Eurasia Journal of Mathematics, Science & Technology Education, 2010, **6**(3), 173-186

Alternative Mathematics Assessment: A Case study of the Development of Descriptive Problems for Elementary School in Korea

Min Kyeong Kim and Sunsook Noh
Ewha Womans University, Seoul, KOREA

Received 11 September 2009; accepted 25 April 2010

A focus on mathematical understanding and problem solving in math education has developed a need to implement alternate ways of testing to better assess the students' understanding and problem solving skills. The seventh national school curriculum (MOE, 1997) for mathematics in Korea was revised to the current curriculum (MOEHRD, 2007) to emphasize alternative assessment methods and particularly the descriptive assessment method. However, the implementation process has been slow due to the lack of support to help teachers to prepare, develop and grade the new descriptive problems. In order to help teachers implement the new assessment framework, we carried out a study with students of grade 2, 4 and 6 to test and refine a framework for developing and grading descriptive problems. Analysis and comparison of the data showed that the grading rubric based assessment of the descriptive problems revealed useful information about the quality of problems and the level of understanding of the students. The study showed that a structured approach to descriptive problem assessment can be a powerful tool for improving math education in the elementary school.

Keywords: Elementary school mathematics; mathematics assessment; descriptive mathematics problems; scoring rubrics

INTRODUCTION

With the advent of knowledge-based society that uses information as its core source for social development, the purpose of education has shifted from simply delivering knowledge and information to student-centered education that focuses on fostering creativity and enhancing problem solving skills. The new educational environment respects the individuality of each student and embraces the variety of students as a source of inspiration and power.

Correspondence to: Min Kyeong Kim, Associate Professor, Department of Elementary Education, Ewha Womans University, Seoul, KOREA
E-mail: *mkkim@ewha.ac.kr*

ISSN: 1305-8223

Today's mathematics education philosophy is influenced by the constructivists' perspective which states that students should learn to connect, organize and integrate mathematical knowledge in order to actively construct their own learning experience. In accordance with this current view, evaluation in mathematics curriculum now focuses more on assessing problem solving skills and advanced mathematical thinking skills such as reasoning, communications and mathematical connection skills (National Council of Teachers of Mathematics [NCTM], 2000).

Fundamental changes in basic educational philosophy bring about not only changes in educational content but also in the overall educational programs which demand changes in the school curriculum. In particular, changes are needed in student evaluation methods to be aligned with the constructivists' view

State of the literature

- The need to make changes in student evaluation methods from ‘assessment of learning’ through ‘assessment for learning’ to ‘assessment as learning’ is widely supported by educators.
- Descriptive evaluation is recognized as an effective method that requires students to write down the problem solving process so that teachers could analyze what the students do not understand and help improve their understanding.

Contribution of this paper to the literature

- This study suggests a method for developing a framework for descriptive assessment including descriptive problems and scoring rubric for grades 2, 4, and 6 in elementary school mathematics in Korea.
- This study shows a detailed method for analyzing the students' response for the descriptive problems based on the scoring rubric established for the problems.
- This study shows the usefulness of analyzing the students’ response to better understand the degree of understanding achieved by the students and to understand the limitations of the problems as well.

point. It seems to have moved from ‘assessment of learning’ through ‘assessment for learning’ to ‘assessment as learning’ (Torrance, 2007, p.292).

In order to respond to such demands, the recently revised 7th national school curriculum for mathematics (Ministry of Education and Human Resources Development [MOEHRD], 2007) of Korea stated the goals in mathematics curriculum as follows; teaching mathematical knowledge and fostering mathematical thinking and communication skills that enable students to use mathematical reasoning in solving problems in everyday context. The 7th national school curriculum for mathematics was published in 1997 by the Ministry of Education (Ministry of Education of Korea [MOE], 1997) of Korea, and the current curriculum is a revision of the 7th curriculum (MOEHRD, 2007). The current national curriculum also emphasizes the importance of assessing the following; students' understanding of the basic concepts and principles in mathematics, the ability to correctly use mathematical definition and symbols, and the ability to observe, analyze, organize and think mathematically in solving problems.

In order to implement the 7th national curriculum in elementary schools, the Seoul Metropolitan Office of Education [SMOE] promoted many projects such as the 'New Wave Movement in Elementary School' and the 'Vision 2002: Creation of New School Culture'. Along with the projects, recommendations were made to fundamentally change the performance assessment methods to enable teachers to better evaluate students' knowledge and skills as a whole (Hur, Baek, Park, Choi, Yang, & Kim, 1999). Recently, SMOE (2007) as well as 12 other school districts in the nation declared its education vision as fostering competent and creative individuals. In order to do that, the school districts made changes in the assessment methods to place less emphasis on evaluating traditional academic achievement and to focus more on assessing students' advanced thinking skills such as creative, logical and critical thinking skills. SMOE started a project in 2005 to implement and expand descriptive evaluations in elementary and secondary schools with a target of over half of the assessment scores coming from descriptive evaluation by 2007 and the teachers have been trying to follow the recommendations suggested by the school district office. In carrying out the program, middle schools in Korea actively implemented descriptive evaluation method in their mathematics curriculum. As of today, majority of middle school teachers say that they are using descriptive problems in written exams and the scores from the descriptive problems counted for more than 40% of the total grade of regular exams as suggested by the school district office.

However, the drive to implement descriptive assessment was not trouble free for teachers and the difficulty is greater for elementary school teachers who are not subject-matter (mathematics) teachers. Many elementary and secondary school teachers are still facing various problems related to implementing descriptive assessment as surveyed in our previous work (Kim, Kwon, Noh, Joo, & You, 2008; Noh, Kim, Cho, Jeong, & Jeong, 2008). The current difficulties for the teachers is at the same level as when the performance assessment was first recommended for the schools during the implementation of the 7th national curriculum (Seong, 2000).

Despite the wide spread efforts to change the overall philosophy of math assessment, many schools still evaluate only a part of the students' ability with multiple-choice and short answer question exams. These types of exams are convenient and objective but they are limited in assessing the student knowledge and ability. The traditional type exams encourage the students to memorize facts rather than encouraging them to actively think and come up with creative approaches to solving the problem. In the recent international PISA-2006 assessment, Korean students performed poorly in complex multiple choice and open-ended descriptive problems. The relatively low performance on open-ended descriptive problems suggests that Korean students are not skilled at effectively presenting their thought process (Lee, Sohn,

& No, 2007). Furthermore, the survey results in the TIMSS-2003 (The Third International Mathematics and Science Study) indicated that Korean students see little connection between mathematics and real life application. This lack of connection between math and real life may further cause difficulties for teachers and students in developing and solving realistic descriptive mathematics problems.

One possible reason for the difficulties encountered in widespread implementation of descriptive assessment maybe the lack of ready-made descriptive-problem type and practical guidelines for developing and grading descriptive problems by the teachers. The same type of situation occurred when the performance assessment was first mandated by the government without providing performance-assessment type of questions and practical guidelines for the teachers. The implementation plan suggested using descriptive problem without providing the enough resource for teachers to use. This lack of resource for teachers calls for the structured development of new assessment methods and descriptive problems that teachers could use to evaluate the mathematical ability of students.

Purpose and Research Question

The current focus in mathematics assessment is to use the assessment as a tool for enhancing teaching and learning for understanding. Descriptive problem assessment method has been recommended widely to help achieve this focus but implementation has been difficult due to a lack of framework for developing and grading the descriptive problems. In this research, a framework for developing descriptive problems and a rubric for assessment are suggested and tested to show how it can help the teachers and students improve teaching and learning for understanding. The research was carried out to determine if the descriptive problem assessment method framework and rubric can reveal useful information about tested problems and students' level of understanding that will be critical to improving math education.

Performance assessment in school mathematics

Information society in which we live in requires individuals to actively select appropriate information and use it creatively to solve problems in everyday life. Herman, Aschbacher, and Winters (1992) claim that school education should help students acquire not just the appropriate knowledge but also the skills that are needed to effectively implement and apply the acquired knowledge. The word 'assess' is originated from a French word 'assidere' which means 'to sit beside'. The word therefore suggests that assessment should be based on accurate observation of students and should be practiced in a way that is effective and applicable. Alberta Education of Canada (2006) defined assessment in schools as 'collecting information about students that help teachers in evaluating students' learning and development process'.

NCTM in its first curriculum guideline book 'Curriculum and Evaluation Standards for School Mathematics' stated that the main goal in mathematics education should be understood (NCTM, 1989). The 1989 Standards by NCTM placed greater emphasis on methods compare to contents and offered a new curriculum guideline that is devoted to improving students' problem solving skills. The Standards defined evaluation as 'a tool for implementing various standards and achieving systematic changes'. An essential goal of evaluation is to help teachers make decisions about instruction that could help students understand better. Furthermore, the Assessment Standards for School Mathematics (NCTM, 1995) presented six standards about exemplary mathematics assessment. They address how assessment should-

- *reflect the mathematics that students should know and be able to do;*
- *enhance mathematics learning;*
- *promote equity;*
- *be an open process;· promote valid inference;*
- *be a coherent process.*

In 2000, NCTM published PSSM (Principles and Standards for School Mathematics) which stated that assessment should support students' learning process and provide both students and teachers with rich and meaningful information. It means that assessment should support the learning of important mathematics and furnish useful information to both teachers and students.

The international trends in mathematics assessment therefore focuses more on evaluating the problem solving process, communication and reasoning skills and focuses less on assessing simple mathematical knowledge. Such advanced thinking skills could not be appropriately evaluated by uniform assessment methods such as multiple-choice questions. In order to assess advanced thinking skills, assessment methods that are specifically designed for evaluation of advanced thought skills are needed (Noh, Kim, & Kim (Eds.), 2003).

Effective use of formative assessment depends on teachers using their pedagogical reasoning so inevitably some teachers will have a better knowledge base to draw on than others (Webb & Jones, 2009, p.182). In this manner, the teachers, researchers, and professional developers view 'knowing what students know' as a critical component of effective classroom practice (Gearhart & Saxe, 2004, p.311).

Based on the international trend, Korea adopted performance-assessment as an alternative to simple knowledge testing assessment methods since the implementation of the 7th national school curriculum for mathematics. Performance assessment allows students to come up with a solution method and describe it in detail to present their thinking process. Performance assessment therefore may track and evaluate students' mathematical thinking ability (NCTM, 2000; Stenmark, 1991). The current descriptive evaluation process grew out of this performance assessment trend to provide a more structured approach to problem development and assessment.

Descriptive assessment as a performance assessment

Descriptive assessment method is the most commonly used performance assessment method in mathematics education in Korea. Descriptive assessment started with the implementation of performance assessment and then it received additional attention because it could help improve students' confidence and interest in mathematics (Noh & Ryu, 2001). In mathematics, descriptive assessment requires students to write down the problem solving process so that the teachers could analyze what the students do not understand and help improve their understanding as suggested by NCTM (2000). Studies have shown that students utilize various knowledge resources when solving open-ended questions than multiple-choice questions (O'Neil & Brown, 1998). Therefore, descriptive assessment involving open-ended questions could enhance student problem solving skills.

Descriptive assessment has several advantages over the traditional assessment methods. First, it is relatively easy to produce descriptive real life problems that would allow students to show their thought process. Secondly, the teachers could better understand the level of understanding achieved by the students from their description of the problem solving process. Assessment framework provides what to evaluate in different content areas of mathematics. Whang (2004) pointed out that mathematics assessment used commonly in Korea includes understanding of problem, problem solving process and the final answer.

Descriptive assessment does not ask simple questions that could be answered with a fact. Instead, they ask students to describe their problem solving process and evaluate the students' advanced thinking skills such as reasoning skills in mathematics. Whang (2004) suggested that descriptive assessment is one of the most effective evaluation methods because it allows teachers to know explicitly the thought process of the students. Descriptive assessment also has the advantage of objectively measuring the reliability and validity of the assessment methods by observing the differences between the responses of the students and the correct answers (Seong, 2000).

Descriptive assessment in school mathematics in Korea

The Seoul Metropolitan Office of Education (2010) recommended that descriptive assessment be implemented in the regular exams to account for 30% in 2010, 40% in 2011, and 50% in 2012 of the total grade in subjects such as Korean, Mathematics, English, Sociology and Science starting from 2005. The percentage of descriptive problems was to be gradually increased up to 50% (SMOE, 2007). After the recommendation of the school district, the adoption of descriptive problem in math courses grew rapidly. By 2007, many middle schools used descriptive assessment to account for about 40% of the total grade (Noh et al., 2008).

The recommendation from the school district to use descriptive assessment caused many practical problems for teachers who were not properly prepared and trained to adopt the new assessment method (Kim et al., 2008; Noh et al., 2008). Teachers commented in high school, descriptive assessment discouraged students who lacked reasoning and communication skills to the extent that they gave up studying mathematics all together. Some students in middle schools gave up on the exams and just turned in a blank answer sheet for the descriptive problems. Jeon (2006) also claimed that the descriptive problems that he used were less effective than the multiple-choice questions in the elementary classroom. Overall, the problems encountered at the school level have made it challenging to prove to the public that descriptive assessment could be objective and reliable.

Such problems rise partly because of the fact that implementation of descriptive assessment system was driven not by the needs and desires of the schools but largely at the request of the Ministry of Education and Human Resource Development of Korea. Most of the teachers agreed to include some portions of descriptive problems in regular mathematics exams even though they were not prepared in implementing the assessment method. To take full advantage of descriptive assessment system, it is important to know what the teachers think of the descriptive methods and to find a way to improve teachers' assessing skills.

In order to answer the questions in current descriptive assessment methods, students typically need to know all the mathematical concepts for the whole curriculum year. Such methods are problematic for both teachers and students. The teachers needed a way to develop descriptive problems that could be used after completion of each chapter in the curriculum.

Previous work on descriptive assessment

The current study was a follow up on the results of a previous study which was executed as a pilot study. The original 7th national school curriculum differentiated the first and second semester curriculum by naming them as Level 1 and Level 2 (MOE, 1997). However, the level names are no longer in used in the current revised national curriculum (MOEHRD, 2007) and the whole year program is called 'Mathematics for Grade 1' through 'Mathematics for Grade 10'. The courses for grade 11 are called 'Mathematics I' for liberal arts track students, and 'Mathematics II' for science track students. The courses for grade 12 are given as electives such as Calculus, Discrete Math, and Probability & Statistics. The 6 years of elementary grade mathematics serve as the basis for developing mathematical understanding which will then be used to build more complex mathematical ideas such as algebra in the seventh grade and other higher level mathematics in the high school grades.

For the previous pilot study, a framework for descriptive assessment was developed for grades 2, 4, 6 in elementary school, grade 7 in middle school, and grade 10 in high school. In Table 1, the overall focus and results of the study for grades 2, 4, 6 during second semester are summarized. For the study, an evaluation framework with example descriptive problems and grading rubrics were developed and tested in schools. The content areas for developing descriptive problems defined by the seventh national curriculum are (A) numbers & operations, (B) figures, (C) measurement, (D) probability & statistics, (E) variables & expressions, and (F) patterns & functions. The descriptive problems

Table 1. Summary of previous work for grades 2, 4, 6

Grade-Level	Reference	Study Focus
2-2	Cho, Kim, Kwon, & Noh (2008)	1) Develop descriptive problems and grading rubrics 2) Analyze grades and answers by students in the 5 content areas
4-2	Hong, Kim, Noh, & Kwon (2008)	1) Develop descriptive problems and grading rubrics 2) Analyze grades and answers by students with respect to the three areas of evaluation (understanding-of-problem, problem-solving-process, and communication/representation-skills)
6-2	Kim, Noh, Kwon, Kim, & Joo (2008)	1) Develop descriptive problems and grading rubrics 2) Analyze grades, answers by students and misconception revealed by students with respect to the three areas of evaluation (understanding-of-problem, problem-solving-process, and communication/representation-skills)

Table 2. Classes for pilot and main study

Grade-Level	Round 1			Round 2		
	class	# of students in class	Test Units	class	# of students in class	Test Units
2-1	2-A	30	Ch. 1~4	2-C	29	Ch. 1~4
	2-B	28	Ch. 5~8	2-D	30	Ch. 5~8
Subtotal	2	58	all chapters in grade 2-1	2	59	all chapters in grade 2-1
4-1	4-A	29	Ch. 1~2, 5~6	4-C	31	Ch. 1~8
	4-B	31	Ch. 3~4, 7~8			
Subtotal	2	60	all chapters in grade 4-1	1	31	all chapters in grade 4-1
6-1	6-A	30	Ch. 1~4, 9	6-C	32	Ch. 1~5
	6-B	31	Ch. 5~8	6-D	32	Ch. 6~9
Subtotal	2	61	all chapters in grade 6-1	2	64	all chapters in grade 6-1
Total	*6*	*179*	*all chapters*	*5*	*154*	*all chapters in grades 2-1, 4-1, 6-1*

in each content area for the targeted grade levels were developed and then the grading rubrics for grading the students' solution in a consistent manner were developed. The grading rubric graded the solutions in 3 specific areas. A total of 10 maximum possible points were awarded to each problem by assigning 2 maximum possible points for understanding the problem, 6 maximum possible points for problem solving and 2 maximum possible points for communication/representation skills. The data from the testing were then summarized and analyzed to better understand and improve both the development of the descriptive problems and the grading rubrics. In Table 1, the focus of the study for each of the grade level and the average scores of the 3 assessment areas are summarized with brief description of the findings from the analysis. The learning from this initial pilot study was then used to design the present study for the first semester classes for grades 2, 4 and 6.

METHODS

Participants

The participants of the present study were 333 students (179 for round 1, 154 for round 2) from 4 classes in the second grade, 3 classes in the fourth grade, and 4 classes in the sixth grade. Table 2 describes in detail the grade-levels, classes, number of students and curriculum units for both the first and second round of tests.

Procedure

After reviewing the findings from the previous pilot study, three descriptive assessment problems for each chapter (unit) were developed and tested in a 40 minute test after the completion of the chapter for the first round of testing. Both round 1 and round 2 of testing were carried out during the 1st semester of the academic year from March 2008 to July 2008. The round 1 exam problems were graded, reviewed and analyzed to make appropriate adjustments to refine the problems and rubric for the round 2 exam problems. For round 2, three descriptive problems for each chapter were tested again in a 40 minute test after the completion of the chapter. The collected answers were graded by three researchers in three categories as suggested in the developed and revised rubric.

Descriptive problems were developed to comply with the goals of the seventh national school curriculum for mathematics that emphasize the improvement of mathematical power and integrated approach to problem solving. Developers of 'Everyday Mathematics' in the U. S. which is an elementary mathematical curriculum based on the Standards, established development principles that include essential principles of design, teaching, learning and evaluation (Bell & Isaacs, 2007). Our research also established the development principles of the descriptive assessment framework:

- *Evaluate problem solving process by assessing how students use and apply mathematical concepts*
- *Evaluate how students express and communicate/represent mathematically in the problem solving process*
- *There is one right answer but various approaches is possible*
- *Present problems that are as closely related to the real world as possible*

Evaluate how students interpret and solve problems

Based on the contents in grade-levels 2-1, 4-1, and 6-1 of the core curriculum in the seventh national curriculum, three descriptive problems for each chapter were developed. Developed problems were reviewed by mathematics education experts and elementary and secondary school teachers. The total number of units was 25, and therefore the total number of problems was 75. There were 30 problems in the strand of numbers & operations, 15 in figures, 12 in measurement, 9 in variables and expressions, 6 in pattern & functions, and 3 in probability & statistics.

The present study was based on the seventh national curriculum for school mathematics (MOE, 1997) even though the revised curriculum was announced in the time of the research but not yet implemented in school sites. There were 6 content strands in the 7th national school mathematics curriculum for core curriculum of grades 1-10:

(A): Numbers & Operations
(B): Figures
(C): Measurement
(D): Probability & Statistics
(E): Variables and Expressions
(F): Patterns & Functions

Advanced concepts in these strands are further discussed in grade 11, and students take one of elective courses such as calculus, discrete mathematics, probability & statistics whatever the courses are offered by the school. Most of the high schools tend to offer Calculus as the elective math course for grade 12 students who are in academic track in the science, mathematics, and engineering in addition to medicinal sciences (MOE, 1997). However, there was a slight change in classification of strands for each grade in the revised seventh national curriculum for school mathematics (MOEHRD, 2007) which was published in the time of the present study without being yet implemented in school sites. The strand connected to

Table 3. Assessment framework for a descriptive problem in mathematics

Area	Understanding-of-Problem (max=2 pts)	Problem-Solving-Process (max=6 pts)	Communication/Representation-Skills (max=2 pts)
Criteria	- degree of understanding mathematical concepts in the problem -degree of using and applying the given information in the problem	- accuracy of setting the strategy and carrying it out - degree of describing and interpreting the answer in the context of the problem after carrying out strategies	- accuracy of contents, ideas and symbols expressed in the process - degree of sufficient explanation in problem solving
Characteristics of sub-criteria			
	2	6	2
	- complete understanding of the mathematical concept given in the problem and application of the concepts in problem solving - ability to select the necessary information from the problem	- no error in completion of the strategy and computation - accurate expression of the answer in the problem context	- sufficiently explained the process that there is no need to presume anything in level of process - accurate representation of mathematical concepts and symbols
		4	
		- used appropriate strategy but answer is not correct a) mistakes in computation or writing down the answer b) the number in the answer is correct but the unit is wrong or omitted c) no answer	
	1	2	1
	- partial understanding of the mathematical concepts in the problem - partial use of information given in the problem that does not lead to complete answer	- used appropriate strategy but executed enough to get the answer - used appropriate strategy and executed enough to get an answer, but the answer is wrong - provided answer, but a) incomprehensible problem solving process or no process at all b) improper strategies are used or execution of strategy is not clear	- there is a partial gap in the logic used in the process - representation of mathematical concept, symbols are not clear in the problem solving
	0	0	0
	- not really understanding what the problem is asking - use of inappropriate concepts due to lack of understanding the problem - misuse of information that is irrelevant to the problem solving process - blank answer sheet, wrong answer or nothing	- selected strategy only leads to low level of achievement and not the right answer - blank sheet, wrong answer or nothing - complete lack of trials	- difficulty in understanding the problem solving process due to lack of clear explanations - difficulty in understanding the problem solving process due to inaccurate use of mathematical concepts and symbols - complete absence of explanations or some explanation irrelevant to the problem solving
	Possible max : 2 points	Possible max : 6 points	Possible max : 2 points

each unit is specified in terms of 6 strands given in the 7th national school curriculum for mathematics. There are more units covering the contents related to the strands in the order of A > B > C > E > F > D in the first semester of grades 2, 4, 6.

Mathematics teachers are supposed to develop test problems in both descriptive and multiple-choice formats twice a semester. That is why we tried to develop exemplary descriptive mathematics problems for all content area that are being taught in school mathematics and to provide them to school teachers so that they can gradually adjust to a new assessment system which has been practiced rather abruptly in the school site. It was another goal of the current study to develop concrete as well as good problems so that teachers can easily modify in their own teaching-learning situation.

TIMSS-2003 test problems were developed based on the '2003 TIMSS Assessment Framework and Specification' (Mullis, Martin, Smith, Garden, Gregory, Gonzalez, Chrostowski, & O'Connor, 2003). We also developed our descriptive assessment framework based on a review of literature. The three assessment areas are the followings:

1) understanding-of-problem
2) problem-solving-process
3) communication/representation- skills

These three categories are thought to be most important based on previous literatures that studied assessment frameworks (Bell & Isaacs, 2007; Kim & Cho, 2006; Leatham, Lawrence, & Mewborne, 2005; Martin, Mullis, & Chrostowski (Eds.), 2004; NCTM, 1995; Whang, 2004).

The first area, understanding-of-problem evaluates how much the student understands the mathematical concepts included in the problem and how well the

Table 4. Internal consistency (Cronbach α)

Grade-Level	Rater	Coefficient
2-1	A	.841
	B	.839
	C	.842
4-1	A	.864
	B	.875
	C	.863
6-1	A	.829
	B	.822
	C	.834

Table 5. Inter-rater reliability in each grade

Grade-Level: 2-1	Rater	Rater		
		A	B	C
Total Score	A	1		
	B	.908**	1	
	C	.913**	.879**	1
Grade-Level: 4-1	Rater	Rater		
		A	B	C
Total Score	A	1.000		
	B	.931**	1.000	
	C	.902**	.907**	1.000
Grade-Level: 6-1	Rater	Rater		
		A	B	C
Total Score	A	1.000		
	B	.913**	1.000	
	C	.833**	.834**	1.000

$^{**}p<.01$

student understands the presented information. The second area, problem-solving-process evaluates how well the student decides on the strategy to solve the problem, how accurately they carry it out, and how well they answer the question. The third area, communication/representation-skills area evaluates how accurately the contents, ideas and symbols are used in the problem solving process and how effectively they explain their process. The detailed criteria for assigning points for each of the category are summarized in Table 3.

Content validity of developed problems and reliability of assessment framework

In the design and development stage, we analyzed the characteristics of descriptive problems and assessment framework from the previous pilot study. The developed problems using the assessment framework were then tested by 3 different raters who were elementary teachers or people with graduate degrees in mathematics education. To guarantee the internal consistency among the 3 raters in each grade, a Cronbach reliability coefficient analysis was used. Table 4 summarizes the coefficients for each of raters ranging from .822 to .875 which indicated good internal consistency for the raters.

In addition, the inter-rater reliability was evaluated by the Pearson correlation coefficient method in each grade. The correlation coefficients ranging from .833 to .931 summarized in Table 5 showed good inter-rater reliability with the level of significance at $p<.01$.

RESULTS AND DISCUSSION

Example of a developed descriptive problem, rubric, and students' responses

The table 6 is an example of a problem for grade 2-1. The sample problem was given when the 4th chapter in the textbook was completed. The 4th chapter covered ‘addition and subtraction of two-digit numbers’ in the strand of (A) numbers & operations. We call the sample problem by the name 'Two-Digit-Number' problem.

To solve the sample problem above, a student must express the situation in a mathematical equation and apply the rule of subtraction. We set up the scoring rubric for the understanding-of-problem, problem-solving-process and communication/representation skills based on students' ability to use the appropriate mathematical equation, solve the subtraction problem with borrowing and use the information given in the problem.

First, for the understanding-of-problem criteria, we intended to evaluate how well students identify the mathematical concepts needed to solve the problem and how well they understand and apply the information given in the problem. The scoring rubric for the understanding-of-problem area is specified into 0, 1, and 2 based on whether the student expressed the problem in mathematical equation and appropriately used the information given in the problem.

To solve the 'Two-Digit-Number' problem in table 6 , it is necessary to understand information such as the number of students, bus seats and teacher and set up the equation where subtraction is needed. The full score 2 was given to students who correctly answered 8 seats through writing and solving the equation $36+1+\square=45$ and $\square=45-1-36=8$. Such an answer deserves the full score for the student showed "complete understanding of the mathematical concepts given in the problem and application of the concepts in problem solving" according to the scoring rubric. The full score 2 was also given to students who did not use the equation directly but expressed the appropriate equation in words

Table 6. 'Two-Digit-Number' Question

Grade-Level	2-1		
Chapter(Unit)	4. Addition and subtraction of two-digit numbers	Content Strand	Numbers & Operations
Subject	Subtraction of two-digit numbers with borrowing		
Standards	Students can subtract two-digit numbers with borrowing		
Evaluation Criteria	Assess whether student understand the problem context, set up the equation to solve the problem, understand the principle of two-digit numbers subtraction with borrowing, and compute the subtraction operation.		
1. Sujin's class is going to an amusement park with their teacher. There are 36 students in her class including Sujin. There are 45 seats in one bus. (1) Describe the solving process to find out how many seats are left unoccupied after <u>all</u> <u>of the students in Sujin's class and the teacher take a seat</u> in the bus. (2) How many seats are left unoccupied?			

and used all the necessary information of the problem. We gave the score 1 to students who intended to compute the number of remaining bus seats but only used a part of the necessary information and failed to completely solve the problem. Students received the score 0 if they did not properly understand the problem and used the inappropriate information.

Second, for the problem-solving-process area, we intended to evaluate how well students set the strategy and correctly carry it out and also how well students describe and interpret their answer. To evaluate such abilities, we specified the criteria based on how coherently and accurately the process of the chosen strategy is connected to reaching the final answer. In detail, we specified the scoring rubric into 0, 2, 4, and 6 based on how well students use the information presented in the problem and express it as a mathematical equation, how precisely students apply the rule of subtraction and solve the subtraction problem with borrowing, and finally how appropriately students present their answer. We gave the full score 6 if a student chose a strategy where he or she subtracted the number of students and the teacher, carried out the computation without any error and reached the final correct answer. We gave the score 4 when a student chose a strategy like the one and showed some errors in the problem solving process such as wrongly adding the teacher instead of subtracting her. In this example, the student managed to get the correct answer but failed to show the appropriate computation process which is 45-36-1 and used the incorrect equation 45-36+1=8. We gave the score 2 when a student intended to use subtraction but did not correctly use the information presented in the problem and as a result set up the wrong subtraction which led to a wrong answer. We gave the score 0 when a student as in Fig. 10-4 did not understand what the problem was demanding and wrote down irrelevant information that has no direct link to solving the problem.

Third, for the communication/representation skills area, we intended to evaluate how precise and clear the mathematical contents, ideas and symbols shows in the problem solving process and whether students offered sufficient explanation. We specified the scoring rubric into 0, 1, and 2 based on the sufficiency and clarity of the mathematical explanation shown in the problem solving process. We gave the full score 2 when a student sufficiently explained why he chose the subtraction as his strategy, revealed how he was going to use the information given in the problem and offered a logical explanation of the solving process. In this example, the student explained in detail why several steps of equations are needed in order to solve the problem. The student managed to logically present his problem solving process based on precise understanding of the problem.

If a student failed to show why he or she chose the subtraction as the strategy or used the equation 45-37 but did not explain why subtracting 37 is necessary, we determined that the explanation is not logical enough and gave the score 1. In this example, the student failed to show that he or she added the number of students (36) and the teacher (1) and got the number 37 which the student then subtracted from 45. The student simply wrote 45-37=8. The student hence answered with lack of sufficient explanation and also had problems with using appropriate mathematical symbols and concepts which altogether led to the score 1. The score 0 was given to students who did not understand the problem properly and simply rewrote the problem or provided information that is irrelevant to answering the problem. The answer received the score 0 for we could not find any evidence of appropriate problem solving process.

Results and discussion of assessment in grades 2, 4, and 6.

The results of the study were analyzed by averaging the breakdown scores for each problem for all students and then comparing the average breakdown scores for each problem and each breakdown category. The overall effectiveness of using the breakdown scores was analyzed by comparing the trends of the three grades tested. In order to summarize and represent all the data in a simple manner, the average scores were plotted as problem-solving-process versus understanding-of-problem as shown in Figure 1, and communication/representation-skills versus understanding-of-problem as shown in Figure 3 for all three grades tested.

In general, the plots show a consistent linear correlation between understanding-of- problem versus problem-solving-process and communication/ representation-skills. Clearly for most of the problems for all three grades, a high score in understanding-of-problem meant high scores in both problem-solving-process and communication/representation-skills. While the overall correlation is good, there were some data points that do not follow the correlation line.

Figure 1 shows that for the second grade, some test problems showed low average score on the understanding-of-problem but a high average score on the problem-solving-process. This deviation from the linear correlation line suggests that the students could solve the problem easily without fully understanding the problem. This suggests that the problem may not be properly setup to assess students' understanding since the students could solve the problem without properly understanding the problem. The problem appears to have been set up so that the students could solve the problem mechanically without truly understanding the problem.

On the other hand, there were few problems with high average scores on understanding-of-problem with lower than the expected correlation average score on problem-solving-process in grades 4 and 6. The characteristic of the score for these problems suggests that the students could understand the problem easily but did not have the proper tool or knowledge to solve the problem. The analysis of the test scores by the structured rubric of breaking down the answers by understanding-of-problem and problem-solving-process reveals a possible issue with the teaching of the subject matter.

Finally, the data in Figure 1 shows that the fourth grade problems were weighed on the lower scores of the correlation line. For the other grades, the distribution across the line was fairly even but the fourth grade tests showed greater number of lower scores than high scores along the correlation line. This suggests that the test

Figure 1. Plot of Understanding vs. Problem-Solving

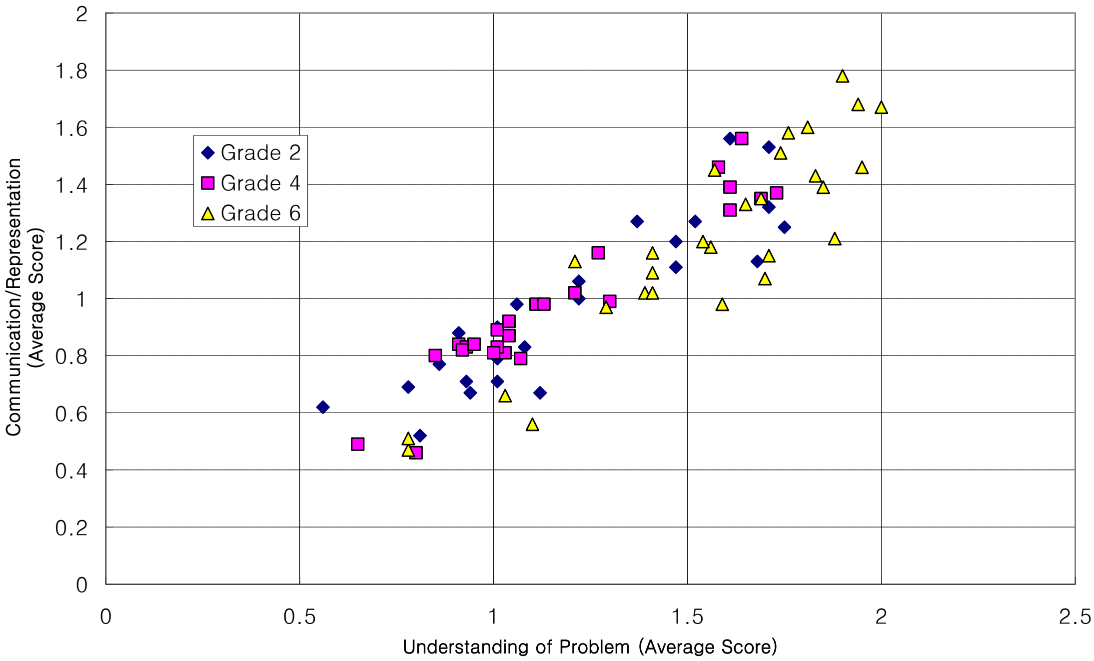

Figure 2. Plot of Understanding vs. Communication/Representation

problems were either more difficult for the students for the subjects covered or that the topics of the problems were not sufficiently covered before the test.

By analyzing and studying the average scores based on the grading scheme for the descriptive problems, more detailed information about the student's ability and the appropriateness of the test question can be derived. This can be a powerful tool for improving the test questions to better evaluate the students and to improve the teaching method to make sure that the students are learning the proper knowledge covered in the section.

The same type of linear correlation trend is also observed for Figure 2 which relates understanding-of-problem with communication/representation-skills. In general, there is a linear correlation for average score for communication/representation-skills with understanding-of-problem. But the data set around the trend line is slightly different from the previous data set around the trend line observed in Figure 1 for understanding-of-problem and problem-solving-process. While Figure 1 showed few data points away from the trend line indicating some issue with the problem, Figure 2 shows a general wider spread around the trend line indicating a poorer correlation in general. Also, there was no instance of low average score on the understanding-of-problem associated with high average score on communication/representation-skills. The data suggests that understanding-of-problem is an absolute requirement for communication/representation-skills. This observation makes consistent with the interpretation that if a student doesn't understand the problem, the student will not be able to communicate or represent the problem mathematically.

On the other side of the plot, higher average score on the understanding-of-problem does not necessarily reflect a higher communication/representation-skills score. Especially for the sixth grade scores, there were many problems where high understanding-of-problem score was associated with a low communication/representation-skills score. This is also consistent with the interpretation that at the sixth grade level, more complex mathematical ideas are difficult to communicate/represent even when the students understand the problem. More emphasis is needed in the classroom to teach the students to communicate effectively when expressing mathematical ideas. This result suggests a point to reconsider the claim made by Cai and Moyer (2008) in their recent paper. They claimed that students in earlier grades can handle generalization of concrete representations and strategies and therefore teachers need to support students with a smooth transition between arithmetic and algebra.

Another interesting observation from the fourth grade analysis showed the lowest communication/representation score in problem number 2 in Unit 7 where they first learn the concept of fraction. Fraction is one of the most difficult concepts to understand for elementary students. In learning fractions, the students have to expand (or generalize) the familiar number system, the ring of integers that they have been taught so far into a totally new and different number system, the field of rational numbers. The average score (1.95/10 points) of problem number 2 in the 'fraction' unit is the lowest in problem-solving-process (1.21/6 points) as well as in communication/representation-skill (0.46/2 points). It is also the second lowest (0.8/2 points) after problem number 2 in the 'problem solving' unit (0.65/2 points) among 24 problems for the fourth grade in the area of understanding-of-problem criteria.

CONCLUSION

The ability to select appropriate information from a vast pool of information and think rationally and mathematically in order to solve problems in the real world is more important than ever in the information society that we live in. Therefore mathematics education needs to deliver more than just mathematical knowledge and skills. It should help students acquire mathematical reasoning, communication skills and ability to rationally solve problems in real life. Mathematics education responded to such changes and needs by adopting performance assessment since the seventh national curriculum for school mathematics (MOE, 1997) was published. Among many types of performance assessment, descriptive assessment is the most widely adopted method for its practical and convenient implementation in Korea.

The revised seventh national curriculum for school mathematics in Korea also recognized the need for change in the purpose of mathematics education (MOEHRD, 2007). The focus was shifted from computation ability to mathematical reasoning that selects and uses appropriate information based on solid mathematical concepts and principles. However, mathematics education in Korea lacked practical and effective descriptive methods that could be readily used in the schools. Our study was designed to offer one possible solution to the problem. We developed descriptive problems and assessment framework for second, fourth and sixth grade mathematics and reviewed and revised the development by implementing them in a sample of schools and getting feedback from teachers and other experts in the field. The developed problems and assessment framework were based on feedbacks from actual implementation of the ideas in a sample of schools to offer concrete assessment guidelines together with detailed problem examples. This aspect of our research can help teachers who experience objectivity issues with descriptive evaluation.

We also recommended teachers to develop evaluation guidelines in order to exclude teachers' subjectivity from evaluation as effectively as possible. Three important implications were found from the study.

First, categorization with understanding-of-problem, problem-solving-process, and communication/ representation-skills as the three main areas of descriptive assessment framework was shown to be effective in analyzing the understanding level of students from three different grade levels. This finding is consistent with previous works in curriculum development literatures that argue that elementary mathematics curriculum developers should think of instruction and assessment as one integrated perspective (Bell & Issacs, 2007). Therefore wider adoption of descriptive assessment will enable further integration between evaluation and learning-teaching process so that assessment could play a more effective educating role in schools.

Secondly, results of our research show that the average score in descriptive assessment testing was the lowest in communication/representation-skill area compared to understanding-of-problem and problem-solving-process areas. Such result implies that Korean students appear to lack the ability to logically explain their thought process and effectively present their ideas even if they know how to derive the correct answers. The revised seventh national mathematics curriculum of Korea emphasizes mathematical communication skills as one of the primary goals of mathematics education. Therefore, we believe that descriptive problems are expected to play a significant role in making mathematical learning process more active in the classroom by giving students a chance to logically present their thoughts and communicate with the class using/representing the mathematical concepts and principles.

The analysis of the data showed many interesting details about the test problems, topics covered and students understanding. The detailed structured assessment was able to reveal that some test problems were not appropriate in terms of testing student's understanding. Some test problems showed that students did not appear to have been properly prepared in the topic to solve the problem even if they understood the problem. The data also showed that more emphasis was needed in preparing the students to effectively communicate mathematical ideas. The descriptive assessment with structured grading clearly showed the benefits of analyzing the students, test problems and teaching effectiveness. A sustained use of the assessment method with continuous improvement on the assessment and teaching method should help to improve the overall effectiveness of math education.

The overall analysis of the descriptive assessment scores with respect to topic areas revealed weaknesses in the curriculum coverage of certain topics throughout the grade levels in preparation for secondary school mathematics. For instance, the fourth graders' average test results showed the lowest score in the 'fraction' unit. The sixth graders' average test results showed the lowest scores in the area of communication/representation-skills (0.47/2 points) in the 'range of numbers' unit. These results provide a basis for evaluating Saul's claim (Saul, 2008) that algebra should be viewed as a generalization of arithmetic, study of binary operation, and study of the rational expressions. The students are first introduced to the field of rational numbers in the 'fraction' unit in grade 4 and then in the 'range of numbers' unit again in grade 6. The descriptive test results suggest that students are weak in both of these units which means that the students will have difficulties in understanding algebra later on in the secondary curriculum. By developing better descriptive problems in elementary grades focusing both arithmetic and algebraic thinking simultaneously, we can expect to reinforce the connection between arithmetic and algebra and to better understand the students' preparation level for understanding algebra in the future. The descriptive testing data should help to make the transition from elementary school mathematics to middle and high school mathematics easier (Cai & Moyer, 2008).

Acknowledgment

The work was supported by the Korea Research Foundation Grant funded by the Korea Government (MOEHR) (KRF-2007-721-B00059).

REFERENCES

Alberta Education of Canada (2006). *Effective student assessment and evaluation in the classroom: Knowledge and skills and attributes*. Edmonton, Alberta, Canada: Author. Availablefrom<http://www.teachingquality.ab.ca/resources/>.

Bell, M., & Isaacs, A. (2007). The case of Every Mathematics. In C. R. Hirsch (Ed.), *Perspectives on the design and development of school mathematics curricula* (pp.9-22). Reston, VA: NCTM.

Cai, J., & Moyer, J. (2008). Developing algebraic thinking in earlier grades: Some insights from international comparative studies. In C. E. Greenes & R. Rubenstein (Eds.), *Algebra and algebraic thinking in school mathematics*: *Seventieth yearbook* (pp.169-180). Reston, VA: NCTM.

Cho, M. K., Kim, M. K., Kwon, J. R., & Noh, S. S. (2008). A study on the development of descriptive problems in grade 2 mathematics. *The Journal of Elementary Education, 21*(3), 437-466.

Gearhart, M., & Saxe, G. B. (2004). When teachers know what students know: Integrating mathematics assessment. *Theory into Practice*, *43*(4), 304-313.

Herman, J. L., Aschbacher, P. R., & Winters, L. (1992). *A practical guide to alternative assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.

Hong, J. Y., Kim, M. K., Noh, S. S., & Kwon, J. R. (2008). A case study on the development of descriptive problems in grade 4 mathematics. *Journal of Educational Research in Mathematics, 18*(3), 335-352.

Hur, K., Baek, S., Park, K., Choi, M., Yang, G., & Kim, K. (1999). *Status and recommendations for performance assessment policy* (Research Report CRE-99-2). Seoul: Korea Institute of Curriculum & Evaluation.

Jeon, S. C. (2006). *Comparison study between descriptive and multiple-choice problems in performance assessment: Focused on the results of the national standardized test in middle school mathematics*. Unpublished master's thesis, Graduate School of Education, SungKyunKwan University, Seoul, Korea.

Kim, M. K., Kwon, J., R., Noh, S. S., Joo, Y. R., & You, H. J. (2008). A survey of the Teachers' perception and the status about the descriptive evaluation in elementary school mathematics. *School Mathematics (A Journal of Korea Society of Educational Studies in Mathematics), 10*(3), 401-422.

Kim, M. K., Noh, S. S., Kwon, J., R., Kim, E., & Joo, Y. R. (2008). Development of teachers' resource for descriptive evaluation in grade 6 mathematics. *Journal of the Korean School Mathematics Society, 11*(4), 543-567.

Leatham, K. R., Lawrence, K., & Mewborne, D. S. (2005). Getting started with open-ended assessment. *Teaching Children Mathematics, April*, 413-419.

Lee, M., Sohn, W., & No, U. (2007). *The results from PISA-2006* (Research Report RRE 2007-1). Seoul: Korea Institute of Curriculum & Evaluation.

Martin, M. O., Mullis, I. V. S., & Chrostowski, S. J. (Eds.). (2004). *TIMSS 2003 technical report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center.

Ministry of Education of Korea (1997). *The 7th national school curriculum for mathematics*. Seoul: Author.

Ministry of Education and Human Resources Development of Korea (2007). *The national school curriculum for mathematics*. Seoul: Korea Institute of Curriculum and Evaluation.

Mullis, I. V. S., Martin, M. O., Smith, T. A., Garden, R. A., Gregory, K. D., Gonzalez, R. J., Chrostowski, S. J., & O'Connor, K. M. (2003). *TIMSS assessment frameworks and specifications 2003*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center.

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

National Council of Teachers of Mathematics (1995). *Assessment standards for school mathematics*. Reston, VA: Author.

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

Noh, S., Kim, M. K., Cho, S., Jeong, Y., & Jeong, Y. (2008). A study of teachers' perception and status about descriptive evaluation in secondary school mathematics. *Journal of the Korean School Mathematics Society, 11*(3), 377-397.

Noh, S., Kim, Y. S., & Kim, M. K. (Eds.). (2003). *Curriculum development for mathematics and information science in a knowledge based society*. Seoul: Ewha Womans University Press.

Noh, Y. S., & Ryu, C. S. (2001). The effect of descriptive assessment as a performance assessment method: Focused on the grade 10 mathematics in high school. *Journal of the Korean School Mathematics Society, 4(*1), 125-136.

O'Neil, H. F. Jr., & Brown, R. S. (1998). Differential effects of question formats in math assessment on metacognition and affect. *Applied Measurement in Education*, *11*(4), 331-351.

Saul, M. (2008). Algebra: The mathematics and the pedagogy. In C. E. Greenes & R. Rubenstein (Eds.), *Algebra and algebraic thinking in school mathematics*: *Seventieth yearbook* (pp.63-79). Reston, VA: NCTM.

Seong, T. J. (2000). Challenges and recommendations for performance assessment in elementary school. *The Journal of Educational Research, 38*(1), 153-184.

Seoul Metropolitan Office of Education (2007). *Development of elementary standardized test problems to improve learning quality* (A text for the elementary teachers training program 2007). Seoul: Seoul Metropolitan Office of Education.

Stenmark, J. K. (1991). *Mathematics assessment: myths, models, good questions, and practical suggestions*. Reston, VA: NCTM.

Torrance, H. (2007). Assessment as learning? How the use of explicit learning objectives, assessment criteria and feedback in post-secondary education and training can come to dominate learning. *Policy & Practice*, *14*(3), 281-294.

Webb, M. & Jones, J. (2009). Exploring tensions in developing assessment for learning. *Assessment in Education*, *16*(2), 165-184.

Whang, W. H. (2004). Mathematics assessment in Korea. Paper presented (by the Korean Presentation Team) at the Tenth International Congress on Mathematical Education, July 4-11, 2004, Copenhagen, Denmark. Available from <http://www.mathlove.com/new3/>.

◈◇◈