These methods are; *i) free problem posing, ii) semi-structured problem*

* This study is a part of doctoral dissertation entitled as *'The Effect of Problem Posing Studies on Mathematical Understanding and Problem Solving Achievements in in Cooperative Learning Environments'* and which is supported by Marmara University Scientific Research Projects Department (Project No: EGT-C-DRP100413-0142).
[1]Corresponding author's address: Kocaeli University, Education Faculty, Kocaeli, Turkey
Telephone: 0 262 303 2446
Fax: 0 262 303 2403
e-mail: yasemin.katranci@kocaeli.edu.tr
DOI: http://dx.doi.org/10.15345/iojes.2016.01.002

posing, iii) structured problem posing (Stoyanova, 2003) and *iv) What if … What if not?* (Abu-Elwan, 2007). In *free problem posing,* students are presented with a situation or a subject from daily life. Students are expected to pose problems by using them. In *semi-structured problem posing,* an open–ended situation is given to students. Students are asked to generate problems about this situation by using their own skills, knowledge and mathematical experiences. In *structured problem posing*, the matter consists of posing a new problem by changing the known information (Akay, 2006). The 'What if? … What if not?' method developed by Brown and Walter (1993) takes place within the scope of structured problem posing (Brown & Walter, 1993).

In problem posing, it is also important to evaluate problems in addition to posing them in order to achieve quality in education. The evaluation of posed problems can be made for two purposes. The first one is to uncover whether or not the desired learning is acquired with problem posing activities and the second one is to evaluate the quality of problems posed by students (Silver & Cai, 2005). At this point, the second purpose was selected in this study with an aim to develop an evaluation instrument for assessing the problems posed by students. In this sense, researchers stated that using rubrics for evaluation is more reliable (Jonsson & Svingby, 2007; Silvestri & Oescher, 2006). The evaluations carried out without using a rubric are inclined to be more subjective and are based on only one rater's personal evaluation. For this reason, teachers state the necessity of using rubrics instead of other evaluation methods (Spandel, 2006). In this regard, using rubrics among alternative evaluation techniques is thought to be more convenient in evaluating posed mathematical problems.

In this context rubrics (graded scoring keys, graded scoring instruction, performance evaluation scale) are learning and scoring instruments which can be used before, during and after classes, which enable teachers to set their objectives and purposes overtly (Goodrich-Andrade, 2000) and which show us the criteria for evaluating the students' performances and to which level students' performances are corresponding to (Kutlu, Doğan & Karakaya, 2009). They are the scoring keys which are developed by teachers or those who are interested in evaluation for guiding students during their learning processes, for helping them to understand how learning products will be analyzed and for supporting their learning (Mertler, 2001; Moskal, 2000; Shepard, 2005; Truemper, 2004). They explain the performance levels and parts of a specific task, product or activity (Hafner & Hafner, 2003) and they are generally used for having quality evaluation (Popham, 2003). All the rubric types can be designed in the shape of a grid or a table as in this study. In a table, the dimensions are given in a descending order and performance levels are listed from left to right (Callison, 2000). In this sense, rubrics are providing criteria regarding students' performances according to various sufficiency levels and exhibit the quality of each criteria/dimension from good work to poor overtly (Rincker, 2002).

The two types of rubrics are defined as holistic and analytical. According to Arter and McTighe (2000) ***holistic rubrics*** instead of separating the product or process into parts by providing a single result for the performance or the product and evaluating each skill focus separately on the whole of a process of product. Holistic rubrics are used when learning products are to be evaluated as in total points (Karip, 2011) and they provide feedback though limited to the students when necessary (Dornisch & Mc Loughlin, 2006; Mertler, 2001; Rincker, 2002; Truemper, 2004). Using a holistic rubric enables us to make quick evaluations. Evaluating products quickly provides an advantage to holistic rubrics (Zimmaro, 2004). In ***analytical rubrics,*** the product or performances are divided into important sections and each section is separately evaluated. Thus, a separate score is obtained for each section (Arter & McTighe, 2000). The total score belonging to analytical rubric is obtained by adding the scores collected from dimensions (Nitko, 2004). The biggest advantage of using an analytical rubric is the fact that it increases the reliability in evaluations. Many researches showed that using analytical rubrics provides a great concordance for the reliability between raters (Al-Fallay, 2000; East & Young, 2007; Knoch, 2009; Nakamura, 2004). Besides, providing more feedback to students and the increase in reliability between raters provide analytical rubrics more advantages (Zimmaro, 2004).

When literature is analyzed, two different ways are followed in evaluating the problems created in mathematics. While the students' problem posing skills, concepts knowledge and sufficiency is evaluated in the first one, the posed problems and problem posing processes are evaluated in the second (Silver & Cai, 2005). Albayrak, İpek and Işık (2006) developed a testing instrument in the studies they realized in this subject which they called as '*an instrument for testing problem posing skills*' and which was composed of 5

questions aimed at determining the skills of prospective teachers regarding problem posing. The students were asked to pose problems according to the problem situations stated in this testing instrument. The posed problems were evaluated by considering the following criteria, *i) writing nothing, ii) using all the given materials, iii) adding new materials* or *iv) the operations of creating samples from the data*.

Ergün (2010) developed an evaluation tool which is called as 'The Rubric for Evaluating Problem Design' and which can be used in evaluating physics problems that are created in his doctoral study. The developed rubric is composed of six dimensions and the following criteria are considered in the evaluation of the posed problems; *i) whether or not the problem can be understood by others, ii) the concordance of the problem with physics principles, iii) the structure of the problem, iv) the number of questions, v) the type of problem* and *vi) the solvability of the problem*.

Grundmeier (2003) evaluated posed mathematical problems by using two different strategies in his study. In the first strategy, it was asked to pose a problem after the problem had been reformulated. In the evaluation, the relation between the original problem and the posed one was analyzed. In the second strategy, new problems were asked to be posed by using given problem situations. In the evaluation of the posed problems, the criteria determined as in the following was used; *i) plausibility and persuasiveness, ii) whether or not the problem involves the necessary information* and *iii) the number of operations required for the solution of the problem.*

In the study realized by Gülten, Ergin and Ergin (2007) problem posing skills were tested with an examination consisting of five questions. In this examination, students were asked to pose problems about given situations and to solve these posed problems. The evaluation was carried out by giving 'one point' when students could not pose any problems about the given situation, 'two points' when they posed the problem incorrectly about given situations and 'three points' when they posed the problem correctly about given situations.

Silver and Cai (1996) gave arithmetic problem situations by using stories to students and they were asked from students to pose mathematical problems by using these situations stated in the story. In the evaluation, the criteria specified as follows were considered; *i) the solvability of the problem, ii) the language of the problem, iii) mathematical complexity* and *iv) the relation between the posed problems.* Silver and Cai (2005) used in another study the following criteria to evaluate problems; *"the quality of the problem"*, *"the originality of the problem"* and *"the number of the posed problems"*.

When we consider the studies that are formerly carried out, it can be seen that some specific criteria were determined for the evaluation of the problems and the evaluation was realized according to these criteria. It was determined that analytical evaluation was not used in the previous studies. Among domestic studies, the in-depth evaluation of the posed problems with different dimensions was conducted only in the study which realized by Ergün (2010). At this point, this evaluation tool developed by Ergün (2010) can be used for evaluating problems posed by students who are studying at higher levels. At this point, the lack of an evaluation tool that can be used to evaluate the students who are studying at lower levels has drawn attention. In this sense, the decision was to focus on this gap in the related literature. Besides, according to Tierney and Simon (2004) there are some problems regarding both in the validity and in the structural use of the rubrics which can be accessed easily through internet. For this reason valid and reliable rubrics are needed. Also it is necessary to focus on the sufficiency and quality of evaluation and reliability and validity studies should be carried out sufficiently (Goodrich-Andrade, 2005). That is why reliability-validity is mentioned as an important point in developing rubrics (Moskal & Leydens, 2000; Thaler, Kazemi & Huscher, 2009).

In this sense, the presence of a gap was noted in the field at this point and the decision to focus on it was taken. By moving from the idea to contribute to the field and literature with this study, the purpose here is to develop a rubric that can be used for evaluating the problems posed by middle school students.

Method

Research Design

In this study, phenomenology design was used among qualitative research techniques in order to determine the performance levels, evaluation criteria and dimensions of the developed rubric.

Phenomenology focuses on the phenomenon that we are aware of but about which we do not have in-depth and detailed understanding. The purpose of phenomenology is to search for the phenomenon that we cannot fully understand but also which are not completely stranger to us (Yıldırım & Şimşek, 2008). At this point, although we are aware of the fact that a rubric can be used as an evaluation tool, using a rubric for the evaluation of the posed problems can be seen as a point that we have not completely understood yet. In this sense, the phenomenon of evaluating the posed problems through rubrics became the model of the study.

Study Group

In order to determine the study group within the scope of this study, the opinions of the three mathematics teachers of the school where the implementation is planned to be carried out were sought. As a result of the interviews, the teachers stated that since the fifth grade students were still in the transition phase from primary school to middle school they were not accustomed, as a consequence, to have a different mathematics teacher other than their class teachers and pointed out that it would not be appropriate to carry out such study at this level with fifth grade students. They also expressed that sixth grade students might have difficulties in adapting themselves to study in cooperative learning environments and for this reason they had fears that the study could last longer than expected. They said that eight grade students would be reluctant to participate because of the examination that they would take at the end of the school year, for this reason the study could be inefficient. Furthermore they also stated that in their opinion the parents of the eight grade students would not be happy to carry out such a study with their children this year. As a result of these interviews, it was concluded that it would be suitable to carry out this study with seventh grade students. Considering that a rubric to be developed for the problems at the specified grade level can also be used for the problems at the other grade levels, the researchers agreed that the study group can be composed of seventh grade students.

The opinions of teachers were taken once more to determine which class would be selected among the seventh grade classes in the school where the study was carried out. While one of the teachers stated that she was new in school, she was still trying to know the students and if she had allowed such study in her class, the students would face difficulties, another teacher stated that she did not want to carry out such a study in her class without specifying any reason. As a consequence, the study was carried out in a public middle school in Kocaeli in the 2012-2013 school year with 29 students from the 7th grade whose teachers did not see any harm to carry out this study. 15 of these students are female and 14 of them are male students.

Data Collection Tools

The data collection tools are worksheets, discussions during lessons and literature about problem posing and rubrics. The data were used in determining the dimensions of the rubric and the criteria for these dimensions. Besides, the data were also used to carry out validity and reliability studies.

Data Analysis

The document analysis includes the analysis of written materials on the issue or issues which are aimed to be searched. The researcher can obtain the data that s/he needs without making any interviews (Yıldırım & Şimşek, 2008). In this sense, document analysis among qualitative research methods was used since the worksheets obtained by students constituted the documents of the study. The document analysis was carried out as in the following; *1) accessing the documents, 2) checking the originality, 3) understanding the documents, 4) data analysis* and *5) using the data* (Forster, 1995, cited in Yıldırım & Şimşek, 2008). The documents should be taken in written if possible from the proper and authorized source (Yıldırım & Şimşek, 2008).

In this study, firstly the worksheets were collected. Secondly, they were checked. They were the original worksheets that were created by students. Thirdly, these worksheets were deeply understood by the researchers. After that, these data were analyzed descriptively. The purpose of such analysis is to present the data in an organized and interpreted way to readers (Yıldırım & Şimşek, 2008). In this study, the data were descriptively analyzed and interpreted and in conclusion the rubric was developed. Besides, the data collected from the discussions during the study were also evaluated descriptively. SPSS 15.0 software was used in calculating the reliability coefficients of the rubric.

The Stages of Developing the Rubric

In this part, the findings and their interpretations which were obtained during the process steps that were followed for the development of the rubric respectively. In order to be able to develop a rubric which is

suitable to the purpose of the study, rubric development stages that were proposed by Goodrich-Andrade (2000) were followed. These stages are as follows; *i) the assigning problem posing tasks to students both in the class and as their homework (the collection of students' worksheets through problem posing studies), ii) during the problem posing studies carried out in class, having discussion like brainstorming, iii) the analysis of the worksheets (listing performance criteria and determining the performance levels), iv) determining the dimensions and evaluation criteria of the rubric, v) creating and implementing the draft rubric* and *vi) doing reliability and validity studies regarding the final version of the rubric*. This process and the findings obtained were interpreted and then presented below.

Data collection process (i and ii). In order to develop the rubric which will be used to evaluate the problems posed by the students. There are 4 hours mathematics lessons per week in middle schools and each lesson lasts 40 minutes. In this sense, the study was carried out within the scope of 25-hour mathematics course. The activities carried out before implementation and during implementation are shown in Table 1.

Table 1. The process of obtaining students' worksheets with problem posing activities

Before Implementation (Preparation)	1. The determination of the mathematical subjects that will be used 2. The preparation of cooperative learning based lesson plans for the specified subjects 3. The analysis of the daily lesson plans and learning environment with the teacher of the selected class 4. Making learning environment suitable to cooperative learning by considering the number of students in the class
Implementation	1. Informing students about the implementation in general (the information about classroom lay-out and groups) 2. Dividing students into groups and classroom lay-out 3. Discussion and explanations about the group names and how the groups will study 4. Teaching classes

The mathematical subjects that will be taught were specified before the implementation. The specified subjects are; "Algebraic Expressions", "Equations", "Circle and Closed Disc", "Angle Measurements" and "Ratio and Proportion". After specifying the subjects, daily lesson plans which are based on cooperative learning were prepared. Cooperative learning environments were preferred to increase the communication between students and to ensure that more fruitful problems are conceived by creating an environment which is based on competition. Whether or not the cooperative learning environment has any effect was not explored during the study. It was decided to have five cooperative based learning groups since there were 29 students in the class. The groups were; *Cahit Arf Group, Mr Salih Zeki Group, Hüseyin Tevfik Pasha Group, Ali Kuşçu Group* and *Tosun Terzioğlu Group*. Later on students were distributed to the groups. After the groups were formed, each group was asked to determine their leader, assistant leader, secretary, spokesman and mailman. The tasks of the group members and how the study will continue were explained in detail to the students.

In order to provide an insight to the reader, four hours of the teaching performed on the subject of 'Algebraic Expressions' was presented as an example below. Similar methods were used in teaching the other subjects within the scope of curriculum. While only a single problem task was covered during the lessons, all three problem posing tasks were used in assignments. In this way, the worksheets that were used in this study which lasted 6 weeks (25 hours). The data were obtained through studies that were carried out as explained above.

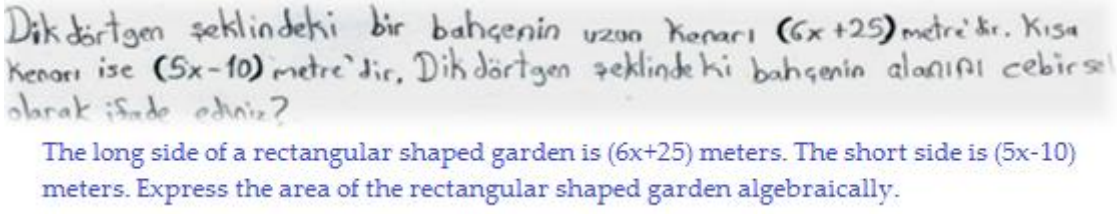
At the first lesson, students were asked what they knew about the 'algebraic expression' concept by the researcher and whether or not they had previously come across this concept. It was perceived from students' expression as '*mathematics with letters instead of numbers*' that students partially knew the concept of algebraic expression. After that, addition and subtraction in algebraic expressions were presented to the students through question-answer method. First of all; term, concepts such as term, coefficient, common factor, commutative property, associative property and neutral element etc. were given to students and previous knowledge of the students regarding these concepts were examined. It was determined that they defined term and coefficient expressions as '*two mathematical expressions that are friends of each other*'. Then, the examples regarding the addition and subtraction in algebraic expressions were written on the blackboard

and students were asked to solve these examples. Furthermore, the students were asked to suppose themselves as teachers and to think how they could pose problems.

At the second lesson, a problem was written on the board by the researcher. It was ensured that all the students solved this written problem. Later on, a student showed the solution at the board. After the student had shown the solution, the groups were asked to think both on the problem and the solution. For this, 2-3 minutes of thinking time was given to groups. After that, each group's spokesman stated their opinions on both the problem and its solution. When the groups had expressed their opinions, the researcher made some explanations about the problem and its solution. For example, they were told that it would be better to make calculations instead saying we solve the problem with our eyes to be sure that the solution is correct. The students were reminded that in posing problems, it was necessary to pay attention to use mathematical terms, expressions and concepts.

Subsequently, students were asked whether they could pose a problem similar to that problem or not. It was observed that students failed to understand what to do since it was the first time they were posing problems. In this sense, the students were asked if a new problem could be obtained by changing the numbers in the problem written on the board. It was seen by the expressions of the students that they could easily pose problems in this way. Furthermore the students were asked whether or not new questions could be added to the given problem. By the expressions of students it was determined that they could easily add or extract some other expressions since the subject was algebraic expressions. Then, the students were given 4-5 minutes to pose their own problems. They were told that each student should also solve the problem that s/he or posed. At the end of this time, each group was told to select the problem that they will send to the other group and write it on a paper. All the posed problems were collected by the researcher (except the ones that would be sent to the other groups). The groups were told to give the problem that they will send the others. Later, groups started mailing their problems. After each group had got their mails, the problems were solved as a group and students were made to think about the problem. After the solution, the spokesmen of each group came respectively and shared their opinions about the solution. It was observed that students were inclined to write the answer directly regarding the problem solutions and an effort was made to overcome this situation during the study. Since all these activities were about students' posing problems according to the task of posing structured problems, effort was made to ensure that students also posed problems according to the other tasks. When the grade levels of the students were considered, they were requested to pose problems only according to a single problem posing strategy during the lessons. However students were expected to pose problems that are suitable to each of the three strategies in their assignments. In this regard, students were informed about what to do and the points to pay attention to in those assignments were explained.

At the third lesson, the lesson started with mailing the group problem among the problem posing tasks assigned as homework to the other group with the mailman. After mailing the problems, the solutions were realized and class discussions were made. After discussion, the subject of Multiplying Two Algebraic Expression was explained to students. At this lesson the task of posing semi-structured problems was emphasized. The expressions "garden, (5x-10) meters, (6x+25) meters, (2x+40) meters" were written on the board by the researcher and the students were made to pose a problem and to solve it. After posing and solving problems, groups mailed their problems. The same things that had been carried out in the previous lesson were repeated. One of the problems that were posed by students at the lesson was given as an example in figure 1.


Dikdörtgen şeklindeki bir bahçenin uzun kenarı (6x+25) metre'dir. Kısa kenarı ise (5x-10) metre'dir. Dikdörtgen şeklindeki bahçenin alanını cebirsel olarak ifade ediniz?

The long side of a rectangular shaped garden is (6x+25) meters. The short side is (5x-10) meters. Express the area of the rectangular shaped garden algebraically.

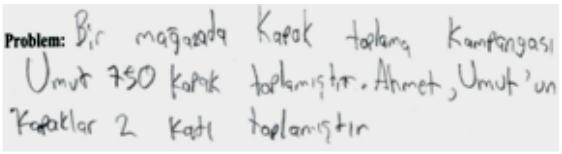
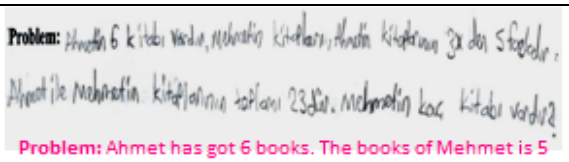
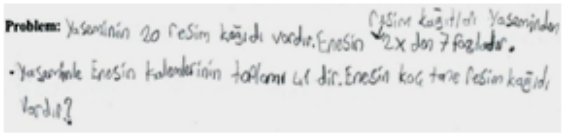
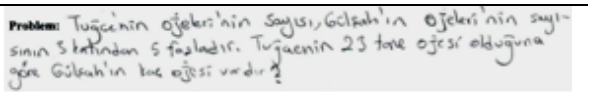
Figure 1. An example from the task of posing semi-structured problems posed in the lesson

At the fourth lesson, the task of posing free problems was emphasized. Furthermore, the students were observed to become more accustomed to the study; they became more aware of their tasks in the group and

they created a competitive atmosphere with the other groups. At the end of the lesson the task of posing problems was assigned as homework.

The analysis of the worksheets. The worksheets obtained through the tasks assigned during the lessons and as homework were analyzed in detail by the researchers. At the end of this analysis, some criteria were specified by the researchers. After reaching a consensus over criteria, the problems posed were divided into three categories as; weak, average and good. The specified criteria and sample problems can be seen in Table 2 below.

Table 2. The criteria for dividing the posed problems into categories and sample problems

Category	Criteria	Sample Problems (Sample Problems that Fit the Criteria)
Weak	• Lack of understandable problem text. • Incomplete problem. • Lack of question expression. • Unclear expectations. • Lack of connection between the real life and the ones given at the problem and in conclusion the problem is too complicated. • Being an exercise type problem which is very easy to solve.	 **Problem:** Umut collected 750 bottle caps in a campaign of a store for collecting bottle caps. Ahmet collected Umut's caps twice as many.
Average	• Although the problem text is understandable, lack of using necessary mathematical expressions (for example; including the expression such as "5 more than 3x"). • Having ambiguous expressions in problem texts. • It is not possible to solve the problem with the information provided in the problem. • In problems which require figures, although the problem text is clear and understandable, there is an inconsistency between the figure and the text. • Being an exercise type problem.	 **Problem:** Ahmet has got 6 books. The books of Mehmet is 5 more than 3x of his books. The total number of Ahmet and Mehmet are 23. How many books has Mehmet got?  **Problem:** Yasemin has got 20 drawing papers. The drawing papers of Enes is 7 more than 2x of the drawing papers. The total number of their pencils is 41. How many drawing papers has Enes got?
Good	• Having a clear and understandable problem text. • Being suitable to mathematical principles. • Stating clearly what the problem is asking. • The problem is solvable.	 **Problem:** The number of nail polishes of Tuğçe is 5 more than the triple of the nail polishes of Gülşah. If Tuğçe has 23 nail polishes, what is the number of nail polishes of Gülşah?

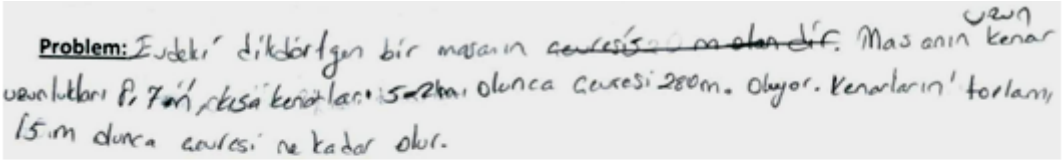
The findings stated in Table 3 were obtained as a result of the detailed analysis of 507 problems collected according to criteria stated in Table 2.

Table 3. The distribution of the problems according to categories

Problem Posing Tasks	Weak	Average	Good	Total
Free Problem Posing	51	35	62	148
Semi-structured Problem Posing	57	60	78	195
Structured Problem Posing	35	38	91	164
Total	143	133	231	507

When Table 3 is analyzed, it is seen that in total 507 problems were posed. As a result of this last analysis, some weaknesses were detected.

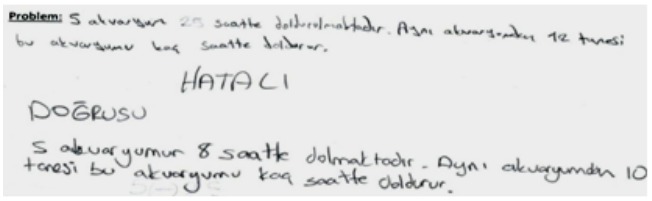
The weaknesses of posed problems. First of all, it was determined that the texts of the problems were not clearly and understandably expressed. It was seen that students could understand their own posed problems but when these problems were read by others, the texts of the problems were not understood. For example;



Problem: The circumference of the rectangular table at home is 20 meters. The length of the long sides of the table is 8, 7 meters; the short sides are 5-2 meters then the circumference 280 meters become. When the total length of the sides is 15 meters, what is the circumference?

Figure 2. The problem posing task of the SB coded student

In the problem stated in Figure 2, the text of the problem is highly complicated. As a consequence, what the student wanted to ask is not clear. In the cases when the texts of the problems were partially understandable, it was determined that the corrections were made when the problem was read by others. However, it was also found that the texts of the problems were not clearly understood in these corrections. For example;



Problem: 5 aquariums are swapped in 25 hours. How many hours does it take for the same 12 aquariums to swap this aquarium?

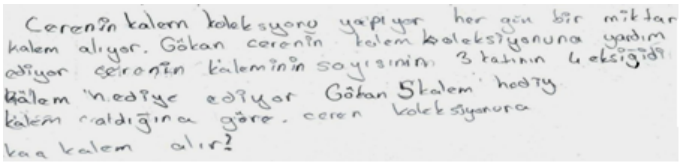
MISTAKEN

THE CORRECT ONE

5 Aquariums are swapped in 8 hours. How many hours does it take for the same 10 aquariums to swap this aquarium?

Figure 3. The problem posing task of the GH coded student

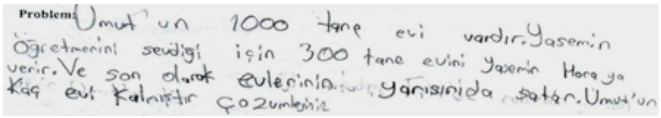
The GH coded student in figure 3 found the problem posed by his/her friend as erroneous. Besides, s/he tried to correct the erroneous problem but s/he also was not able to write the problem text clearly and explicitly. Furthermore, the texts of the problems were observed to be too long since detailed explanations were used in these texts and this length made the problems difficult to understand. For instance;



Ceren collects pencils and buys some pencils every day. Gökan helps Ceren to collect pencils. He presents her 4 less pencil than triple of her total pencils. Since Gökan bought 5 pencils as a present, how many pencils are there in Ceren' s collection?

Figure 4. The problem posing task of NY coded student

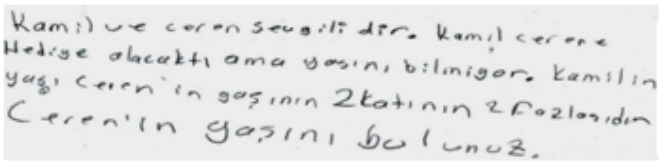
In figure 4, there are some expressions which detailed the problem such as Ceren collects pencils and Gökhan helps her. It is appeared that such details make it difficult for the student to solve the problem. Another weakness is the fact that although the problem can be solved by using the numerical data provided in the problem, these data are not suitable with the everyday life. For example;



Problem: Umut has got 1000 houses. As he loves his Yasemin teacher, he gave 300 of them to Yasemin teacher. And finally, he sold half of his houses. How many houses left for Umut? Please find.

Figure 5. The problem posing task of the UK coded student

As it can be seen in figure 5, the numbers provided in the above problem are not suitable to everyday life. Another weakness that attracts attention is to use emotional expressions (such as love, like, want to hang around) in the problems in addition to mathematical expressions. For example;



Kamil and Ceren are in love with each other. Kamil wanted to buy a gift to Ceren but he does not know her age. The age of Kamil is 2 more than the twice of the Ceren's age. Please find how old is Ceren.

Figure 6. The problem posing task of the BK coded student

As it can be seen in figure 5 and figure 6, it was determined that students used expressions such as to love or to be in love with. Using such expressions is certainly not erroneous mathematically. But, it was noticed that problems including such expressions distracted the students' attention and there were discussions between students during class debates. In this sense, the problems including such expressions were accepted as emotional rather than scientific.

Another weakness is the fact that some students used mathematical expressions incorrectly while posing problems. For example, it appeared that problems such as *'The circumference of a rectangular is 310 and if the sides of this rectangular is proportional with the numbers of 3 and 4, then what is the long side of this rectangular?'* were posed. In this sample problem, it was found that the student was not paying attention at all to length expressions (for instance; 'writing only 310 instead of 310 m²'). Another point is that students found it sufficient to put a question mark to the point where the answer was required on the figure without using problem text while posing problems requiring the use of figures. Additionally, in problems involving figures, any correlation was not established between the figure and the text of the problem. Besides, inconsistencies were detected between the given figure and the figure asked in the text of the problem. For instance;

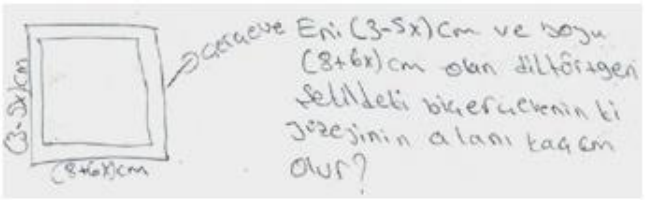


Figure 7. The problem posing task of the SO code ctudent

As it can be seen in the problem stated in figure 7, a figure was drawn by the student but it was not clear which of the surfaces of the figure were referred to in the text of the problem. Although it can be deduced that the purpose is to ask what the area remaining between two inter-wined rectangles is, it is not clearly stated in the text of the problem. Besides, it was observed that as a result of the students' confusion in calculating the area of the some of the mathematical figures, they posed erroneous problems. For instance;

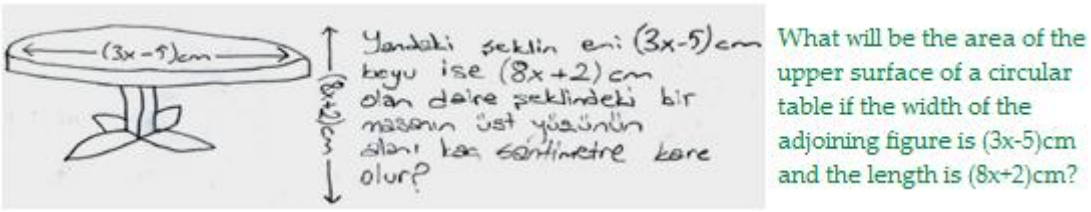


Figure 8. The problem posing task of the MSK coded student

In the problem given in figure 8, the student drew a circular table and asked to calculate the area of this surface. However, it is seen that s/he did not notice the fact that it is not possible to get the desired result with the data s/he provided in the figure. Besides, it appeared that with expressions such as 'length and width' the student lacked the knowledge to calculate the area of a circle and had misconceptions.

In addition to such noted weaknesses, some problems were found to be too simple and in the form of an exercise. When the age and grade levels of the students were considered, it was seen that they paid much attention to numerical data given in problems. It clearly appeared that they preferred numbers that could be divided without remainder while making divisions and urged to do so their friends who did not prefer such numbers. It was found that some students were not able to solve their own problems and noted as '*I guess it is incorrect, I could not solve it'*. Furthermore, multiple choice answers were added to some problems.

The features of the problems stated under 'good' category. In problems grouped under the category "Good" the texts of the problems were found to be stated clearly and understandably, the problems involving a figure were compatible with the given figure, the given numerical data were suitable to everyday life and the problems were solvable. For instance;

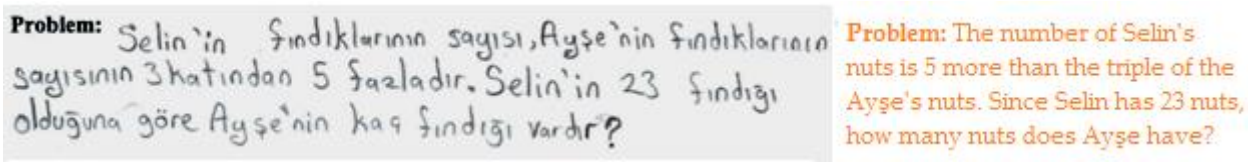


Figure 9. The problem posing task of the EK coded student

As it can be seen in Figure 9, the text of the posed problem is clear and explicit. Besides, the numbers and expressions given in the problem were also suitable to everyday life.

The findings obtained as a result of class discussions. After problem posing tasks, discussions in the form of brainstorming were carried out with students. In these discussions, the students were asked to evaluate the problems posed by their peers and their opinions were taken regarding the mistakes during problem posing and things to do for posing better problems. Based on these evaluations, an effort was made to determine the points to be used in a problem which can be considered as 'good'. In this process, the researcher tried to increase the awareness of students about problem posing and to keep the competitive environment at a reliable level in discussions by taking the guidance role. The opinions that came out of the class discussions carried out in the form of brainstorming were set out as follows;

1. The numerical data should be carefully considered while posing problems. Large numbers should not be preferred and they should be related with everyday life,
2. The given number should be fully divisible,
3. It is necessary to write problems more clearly,
4. Complicated problems which are difficult to understand should not be posed,
5. The one who posed the problem should also solve it. Unsolvable problems should not be asked,
6. The posed problems should not be so easy,
7. In some problems, the choices provided for the answers should be carefully considered,
8. The results obtained with the solution of the problems should be consistent with the real life situations,
9. The solvability of the problem should be tested during problem posing process. For this reason, the one who posed the problem should also solve the problem and
10. It is necessary to have a sound knowledge regarding the mathematical concepts about the given subject in order to be able to pose the problem.

In accordance with the findings obtained and the findings resulting from class discussions, the features of good and weak problems were determined (Is the posed problem in the form of an exercise?, Is the problem sentence understandable? Is the problem suitable to mathematical principles? etc.). Under the light of these determined features, it was concluded that problem posing activities involved multiple dimensions and for this reason the decision to develop an analytical rubric was taken.

The determination of the evaluation criteria and the dimensions of the rubric. Decision was taken to develop an analytical rubric in the light of the findings obtained through class discussions and worksheets. The features of the good problems (for example; using mathematical expressions properly, the solvability of the problems) were taken as a basis in determining the dimensions of the rubric. It was decided that the rubric should include six dimensions by considering the posed problems and grade levels of the students. After deciding on the dimensions of the rubric, evaluation criteria/performance criteria were determined. Later on, the scoring of this prepared rubric was made. Thus, the dimensions and evaluation criteria of the draft rubric took their final form. The rubric was finalized by considering the opinions of the grammarian. This draft rubric was used by the researchers in the evaluation of the students' problems.

The implementation of draft evaluation rubric. The draft rubric composed of six dimensions was developed as a result of the findings obtained through class discussions and worksheets. This draft rubric was used in evaluating the students' problems. As a result of these evaluations it was determined that some dimensions and criteria of the rubric were not functioning and did not have any discriminative feature in the evaluation of the performance. The findings obtained as a result of implementing the draft rubric are stated below.

The dimensions which were used in the draft version but were not used in the final rubric. It was seen as result of the implementation that two dimensions which were used in draft version of the rubric did not have any function. These dimensions are; *The Degree of the Problem* and *The Originality of the Problem*. "The Degree of the Problem" dimension is as in the following.

The Degree of the Problem *(What is the level of difficulty of the problem?)*
1: It is an exercise type problem which can be solved very easily.
2: It is an exercise type problem which can be solved easily.
3: It is a normal problem which can be solved easily.
4: It is a normal problem which is difficult to solve.

The purpose of this dimension was to determine difficulty levels of the problems posed by students. But, in the implementation process of the draft rubric, it was observed that students who solved the problems did not have any difficulties as students posed problems which could be solved in maximum 2 or 3 steps. In this sense, since it was determined that this dimension which was used in the draft rubric did not have any discriminative feature in the evaluation of performance, it was determined that this dimension shall not be used in the final version of the rubric. "The Originality of the Problem" dimension is as in the following.

The Originality of the Problem *(Is this an original problem?)*
1: It is not an original problem.
2: It is partially an original problem.
3: It is an original problem.

The purpose of this dimension was to determine the originality of the posed problems. When students were asked to pose original problems in problem posing process, it was observed that they posed problems which did not have any connection with the real life. For this reason, since it was determined that this dimension used in the draft rubric did not have any discriminative feature in the evaluation of performance, it was determined that this dimension shall not be used in the final version of the rubric.

The dimensions which were used in the draft version and were edited in the final version of the rubric. "Type/Structure of the Problem" dimension was edited after implementation and in accordance with the opinions of an expert. This dimension was used in the draft rubric as in the following.

The Type/Structure of the Problem *(What is the type/structure of the posed problem?)*
1: It is in simple exercise form.
2: It is an exercise.
3: It is a normal problem.
4: It is a non-routine problem.

It was seen while implementing the draft rubric that the posed problems were in the form of an exercise or normal/verbal problem. In this sense, 'the Type / Structure of the Problem' dimension were take the following below in the final version of the rubric including the dimension which was used in the draft version but was not used in the final version of the rubric for determining the difficulties of the problems.

The Type/Structure of the Problem *(What is the type/structure of the posed problem?)*
1: It is in simple exercise form.
2: It is an exercise.
3: It is a simple normal/verbal problem.
4: It is a normal/verbal problem.

The other edited dimension is the dimension of "The Solvability of the Problem". This dimension was stated in the draft rubric as in the following.

The Solvability of the Problem *(Is there a solution for the posed problem?)*
0: The data and information provided in the problem is not sufficient for any solution.
1: The problem is too complicated to be solved.
2: The problem has a solution but the data is mistaken.
3: The problem has a solution but some data is missing.

4: The problem has a solution since all the data and information of the problem is complete and proper.

As a result of the implementation of the draft rubric, it was seen that the cases of erroneous and missing data in the posed problems were overlapping. For this reason, it was decided that it was not necessary to evaluate these two criteria separately. In this case, this dimension was stated in the final version of the rubric as in the following.

The Solvability of the Problem *(Is there a solution for the posed problem?)*
0: The data and information provided in the problem is not sufficient for any solution.
1: The problem is too complicated to be solved although the data is sufficient.
2: The problem has a solution but the data is mistaken and missing.
3: The problem has a solution since all the data and information of the problem is complete and proper.

This draft rubric was also implemented by a different expert. In accordance with the findings, the things that should be done in the draft rubric, omitting dysfunctional or editing dimensions and criteria were all discussed. The rubric was also edited in accordance with the opinions of a grammarian. For example; the sentence 'the text of the problem not clear and understandable' was changed with 'the text of the problem is not clear and understandable'. Similarly, the other criteria were also edited.

Besides, in order to use the rubric more efficiently, the shape of the rubric was also discussed and edited and rubric was put in a different shape from the one used in the draft. The field experts' opinions were taken again for all editions and changes made in rubric. The final version of the rubric was given in Appendix 1. The rubric was used once again after it had taken its final shape.

The reliability and validity studies of the final version of the rubric. *The validity of the rubrics* depends on the objectives of the assessment. First of all, the teacher needs to express clearly that what performance s/he is expecting from the students at which level. There are three types of evidences which prove the validity of the evaluation tool. These are; content validity, construct validity and criterion validity (Moskal & Leydens, 2000). There are some questions that are needed to be answered in order to control the validity of a rubric.

The following questions should be answered for *content validity*; Do the evaluation criteria define any irrelevant content?, Do the evaluation criteria define all the dimensions of the designed content?, Are there any undefined content domain in the activity/assignment which will be evaluated by using rubric?, for *construct validity*; Whether all the important areas of the designed structure which is evaluated with scoring criteria are available or not?, Are there any evaluation criteria which is not related with the structure? and *for criterion validity*; How is scoring criteria reflecting the elements of the related performance or the future success?, What are the important elements of the related performance or the future which is evaluated by using a rubric? How is scoring criteria assessing the important elements of the related performance or the future? and Are there any areas of the related performance or the future which is not reflected in the scoring criteria? (Tuncel, 2011).

In this study, the rubric will be used solely with the purpose to evaluate the posed problems and it does not have any irrelevant content. It was ensured that evaluation criteria defined all the aspects of the designed content. Similarly, it was tried not to leave any part of the task or performance that will be evaluated out of evaluation with the help of the editions done in the draft rubric. For the evaluation of the problems, an attempt was made to ensure for the rubric to cover all the major surfaces of the performance by forming it with four sub-dimensions. The answer the second question about the construct validity was sought without using the irrelevant evaluation criteria of the draft version in the final version of the rubric. The questions on the criterion validity were tried to be answered by including sub-dimensions both in the draft and final version of the rubric, specifying the evaluation criteria of these sub-dimensions between 0 and 4 points and determining the coefficients for the evaluation criteria of each sub-dimensions. In a study carried out by Jonsson and Svingby (2007) 75 rubric development studies were reviewed. The validity studies were analyzed in only 25 of these studies. In this review, it was observed that only one type of validity had been examined regarding the validity of the rubrics. It appeared that providing more than one signified validity types had been considered in only a few studies. In this sense, the validity studies were accepted as sufficient since the aim was to provide three different types of validity.

The reliability in rubrics means having the same evaluation score after the multiple evaluations of the performance. Two different reliabilities are accepted as important for the evaluation of the performances. They are; *intra-rater reliability* and *inter-rater reliability* (Moskal, 2000).

It is the consistency between the scores obtained through the evaluation of the same performance by the same raters in different time intervals (Moskal & Leydens, 2000). In most of the studies, Cronbach Alfa value is calculated for intra-rater reliability (Jonsson & Svingby, 2007) and the value which is equal to or above 0,70 is accepted as sufficient (Brown, Glasswell & Harland, 2004). When an evaluation is carried out by using a rubric, it is stated that the calculated intra-rater reliability is higher than inter-rater reliability (Jonsson & Svingby, 2007). It is the correlation between the scores obtained through the evaluation of the same performance task by two different raters (Hafner & Hafner, 2003; Newell, Dahm & Newell, 2002). There are three general categories in inter-rater reliability. These categories are; consensus estimates, consistency estimates and measurement estimates (Stemler, 2004).

In *consensus estimates*, independent raters should reach a consensus about how they will use rubrics in evaluating students' performances. Consensus estimates are helpful with likert type rubrics in the reliability of the evaluations (Stemler, 2004). It is stated that consensus estimates are the most suitable models for calculating inter-rater reliability (Oakleaf, 2009). There are three different methods for calculating consensus estimates. The first and the most popular of them is the percentage agreement between the raters. In this method, percentage agreement is calculated by dividing the same scores given for the same performances into the total number of performances (Stemler, 2004). For example; suppose that 50 students' performances were evaluated by two different raters. Let's accept that these two different raters gave the same score to the 40 performance task out of 50. According to this, the percentage agreement is calculated as 40/50=0,80. The second method for calculating consensus estimates is Kendall's coefficient of concordance. Kendall's coefficient of concordance is used for calculating the agreement between multiple raters and ve it is suitable for the ordered responses which are coded with (SAS, 2006). The third method for calculating the consensus estimates is Kappa statistics. Kappa statistics shows whether the agreement between raters is better than the estimated chance or not (Stemler, 2004). Landis and Koch (1977) stated the limits of acceptance for the Kappa value in Table 4.

Table 4. The acceptance values for **κ**

κ	**Strength of Agreement**
< 0,00	Poor
0,00 – 0,20	Slight
0,21 – 0,40	Fair
0,41 – 0,60	Moderate
0,61 – 0,80	Substantial
0,81 – 1,00	Almost Perfect

It appears that there are three methods for calculating *consistency estimates*. These methods are; Pearson correlation coefficient, Spearman correlation coefficient and Cronbach Alfa value. In consistency estimates, it is accepted as sufficient when the calculated value is equal to or above 0,70 (Brown, Glasswell & Harland, 2004; Stemler, 2004). It is seen that generalized theory and Many-facet Rasch model is used for calculating the reliability of the assessment. It is expressed that the consistency value is accepted when it is equal to or above 0,80 (Brown & others, 2004; Marzano, 2002).

In this study, for the reliability of the rubric, 100 problems were selected proportionally among 507 posed problems which had been previously classified by students as weak-average-good. These selected problems were evaluated by two different raters and by the same raters every other three week. For the consistency of the obtained scores, consensus estimates, Kappa values, Pearson correlation coefficients, Spearman correlation coefficients and Cronbach alpha values were calculated. The consistency coefficients of the *"Rubric for Evaluating Problem Posing (REPP)"* were given in Table 5.

When Table 5 is analyzed; all consistency coefficients of the rubric are accepted as sufficient.

Table 5. Consistency coefficients of the REPP

Consistency coefficients	Different Raters	Same Raters
Consensus estimates	,720	,830
Kappa value (κ)	,700	,817
Pearson correlation coefficient	,873	,921
Spearman correlation coefficient	,865	,912
Cronbach alpha	,932	,959

Discussion, Conclusions and Suggestions

A rubric which has four dimensions was developed in order to be able to evaluate the problems posed by students. While designing the draft rubric, the number of criteria used in each dimension was not fixed so as to have a purposeful assessment and 3 or 5 evaluation criteria was specified for each dimension. In the final version of the rubric, there were four criteria regarding each dimension. The first of these dimensions is '*The Text of the Problem* (*Language and Expression*). This dimension was used in the rubric in order to test whether or not the text of the posed problems were expressed clearly and understandably. The second dimension is *"The Compatibility of the Problem with the Mathematical Principles"*. This dimension was used in the rubric in order to test students' ability to use and express mathematical concepts. The third dimension is *"The Type/Structure of the Problem'*. This dimension was used in the rubric in order to determine both the types and difficulty levels of the posed problems. The last and the fourth dimension is *"Solvability of the Problem"*. This dimension was used in the rubric to determine whether or not the problem has a solution and the concordance between the asked and given at the problem.

In the reliability studies of the developed rubric, both inter-rater and intra-rater reliability scores were calculated (Moskal, 2000). The intra-rated consensus percentage was calculated as ,830; Kappa value as ,817; Pearson correlation coefficient as ,921; Spearman correlation coefficient as ,912 and Cronbach Alpha value as, 959. In the studies, Cronbach Alpha Value is calculated for intra-rater reliability (Jonsson & Svingby, 2007) and it is accepted as sufficient when this value is or over 0,70 (Brown, Glasswell & Harland, 2004). For this study, the intra-rater Cronbach Alpha values was calculated as ,959. In this sense, it was concluded that the intra-rater reliability coefficient of this developed rubric was valid.

In this study, for the rubric the consensus percentage was calculated as ,720; Kappa value as ,700; Pearson correlation coefficient as ,873; Spearman correlation coefficient as ,865 and Cronbach Alpha value as, 932. In Ergün's (2010) study, 90 problems posed by students were evaluated by two different raters and intra-raters in between twenty days and Pearson correlation coefficient which shows the consistency was calculated. While inter-rater Pearson correlation coefficient was calculated as r=0,86 the intra-rater Pearson correlation coefficient was calculated as r=0,92. The results of this study also show parallelism with the studies of Ergün. In this sense, the rubric was found to be reliable since the calculated value is and over the value of 0,70. In conclusion, according to the results of the analyses carried out, it can be said that the developed rubric is a valid and reliable scale. Although it was determined that the rubric is valid and reliable, it is necessary not to miss the fact that the reliability and validity studies were carried out on the same problems. In this sense, it is thought and suggested that it is necessary to repeat the validity and reliability studies by using the developed rubric for evaluating different problems. The researchers and teachers who will be using the rubric should not ignore this point and it is advised to repeat validity and reliability studies before using it in their studies. Besides, there will be studies as the continuation of this study about the specified point by the researchers.

It is thought that this rubric developed with the purpose of evaluating problems posed by middle school students in their lessons can also be used for evaluating problems posed by elementary school students. Besides, it can be said that this rubric can also be used for evaluating problems posed within the scope of middle school science lessons. At this point, using this rubric for evaluating problems at the levels and lessons stated above can be investigated. Additionally, it is thought that re-developing this rubric will be useful for evaluating problems that will be posed at high school and university levels by considering the students' grade levels and level of problems posed by them and by editing or adding new dimensions to the developed rubric.

References

Abu-Elwan, R. (2007). The use of web quest to enhance the mathematical problem-posing skills of pre-service teachers. *The International Journal for Technology in Mathematics Education, 14*(1), 31. (ERIC No: EJ847732)

Akay, H. (2006). *Problem kurma yaklaşımı ile yapılan matematik öğretiminin öğrencilerin akademik başarısı, problem çözme becerisi ve yaratıcılığı üzerindeki etkisinin incelenmesi*. Yayınlanmamış doktora tezi, Gazi Üniversitesi, Ankara.

Akay, H., & Boz, N. (2010). The effect of problem posing oriented analyses-II course on the attitudes toward mathematics and mathematics self-efficacy of elementary prospective mathematics teachers. *Australian Journal of Teacher Education, 35*(1), 59-75. doi: 10.14221/ajte.2010v35n1.6

Albayrak, M., İpek, A. S., & Işık, C. (2006). Temel işlem becerilerinin öğretiminde problem kurma-çözme çalışmaları. *Erzincan Eğitim Fakültesi Dergisi, 8*(2), 1-11.

Al-Fallay, I. (2000). Examining the analytic marking method: Developing and using an analytic scoring schema. *Language & Translation, 12*(1), 1-22.

Arter, J., & McTighe, J. (2000). *Scoring rubrics in the classroom: Using performance criteria for assessing and improving student performance.* Thousand Oaks, CA: Corwin Press.

Brown, S. I., & Walter, M. I. (1993). *Problem posing: Reflection and applications*. Hillsdale, NJ: Erlbaum.

Brown, G. T. L., Glasswell, K., & Harland, D. (2004). Accuracy in the scoring of writing: Studies of reliability and validity using a New Zealand writing assessment system. *Assessing Writing, 9*, 105-121, doi: 10.1016/j.asw.2004.07.001.

Callison, D. (2000). Rubrics. *School Library Media Activities Monthly, 17*(2), 34-36.

Chang, K. E., Wu, L. J., Weng, S. E., & Sung, Y. T. (2012). Embedding game-based problem solving phase into problem posing system for mathematics learning. *Computers & Education, 58*(2), 775-786. doi: 10.1016/j.compedu.2011.10.002.

Dornisch, M. M., & Mc Loughlin, A. S. (2006). Limitations of web-based rubric resources: Addressing the challenges. *Practical Assessment, Research & Evaluation, 11*(3), 1-8.

East, M., & Young, D. (2007). Scoring L2 writing samples: Exploring the relative effectiveness of two different diagnostic methods. *New Zealand Studies in Applied Linguistics, 13*(1), 1-21.

Ergün, H. (2010). *Problem tasarımının fizik eğitiminde kavramsal öğrenmeye ve problem çözmeye etkisi*. Yayınlanmamış doktora tezi, Marmara Üniversitesi, İstanbul.

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

Goodrich-Andrade, H. (2005). Teaching with rubrics: The good, the bad, and the ugly. *College Teaching, 53*(1), 27-30. doi: 10.3200/CTCH.53.1.27-31

Grundmeier, T. A. (2003). *The effects of providing mathematical problem posing experiences for K-8 pre-service teachers: investigating teachers' beliefs and characteristics of posed problems*. Unpublished doctoral dissertation, University of New Hampshire, Durham, ABD.

Gülten, D., Ergin, H., & Ergin, T. (2007). *İlköğretim 3. sınıf öğrencilerinin problem kurma becerileri ile bilişsel işlemlerden eşzamanlılık ve planlama arasındaki ilişki.* Anadolu Üniversitesi VI. Ulusal Sınıf Öğretmenliği Eğitimi Sempozyumu, Eskişehir, Türkiye.

Gür, H., & Korkmaz, E. (2003). *İlköğretim 7. sınıf öğrencilerinin problem ortaya atma becerilerinin belirlenmesi*. Matematikçiler Derneği Bilim Köşesi. http://www.matder.org.tr adresinden 06.04.2015 tarihinde edinilmiştir.

Hafner, J. C., & Hafner, P. M. (2003). Quantitative analysis of the rubric as an assessment tool: An empirical study of student peer-group rating. *International Journal of Science Education, 25*(12), 1509-1528. doi: 10.1080/0950069022000038268

Işık, C., Kar, T., Yalçın, T., & Zehir, K. (2011). Prospective teachers' skills in problem posing with regard to different problem posing models. *Procedia Social and Behavioral Sciences, 15*, 485-489. doi: 10.1016/j.sbspro.2011.03.127

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review, 2*(2), 130-144. doi: 10.1016/j.edurev.2007.05.002

Karip, E. (Ed.) (2011). *Ölçme ve değerlendirme*. Ankara: Pegem Akademi Yayıncılık.

Knoch, U. (2009). Diagnostics assessment of writing: A comparison of two rating scales. *Language Testing*, *26*(2), 275-304. doi: 10.1177/0265532208101008

Kutlu, O., Doğan, C. D., & Karakaya, I. (2009). *Öğrenci başarısının belirlenmesi: Performansa ve portfolyaya dayalı durum belirleme*. Ankara: Pegem Akademi Yayıncılık.

Landis, J. R., & Koch, G. G. (1977). The measure of observer agreement for categorical data. *Biometrics, 33*(1), 159-174.

Lavy, I., & Bershadsky, I. (2002). *"What if not?" problem posing and spatial geometry-a case study*. 26th Conference of the International Group for the Psychology of Mathematics Education: PME 26, Norwich, UK.

Marzano, R. J. (2002). A comparison of selected methods of scoring classroom assessments. *Applied Measurement in Education*, 15(3), 249-267. doi: 10.1207/S15324818AME1503_2

Mertler, C. A. (2001). Designing scoring rubrics for your classroom. *Practical Assessment, Research & Evaluation, 7*(25). Retrieved on 10.09.2015 from http://PAREonline.net/getvn.asp?v=7&n=25

Mestre, P. J. (2002). Probing adults' conceptual understanding and transfer of learning via problem posing. *Applied Developmental Psychology, 23*, 9-50. doi: 10.1016/S0193-3973(01)00101-0

Moskal, B. M. (2000). Scoring rubrics: What, when, and how? *Practical Assessment, Research & Evaluation*, 7(3). Retrieved from http://PAREonline.net/getvn.asp?v=7&n=25 on 10.09.2015

Moskal, B. M., & Leydens, J. A. (2000). Scoring rubric development: Validity and reliability. *Practical Assessment, Research & Evaluation*, 7, 71-81.

Nakamura, Y. (2004). *A comparison of holistic and analytic scoring methods in the assessment of writing*. The Interface between Interlanguage, Pragmatics and Assessment: Proceedings of the 3rd Annual JALT Pan-SIG Conference, Tokyo, Japan. Retrieved from http://jalt.org/pansig/2004/HTML/Nakamura.htm on 06.04.2015

Newell, J. A., Dahm, K. D., & Newell, H. L. (2002). Rubric development and interrater reliability issues in assessing learning outcomes. *Chemical Engineering Education*, 36(3), 212-215.

Nitko, A. J. (2004). *Educational assessment of student*. Upper Saddle River, NJ: Pearson Education.

Oakleaf, M. (2009). Using rubrics to assess information literacy: An examination of methodology and interrater reliability. *Journal of the American Society for Information Science and Technology*, 60(5), 969-983. doi: 10.1002/asi.21030

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

Rincker, S. J. (2002). *Developing core French rubrics to evaluate student progress and performance: an action research study*. Unpublished master thesis, University of Regina, Regina, Canada.

SAS. (2006). Compute estimates and tests of agreement among multiple raters. Retrieved from http://support.sas.com/kb/25/006.html on 06.04.2015

Shepard, M. M. (2005). *The effect of the use of a rubric in teacher assessment*. Unpublished doctoral dissertation, Boston College, Chestnut Hill, ABD.

Silver, E. A., & Cai, J. (1996). An analysis of arithmetic problem posing by middle school students. *Journal of Research in Mathematics Education*, 27(5), 521-539.

Silver, E. A., & Cai, J. (2005). Assessing students' mathematical problem posing. *Teaching Children Mathematics*, 12(3), 129-135. (ERIC No: EJ749674)

Silvestri, L., & Oescher, J. (2006). Using rubrics to increase the reliability of assessment in health classes. *International Electronic Journal of Health Education*, 9, 25-30. (ERIC No: EJ794114)

Spandel, V. (2006). In defense of rubrics. *English Journal*, 96(1), 19-22.

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

Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research & Evaluation*, 9(4). Retrieved on 10.09.2015 from http://PAREonline.net/getvn.asp?v=9&n=4

Thaler, N., Kazemi, E., & Huscher, C. (2009). Developing a rubric to assess student learning outcomes using a class assignment. *Teaching of Psychology*, 36, 113-116. doi: 10.1080/00986280902739305.

Tierney, R., & Simon, M. (2004). What's still wrong with rubrics: Focusing on the consistency of performance criteria across scales levels. *Practical Assessment, Research & Evaluation*, 9(2). Retrieved on 10.09.2015 from http://PAREonline.net/getvn.asp?v=9&n=2

Truemper, C. M. (2004). Using scoring rubrics to facilitate assessment and evaluation of graduate level nursing students. *Journal of Nursing Education*, 43(12), 562-564.

Tuncel, G. (2011). Sosyal bilgiler dersinde rubriklerin etkili kullanımı. *Marmara Coğrafya Dergisi*, 23, 213-233.

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

Zimmaro, D. M. (2004). Developing grading rubrics. *Measurement and Evaluation Center*. Retrieved 06.04.2015 from http://slo.sbcc.edu/wp-content/uploads/developing_grading_rubrics.pdfs on

Appendix 1: Rubric for evaluating problem posing (REPP)

Problem Oluşturmayı Değerlendirme Rubriği (PODER)

Öğrencinin Adı-Soyadı:				**Problem No:**	**Katsayı**	**Puan**
I **Problem Metni (Dil ve Anlatım)**	**1** Problem metni, açık ve anlaşılır değildir. ◯	**2** Problem metni, kısmen anlaşılırdır. ◯	**3** Problem metni, anlaşılır fakat istenen belli değildir. ◯	**4** Problem metni, açık, net ve anlaşılırdır. ◯	7	
II **Problemin Matematik İlkeleriyle Uyumu**	**0** Problem, matematik ilkelerine uygun değildir. ◯	**1** Problem, matematik ilkelerine kısmen uygundur (yanlış kavram kullanımı gibi). ◯	**2** Problem, matematik ilkelerine uygun fakat gerçek hayata uygun değildir. ◯	**3** Problem, matematik ilkelerine ve gerçek hayata uygundur. ◯	8	
III **Problemin Türü/Yapısı**	**1** Basit alıştırma türündedir. ◯	**2** Alıştırma türündedir. ◯	**3** Basit normal/sözel problem türündedir. ◯	**4** Normal/Sözel problem türündedir. ◯	6	
IV **Problemin Çözülebilirliği**	**0** Problemdeki veriler ve bilgiler, problemin çözümü için yeterli değildir. ◯	**1** Problemde verilenler yeterli olmasına rağmen çok karmaşık olduğundan çözülemez. ◯	**2** Problem çözülebilir fakat veriler hatalıdır/eksiktir. ◯	**3** Problemdeki veriler ve bilgiler tam ve uygun olduğundan problem çözülebilirdir. ◯	8	
NOT: *II ve IV numaralı boyutlardan sıfır alan problemler değerlendirmeye alınmayacaktır.*					**Toplam Puan**	