

# The Study of Constructed-Response Assessment of Elementary Mathematics Education in Korea

Min Kyeong Kim & Mi Kyung Cho  
Ewha Womans University, SOUTH KOREA

Received 13 January 2015; accepted 19 January 2015; published 25 March 2015

Recently, many countries have considered various assessment methods in order to measure multiple ways of students' thinking skills and problem solving ability. Recently, the Ministry of Education, Science and Technology [MOEST] (2009) in Korea revised the mathematics curriculum, which now more focuses on enabling students to explain mathematically as well as to reflect on their thinking. This method allows students to express or debate on their own ideas in order to extend their mathematical communication skills. The purposes of this study are to introduce the background of such implementation of a constructed-response assessment in Korea as well as to investigate how elementary teachers implement a constructed-response assessment in developing and refining a framework. The developed grading rubric for the constructed-response problems in terms of a holistic approach was introduced. Further, the implication of applying constructed-response items to third-grade students in Korea was discussed.

*Keywords:* case of Korea, constructed-response assessment, elementary education, holistic approach, mathematics education.

## INTRODUCTION

In the 21<sup>st</sup> century, a knowledge- and information-based society, which is represented by keywords such as sudden change, diversification and pluralism, requires members of society to have the ability to generate creative ideas, cope with and solve problems and adapt to the changes of society. In order to accomplish this goal, the role of education is to teach citizens to fulfill their roles as members of society and furthermore, be able to develop a system where their potential can be cultivated. In that sense, education in school, from curriculum design to evaluation, must interact with and be complementary to each other. Overall, educational assessment undertakes the primary roles (William, 2007) such as supporting learning (formative), certifying

achievements (summative) and evaluating educational programs (evaluative).

Formative assessment is an assessment for learning, whereas summative assessment is that of learning (NCTM, 2011b; Slavin, 2009). Formative assessment provides teachers information about the degree of students' learning in order to make informed decisions on how to move on to the next step. Evaluations conducted at school can diagnose and prescribe the process of students' growth as well as foster their potential and talent.

Recently, evaluation performed at schools tends to combine quantitative assessment with qualitative assessment. Quantitative assessment is conducted in order to select, classify and allocate students; this realm of assessment received more interest in the past. Qualitative assessment can provide advice on growth and development by gathering and analyzing information regarding the progress and results of teaching and learning. These changes have also been applied to the evaluation of mathematics education in school. Accordingly, the evaluation includes problem solving skills and higher-order thinking, including reasoning, communication and connection skills.

Correspondence to: Min Kyeong Kim,  
Ewha Womans University, 52 Ewhayeodae-gil, Seodae  
mun-gu,, South Korea  
E-mail: mkkim@ewha.ac.kr  
doi: 10.12973/eurasia.2015.1331a

**State of the literature**

- Meaningful assessment makes it possible to provide feedback between students and teachers as well as motivate students.
- The assessment qualitatively evaluates focusing more on the higher-order thinking ability, which includes abilities such as problem solving, mathematical reasoning, mathematical communication and creativity of connecting mathematical knowledge with the knowledge of other subjects.
- Constructed-response assessment is a proper type that can evaluate the thinking process of students by asking them to write down how students understand the problem and how they solve the problem.

**Contribution of this paper to the literature**

- This paper reviews the status of the implementation of assessment in elementary mathematics education in Korea.
- This study shows a constructed-response assessment framework with a holistic rubric by combining it with an analytic rubric of three areas, such as understanding-of-problem, problem-solving-process and communication/representation-skills.
- This study shows the rubric for holistic scoring, which is applied to each item by transforming 4 levels ('Excellent', 'Good', 'Fair' and 'Need improvement') to 3, 2, 1 and 0 points in 3rd grade mathematics

With the advent of new learning theories, such as constructivism, mathematics curriculum reforms are taking place in most countries. The new curriculum places emphasis on the learning process, which also requires alternative modes of assessment. The current trend in assessment is moving from 'assessment of learning' through 'assessment for learning' to 'assessment as learning' (Black, Harrison, Lee, Marshall & Williams, 2003; Brookhart, 2005; Leahy, Lyon, Thompson & Wiliam, 2005). Today, such assessment makes it possible to provide feedback between students and teachers as well as motivate students. Further, it evaluates qualitatively by focusing more on the higher-order thinking ability, which includes problem solving ability, mathematical reasoning ability, mathematical communication ability and creativity of connecting mathematical knowledge with the knowledge of other subjects.

Traditional assessment, such as typical multiple-choice, true or false, short answer questions, etc., has been criticized because it did not properly evaluate

students' learning, but only measured their test-taking skills. Moreover, such assessment only covered compartmentalized knowledge-based questions, which the students could answer by memorization or recall. Eventually, students who were accustomed to these assessments became passive learners (Diane, 1994). To overcome these critics, a number of new approaches toward assessment had been generated under such names as alternative assessment, performance assessment, authentic assessment, etc. As Gearhart & Saxe (2004) stated, it must be important that 'knowing what students know' is a critical component of effective classroom practice.

Many countries use the constructed-response assessment in mathematics because it can evaluate students' problem solving process and furthermore, it is easier to create constructed-response problems compared to any other assessment types for the test (Hur et al., 1999; Seong, 2000). Bell & Isaacs (2007), who are elementary mathematics curriculum developers, noted that a constructed-response assessment is a proper type, which can evaluate the thinking process of students by asking them to write down how students understand the problem and how they solve the problem. Park & Pang (2008) also presented some specific types of a constructed-response assessment, such as explaining mathematical concepts and principles with logically reasonable evidence along with the meaning of the mathematical formula with diagrams or writings, explaining the mathematical sequence and connectedness, problem posing, explaining problem solving strategies, using mathematical words in real life situations in order to solve problems, and writing down how students interpret the data with diagrams or sentences, etc., in order to support the suitability of the constructed-response assessment for evaluating the thinking process.

The existing traditional evaluation methods do not allow for the students' mathematics learning process to be understood. Also, such methods are not suitable to implement and strengthen the creativity. With a newly released personality-based curriculum, constructed-response assessment have come out to measure how much students enhance in the learning of mathematics. Assessment is a process of gathering information with regard to mathematical, conceptual and procedural knowledge and attitudes that a student has ([NCTM], 1995). Thus, assessment is a factor that completes the mathematics class as well as an important criterion that enhances students' mathematics learning (NCTM, 1995, 2000). So far, assessment has been understood as a tool to diagnose a student's mathematics achievement through exam papers rather than as a key element of math class. NCTM's documents (1995, 2000) containing a number of "principles and standards" suggest what should be taught in school math class as well as how to

teach it. In particular, they emphasize the math learning processes, such as learners' problem-solving skills, reasoning skills, communication skills, representation skills, and connection skills.

This change is reflected in the international academic achievement test. In TIMSS 2007, 45.6% of the problems were constructed-response assessment problems in the constructed response items, and some of them were constructed with multi-part questions asking to explain the problem solving process considering the problem situation. Interest regarding constructed-response assessment has been increasing since 2008. Further, constructed-response items have a tendency of questions that are needed to apply some specific mathematical knowledge or principle in real life situations as well as to reason logically. Moreover, they also contain several sub-problems that can convey students' problem solving strategies and logical thinking ability step by step (Cho et al., 2011).

TIMSS includes constructed-response items that apply to everyday life or reasoning based on the understanding about mathematical concepts or principles. The scoring system of constructed-response items consists of recording what students describe in the process of problem solving, so that what they know and what they can do will be accounted in their thinking process (Mullis & Martin, 2011). In the case of fourth-grade students, constructed-response items comprised 45.6% (year of 2007) and 47.2% (year of 2011) of the problems (Cho et al., 2011), and in 2015, the figure will increase to 50% (Martin, Mullis, & Foy, 2013).

In PISA 2009, 31.5% of the problems were constructed response items. One type included closed-constructed-response items, whose answer scope was restricted, and the other comprised of open constructed-response items, which required the skill to link the information and ideas of problem passages with students' experience or opinion to solve the problem (Kim et al., 2010). PISA 2012 included 63.5% of open-constructed response items. Overall, because students need to solve constructed response items using their knowledge and experience of mathematical concepts and skills, the items are appropriate as a method of investigating how to connect the knowledge acquired in school to everyday life.

### **The current implementation status of constructed-response assessment in Korea**

In order to respond to demands, the Ministry of Education, Science and Technology [MOEST] (2009) in Korea recently revised the national school mathematics curriculum for mathematics, which now more focuses on enabling students to use mathematical definitions and symbols correctly and to explain mathematical ideas by talking or writing. This method allows students to

express or debate on their own ideas in order to extend their mathematical communication skills.

The assessment encourages the use of various types of assessment as well as emphasizes the characteristics of problems. Moreover, it supports the notion of including high quality problems on tests because they can evaluate students' thinking process and output. In particular, students require useful feedback on time in order to organize several factors needed for problem solving and to make decisions as well as apply them to problems (Slavin, 2009). Therefore, understanding the strengths of a constructed-response assessment and investigating how it is implemented in a school environment are essential.

Constructed-response assessment evaluates how students understand, explain and interpret the problem situation with their already known information and knowledge; moreover, it measures how they solve the problem creatively, unlike the traditional test, which only measures how many students could memorize knowledge. The 7th revised national school mathematics curriculum of Korea states that not only students' problem solving process but also the results should be evaluated in order to increase their mathematical thinking ability in a cognitive domain. Further, the specific goals are as follows; the ability of understanding and applying fundamental concepts, principles and rules of mathematics; the ability of using mathematical definitions and symbols correctly; the ability of reasoning properly; the ability of solving problems by understanding the problem situation mathematically; the ability of problem solving by thinking mathematically in real life situations, social phenomenon and natural phenomenon; and the mathematical ability of communicating the thinking process and output reasonably (MOEST, 2009, Seoul Metropolitan Office of Education, 2010, 2011). A constructed-response assessment helps to evaluate how students formulate, organize, internalize and explain their mathematical concepts and thinking process (Baek, 2000). Also, it helps to understand students' level of understanding the problem as well as correct errors by analysing students' answers and problem solving processes (NCTM, 2001).

This study focused on the characteristics of problems or tasks. Lappan, Phillips & Fey (2007) emphasized that 'good problems' should be developed, and defined good problems as including important mathematical ideas and needing to use higher-order thinking skills and various solving methods. In Korea, teachers should reflect on the characteristics of 'good problems' in order to develop constructed-response assessment problems. The revised 7th national school mathematics curriculum of Korea says that the results of the assessment should provide useful information for both teachers and students, and should contribute to the

teaching-learning process. In other words, the assessment results need to convey students' strength and weakness after learning, followed by analysing the results of errors, which will help to decide the direction of the next class (MOEST, 2009).

The most important point here is the connection between the teaching-learning process and assessment. Traditionally, students were classified according to their scores and students' ranks were compared; however, today, it is important to verify whether the students understand what they have learned or not and whether they can reach their goals or not. With this view of assessment, in Korea, traditional assessment forms, such as the mid-term exam or final exam, has been changed to a formative assessment that can help teachers to verify students' understanding during class and further, these exams can often be implemented when teachers need them. This was supported in the same context of Kim, Cho & Joo (2012) and Kim, Cho, Kim, Kim & Noh (2012) that the constructed-response assessment could help to verify the results of the teaching-learning process and make improvement in a mathematics class. Because the constructed-response assessment is a method in which students are required to expose the types of concepts and strategies used in expressing their problem solving process, the assessment would be a type of method to implement formative assessment.

## DEVELOPMENT AND APPLICATION OF CONSTRUCTED-RESPONSE ASSESSMENT ITEMS

### Principles of constructed-response assessment items

Through a review of the existing studies with regard to the constructed-response assessment, this study defines it as an 'assessment in which various strategies can be used with problems by evaluating how students organize and internalize a concept or a process instead of asking for their fragmentary knowledge'.

The developmental process of a constructed-response assessment problem and rubric was designed, as shown in Figure 1, from a comprehensive point of view of preliminary studies (e.g., Cho Kim, Kwon & Noh, 2008; Kim & Noh, 2010). According to the developmental process, the developed preliminary items were examined for content validity, term relevance and field applicability of items by elementary math education experts as well as by teachers with extensive teaching experience. In addition, the categories and scales of the rubric were settled after a professional conference, completing a concrete rubric for every problem through developing the achievement criteria, in order to divide the cases and characteristics by measure with an example answer paper by category. As a result of the

survey for elementary school teachers (Kim, Cho & Joo, 2012), the item difficulty of the constructed-response items indicates that the average level is 45.1%. Hence, the level of the developed questions was adjusted as average~intermediate-advanced. Three items were developed for each unit, and the suitability was examined through an expert review and pilot test.

Nowadays, there are troubles that instructional objectives are unclear and some classroom activities are kept going without any educational purpose. Those problems can be overcome by organizing lessons with a focus on big ideas and essential questions, which are based on the core concepts and principles of discipline; this method will allow students to endure the process of gaining an understanding (Wiggins & McTighe, 2005). NCTM (2011a) addressed the essential understandings that are central in mathematics topics in order to help students with deep understanding overarching concepts to connect big ideas to several little mathematical ideas.

With this perspective, this study developed constructed-response assessment items and criteria that could be solved when students understand the relation between objectives and mathematical ideas, which will ultimately help students to gain a deep mathematical understanding.

### Framework of constructed-response assessment rubric

This study intended to develop constructed-response assessment items corresponding to the aims of the Revised Math Curriculum of Korea in order to enhance mathematical power as well as comprehensively pursue approaching problem-solving instruction related to the teaching contents of every scope. To develop the rubric, in the 2008 research, which derived the frame of the constructed-response assessment with reference to "2003 TIMSS Assessment Framework and Specification", a precedent research report on the development and evaluation of math achievement test items to be used for TIMSS-2003 by Mullis et al. (2003), it set 3 analytic assessment scopes, problem understanding, problem solving process and communication, as the frame of the rubric (Cho et al., 2008).

In the case of U.S.A, for example, the state of Alaska set an analytic performance grade list, including the scopes of concept understanding, problem solving strategy, communication, logic and inference, and then presented it to students after restating them with students' terms to help students understand it. Furthermore, there are cases for each state to use the analytic scoring rubric table in order to include each sub-scope in the rubric evaluating a problem solving process. However, there are also many cases of using a

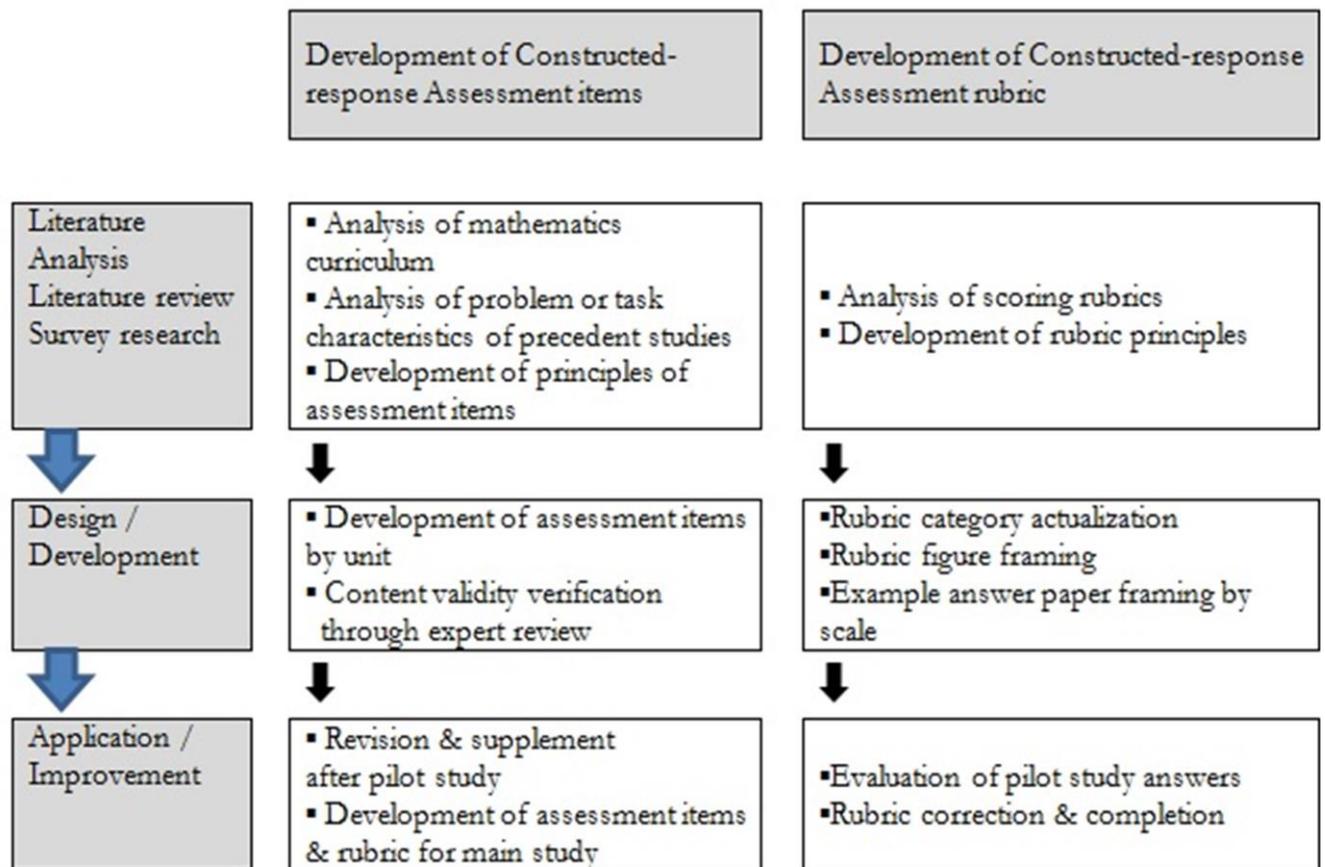


Figure 1. Procedures in the design of constructed-response assessment items & the rubric

holistic scoring rubric table, such as in the states of Maryland, Kentucky, California, North Carolina and Maine (Chicago Public Schools, n.d.).

The analytic scoring rubric is to score the achievement behaviour according to the elements of the rubric and allotting the marks on the assessment rubric table as well as to give marks by element with the aim of understanding and making a diagnosis of students' learning. Whereas it takes too much time to score and train a marker, it has advantages in grasping the student's strength and weakness by element or scope, finding out concrete information for improvement of the program, and giving marks according to the various aspects of performance (Charles, Lester & O'Daffer, 1987; Chi, 2000; Herman, Aschbacher & Winters, 1992).

Reflecting on the situation of school fields more often used a holistic scoring in Korea such as excellent/good/fair/need improvement, this study established the holistic scoring rubric, as presented in Table 1. The holistic scoring rubric is a method in which a marker grades judging from the holistic viewpoint with the scoring unit of whole achievement behaviours. It is possible to score relatively rapidly and give one kind of mark to the entire answer paper while attaching

importance not only to the answer, but also to the solving process.

### Applying constructed-response assessment items to 3rd graders

An example of constructed-response assessment items developed for Grade 3 is shown below. The unit covered the topics of 'volume and weights' in the strand of measurement (Table 2). Strand of measurement is the most ubiquitous topic and connects to other content knowledge such as geometry and numbers, which are main strands in school mathematics. In particular, for the lower level graders, the measurement lessons should be organized to intuitively sense and experience measuring itself, and to make sense of the meaning and attributes of measuring.

In Korea's mathematics curriculum, the strand of measurement deals with utilization of time, length, capacity, weight, angle, area and volume. The items exemplified in this paper are developed for the 'capacity and weight' chapter. Before learning about the capacity and weight, students learn the comparison of amounts, value indication, units of length and calculation in the measurement strand. This chapter allows students to

**Table 1.** Assessment Framework of the Constructed-response Items for Holistic Scoring

Level	Criteria
A (Excellent)	<p>Completely understood mathematical concepts presented in the problem</p> <p>Correctly selected the necessary information from the problem and applied the concepts and information in problem solving</p> <p>Accurately set the strategy and carried it out correctly, and expressed the answer in the problem context</p> <p>Solved the problem through appropriate reasoning</p> <p>Sufficiently explained the process such that there is no need to presume anything in the level of process</p> <p>Accurately represented mathematical concepts and symbols</p>
B (Good)	<p>Completely understood mathematical concepts presented in the problem</p> <p>Correctly selected the necessary information from the problem and applied the concepts and information in problem solving</p> <p>Accurately set the strategy and carried it out correctly, answer is not correct</p> <p>a) mistakes in computation or writing down the answer</p> <p>b) the number in the answer is correct but the unit is wrong or omitted</p> <p>c) no answer</p> <p>Sufficiently explained the process logically, but included a partial gap in the logic used in the process</p> <p>Represented mathematical concepts and symbols with little mistakes</p>
C (Fair)	<p>Partially understood the mathematical concepts in the problem</p> <p>Partially used information presented in the problem with a lack of problem situation</p> <p>Appropriately used strategy, but did not sufficiently execute to get an answer</p> <p>Did not clearly represent mathematical concepts and symbols in problem solving</p> <p>Provided the answer, but incomprehensible problem solving process or no process at all</p>
D (Need Improvement)	<p>Did not really understand what the problem is asking and what mathematical concepts are presented in the problem</p> <p>Misused information that is irrelevant to the problem solving process due to a lack of understanding the problem</p> <p>Difficulty in understanding the problem solving process due to a lack of clear explanations</p> <p>Complete lack of trials</p> <p>a) blank sheet, wrong answer or nothing</p> <p>b) some explanation irrelevant to the problem solving</p>

feel the inconvenience of using a random unit in the process of comparison through examples of daily life; further, it introduces the universal units. Next, by learning how to estimate the weight and how to obtain the sum and difference of weights, students develop a sense of volume for capacity and weight, indicate values using the appropriate unit, and properly perform calculation which is the ultimate goal. Subsequently, students recognize the relationship between mathematics and daily life by learning a variety of measurements, such as width, volume, etc. This study intended to evaluate whether students understand the necessity of the universal unit of weight in the process of comparing weights and the inconvenience of using a random unit. The items are differentiated from the arithmetic operation-centered items that calculate the sum and difference, as primarily suggested in the conventional test items.

The developed problems using the holistic assessment framework were then reviewed by 3

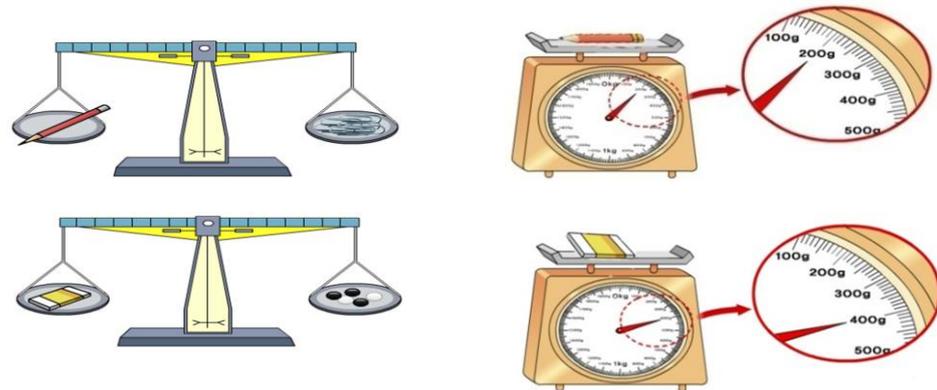
different raters who were elementary teachers or people with a graduate degree in elementary mathematics education. The participants of the present study included 74 students (27 for grade-level 3-1, 47 for grade-level 3-2) from 3 classes in the 3rd grade (Table 3). The rubric for holistic scoring was applied to each item by transforming 4 levels ('Excellent', 'Good', 'Fair' and 'Need Improvement') to 3, 2, 1 and 0 points.

Cronbach reliability coefficient analysis was used to guarantee the internal consistency among the 3 raters in the grade. The coefficients for each rater ranged from .864 to .923, indicating good internal consistency among the raters (Table 4). Pearson correlation coefficient method was used for the constructed-response assessment items from the participants in the 3<sup>rd</sup> grade for inter-rater reliability. The correlation coefficients ranging from .864 to .923 showed good inter-rater reliability with the level of significance at  $p < .01$  (Table 5).

**Table 2.** Example of Constructed-response Assessment Item for Grade 3

Grade level	Unit	Content
3-2	Volume and Weight	Measurement
Key Concepts and Skills	Comparing Weights	
Objectives	To compare weights directly or indirectly, and to explain the comparing method To feel the necessity of the measure(unit) of weight and the inconvenience of random measure(unit) by comparing weight activities	
	There are two ways of comparing weights of a pencil and an eraser.	
	A. Measuring the weight with a paperclip and a stone	B. Putting a pencil or an eraser on a graduated scale

Problem example



Which method should be used to precisely weigh the difference between a pencil and an eraser? Choose a method and explain your choice. Find out how much heavier the weight of an eraser is than the weight of a pencil from the reading on the graduated scale, and explain how you can solve this problem.

**Table 3.** Classes and Number of Students for Study

Grade-level	class	# of students in class	Chapters
3-1	A	27	1~8
3-2	B	21	1~8
	C	26	1~8

**Table 4.** Internal consistency (Cronbach  $\alpha$  Coefficient)

Rater	Grade-Level	Grade-Level
	3-1	3-2
A	.864	.923
B	.871	.919
C	.881	.904

Based on the contents in grade-levels 3-1 and 3-2 of the core curriculum in the 7th revised national school mathematics curriculum of Korea, three constructed-response assessment items were developed for each chapter from 5 content strands, such as number & operations, figures, measurement, probability &

statistics, and pattern & problem solving (Table 6). Developed items were reviewed by mathematics education experts and elementary school teachers. The total number of units was 16, and the total number of items was 48. The results of applying 24 items per semester are summarized in Table 7.

**Table 5.** Inter-rater reliability in each semester of 3<sup>rd</sup> graders-Pearson Coefficient

	Grade-Level: 3-1 (N=573)			Grade-Level: 3-2 (N=1116)		
	rater A	rater B	rater C	rater A	rater B	rater C
rater A	1			1		
rater B	.918**	1		.920**	1	
rater C	.976**	.946**	1	.843**	.858**	1

\*\* p<.01

**Table 6.** Numbers of items from content strands

Content Strands	Grade level: 3-1			Grade level: 3-2		
	# of items	%	Mean (SD)	# of items	%	Mean (SD)
Number & Operations	15	62.5	1.78(1.11)	12	50	1.59(1.14)
Figures	6	25	1.97(1.05)	3	12.5	1.13(1.15)
Measurement	3	12.5	1.85(1.15)	3	12.5	1.26(1.20)
Probability & Statistics	None	.	.	3	12.5	1.79(1.02)
Pattern & Problem Solving	None	.	.	3	12.5	0.67(0.87)
Total	24	100.00	1.83(1.11)	24	100.00	1.41(1.16)

The tests for the constructed-response assessment were given to third-grade students for two semesters. As a result, the scaled score of the Number & Operation area (56.25 percent), which accounts for the largest amount among five content areas of the curriculum, was 1.695 out of a total of 3. The Measurement area (6.25 percent) had the highest score of 1.853. The Pattern area (6.25 percent) had the lowest score of 0.670. Overall, the results indicated that the students had similar levels in four content areas of the curriculum, except for the Pattern area. Because the Pattern area requires more advanced and comprehensive thinking skills compared to the other areas, it can be assumed as to why the Pattern area has the lowest score among the five. Examples of specific rubric and participants' answers from items, as shown in Table 2, are attached in each level in Table 8.

According to the results presented in [Table 9], for the sample items presented in [Table 2], 48.1% of students understand the mathematical concepts implied in the questions, and 47.0% of students need help for conceptual understanding. Therefore, this study intends to suggest a way of helping students' learning depending on the score.

As suggested in the example of [Table 8], students who obtained a B score understand the mathematical concepts implied in the situation of the item because they know the difference between Method A and Method B; however, they cannot explain the strength and weakness of each method sufficiently by connecting the information given in the problem situation. For these students, we can help by creating many

opportunities to explain the relationship between the property of measurement object and unit.

Students who obtained a C score know the difference between Method A and Method B, as it is exemplified, but are confused about the concepts implied in the item or cannot describe the answer adjusting to the situation of the item. Moreover, the students cannot read the given scale correctly. Therefore, for students who obtained a C score, we should help them to understand the relationship between the property of the object and measurement by continuing with learning activities about what instrument and unit can be used depending on the property of measurement.

The example in [Table 8] shows that students who obtained a D score do not recognize the difference between Method B, placing on a graduated scale labelled with the universal unit, and Method A, measuring the weight by stone checkers and clips using a random unit, or do not understand the inconvenience of using a random unit. Therefore, for these students, it is required to offer experiences that help them to understand why a universal unit is necessary in order to communicate the measurement results.

## DISCUSSION AND IMPLICATION

This study intends to examine the policy and actual condition of a constructed-response assessment, performed from the level of the nation, district offices of education and schools, in order to enhance students' creativity and problem solving ability as well as stabilize

**Table 7.** Participants' results of Constructed-response Assessment Items

Grade Level: 3-1				Grade Level: 3-2			
Chapter (Unit)	Strand	Item #	mean(SD)	Chapter (Unit)	Strand	Item #	mean(SD)
1. Number to 10000	Number & Operations	3-1-1-1	1.75(.92)	1. Addition & Subtraction	Number & Operations	3-2-1-1	1.17(1.07)
		3-1-1-2	.75(.88)			3-2-1-2	1.77(1.01)
		3-1-1-3	2.24(.92)			3-2-1-3	1.51(1.29)
2. Addition & Subtraction	Number & Operations	3-1-2-1	2.86(.29)	2. Multiplication	Number & Operations	3-2-2-1	1.23(1.17)
		3-1-2-2	2.58(.84)			3-2-2-2	1.41(.87)
		3-1-2-3	1.33(1.08)			3-2-2-3	1.37(1.00)
3. Plane Figure	Figures	3-1-3-1	2.09(.80)	3. Circle	Figures	3-2-3-1	1.12(.86)
		3-1-3-2	1.50(1.08)			3-2-3-2	.97(1.08)
		3-1-3-3	2.09(.97)			3-2-3-3	1.28(1.36)
4. Division	Number & Operations	3-1-4-1	1.79(.94)	4. Division	Number & Operations	3-2-4-1	2.28(.91)
		3-1-4-2	1.42(.83)			3-2-4-2	1.56(1.27)
		3-1-4-3	1.87(1.18)			3-2-4-3	1.56(1.05)
5. Transformation of Plane Figure	Figures	3-1-5-1	2.09(1.04)	5. Volume & Weight	Measurement	3-2-5-1	1.11(1.06)
		3-1-5-2	2.19(.94)			3-2-5-2	1.15(1.30)
		3-1-5-3	1.91(1.24)			3-2-5-3	1.53(.95)
6. Multiplication	Number & Operations	3-1-6-1	2.51(.93)	6. Decimal	Number & Operations	3-2-6-1	2.12(.84)
		3-1-6-2	2.24(.97)			3-2-6-2	2.19(1.11)
		3-1-6-3	.72(.83)			3-2-6-3	.96(.78)
7. Fraction	Number & Operations	3-1-7-1	1.96(1.05)	7. Data Arrangement	Probability & Statistics	3-2-7-1	1.55(.81)
		3-1-7-2	1.47(.86)			3-2-7-2	1.65(1.13)
		3-1-7-3	1.15(.81)			3-2-7-3	2.18(.70)
8. Length & Time	Measurement	3-1-8-1	2.50(.83)	8. Pattern & Problem Solving	Pattern & Problem Solving	3-2-8-1	1.05(.79)
		3-1-8-2	1.64(1.15)			3-2-8-2	.37(.63)
		3-1-8-3	1.42(1.17)			3-2-8-3	.59(.81)
Average of Total			1.84(.46)	Average of Total			1.41(.62)

basic education. Also, it aims to find out a suggestion for effective application of constructed-response assessment performed on a large scale by continual needs and demands.

First, as a result of analysing the purpose of assessment in the 2007 and 2009 7th revised national school mathematics curriculum of Korea, this study aims to assess by not only focusing on the outcome, but on the process, and therefore uses the assessment results to improve teaching-learning. In addition, as depicted from the findings of Kim, Kwon, Noh, Joo & You (2008) and Noh, Kim, Cho, Jeong & Jeong (2008), the demand for item development has been increasing continually since the beginning of performance assessment. Thus, it is required to develop and spread constructed-response assessment items of various levels and forms, which can be used in the classroom, along with the rubric which can be helpful for the

improvement of teaching-learning; eventually, it will also help in the study of learners through objective scoring.

Second, the quantitative assessment of the past valuing administrative functions to measure the ability of students and rank them relatively is changing to a qualitative assessment, which values instructional functions in order to obtain information on students' learning and improve the teaching-learning process with the aim of increasing the understanding of students, thereby taking on its added significance. The importance of the assessment reveals itself from Bell & Isaacs (2007), saying that the entire thinking process of learners should be evaluated synthetically in elementary math curriculum. Also, the questionnaire of this study also found that students' mathematical thinking power, creativity and problem solving ability may increase by using a constructed-response assessment in the mathematics subject. Therefore, it is critical to operate

assessment, instruction, learning objective and teaching-learning process of mathematics consistently.

As for the mathematics curriculum development and implementation, assessment should be in

accordance with the goal of teaching and learning mathematics. Moreover, the process of teaching and learning mathematics should be consistent. More specifically, the nature of the questions from the

**Table 8.** Example #1 of Rubric and students' answers

Level	Criteria
A (Excellent)	<p>Completely understood mathematical concepts (comparing weights) presented in the problem</p> <p>Correctly selected the necessary information (choosing the precise method between measuring the weight with a paperclip and a stone and using a graduated scale, and explaining the difference of objects by reading the gradation on the scale) from the problem and applied the concepts and information in problem solving</p> <p>Accurately set the strategy and carried it out correctly (choosing the precise method for comparing weights), and expressed the answer in the problem context</p> <p>(a) Method A is not appropriate because a paperclip and a stone used as the comparing unit have different weights.</p> <p>(b) Method B is appropriate because the graduated scale can show the precise weight of the object.</p> <p>Solved the problem through appropriate reasoning</p> <p>Sufficiently explained the process that there is no need to presume anything in the level of process</p> <p>Accurately represented the mathematical concepts and symbols</p> <p>1) 연필과 지우개 무게의 차이를 정확히 알아보고 할 때 어떤 방법을 사용할 것 인지 선택하고 그 까닭을 설명하십시오. -선택한 방법 : 눈금 있는 저울을 사용함 -그 까닭 : 일단 무게를 비교하려면 같은 종류의 것을 비교 해야 한다 하지만 클립과 바둑돌은 무게가 다르기 때문에 정확히 비교할 수 없다.</p> <p>2) 저울의 눈금을 보고 지우개의 무게는 연필 무게의 몇 배인지 찾고, 풀이 과정을 설명하십시오. -풀이과정 : 연필 : 200g, 지우개 : 400g, <math>400 \div 200 = 2</math> -정답 : ( 2 )배</p> <p>1) Method B: Placing on a graduated scale Because we have to use 'the same thing' to compare weights. A paperclip and a stone have different weights to use as the unit, so we couldn't compare the weights precisely.</p> <p>2) A pencil's weight: 200g An eraser's weight: 400g <math>400 \div 200 = 2</math> Answer: 2 times</p>
B (Good)	<p>Completely understood mathematical concepts (comparing weights) presented in the problem</p> <p>Correctly selected the necessary information (choosing the precise method between measuring the weight with a paperclip and a stone and using a graduated scale, and explaining the difference of objects by reading the gradation on the scale) from the problem and applied the concepts and information in problem solving</p> <p>Accurately set the strategy and carried it out correctly, answer is not correct</p> <p>a) mistakes in computation or writing down the answer</p> <p>b) the number in the answer is correct but the unit is wrong or omitted</p> <p>c) no answer</p> <p>Sufficiently explained the process logically, but included a partial gap in the logic used in the process</p> <p>Represented mathematical concepts and symbols with little mistakes</p> <p>1) 연필과 지우개 무게의 차이를 정확히 알아보고 할 때 어떤 방법을 사용할 것 인지 선택하고 그 까닭을 설명하십시오. -선택한 방법 : 눈금 있는 저울을 사용함 -그 까닭 : 영구체와 돌은 정확히 알 수 없는 무게라 하여 '인수' '인수'라는 눈금에 있는 저울을 사용함 이 정답을 알 수 없습니다.</p> <p>2) 저울의 눈금을 보고 지우개의 무게는 연필 무게의 몇 배인지 찾고, 풀이 과정을 설명하십시오. -풀이과정 : <math>200 \times \square = 400</math>, <math>400 \div 2 = 200</math> -정답 : ( 2 )배</p> <p>1) Method B: Placing on a graduated scale While weighing the machine for Method A just lets us know the approximate difference of weights, weighing the machine for Method B lets us know the precise weights.</p> <p>2) <math>200 \times \square = 400</math> <math>400 \div 2 = 200</math> Answer: 2 times</p>
C (Fair)	<p>Partially understood the mathematical concepts in the problem</p> <p>Partially used the information presented in the problem with a lack of problem situation</p> <p>Appropriately used strategy, but did not sufficiently execute the problem in order to get an answer</p> <p>a) Could not read the gradation correctly</p> <p>b) For Method A, incorrect explanation, such as 'the unit of volume', etc.</p>

<p>C (Fair)</p>	<p>Partially understood the mathematical concepts in the problem Partially used the information presented in the problem with a lack of problem situation Appropriately used strategy, but did not sufficiently execute the problem in order to get an answer a) Could not read the gradation correctly b) For Method A, incorrect explanation, such as ‘the unit of volume’, etc. Did not clearly represent the mathematical concepts and symbols in problem solving Provided the answer, but incomprehensible problem solving process or no process at all</p>	<p>1) Method B: Placing on a graduated scale A paperclip and a stone have different volume and weight, so we couldn't know the precise weight and the difference of them. 2) <math>400(\text{an eraser}) \div 2 = 200(\text{a pencil})</math> <math>200(\text{a pencil}) \times 2 = 400(\text{an eraser})</math> Answer: 2 times</p>
<p>D (Need Improvement)</p>	<p>Did not really understand what the problem is asking and what mathematical concepts are in the problem Misused information that is irrelevant to the problem solving process due to a lack of understanding the problem Difficulty in understanding the problem solving process due to a lack of clear explanations Complete lack of trials a) blank sheet, wrong answer or nothing b) some explanation irrelevant to problem solving</p>	<p>1) An electronic scale To find out quickly and precisely 2) A pencil: 4 An eraser: 5 Answer: 1 time</p>

**Table 9.** Frequency of participants' response on items # of response (%)

Level	A (Excellent)	B (Good)	C (Fair)	D (Need Improvement)	value unknown at present	Total
rater A	12(44.4)	1(3.7)	6(22.2)	4(14.8)	4(14.8)	27(100)
rater B	12(44.4)	1(3.7)	6(22.2)	4(14.8)	4(14.8)	27(100)

Note: Frequency of participants' response regarding the example of constructed-response assessment item for Grade 3, as shown in Table 2, is attached the table.

constructed-response assessment conducted in schools should be applied to that of textbooks or other resources utilized in teaching and learning math. For this, more attention is needed to develop assessment items included in all future curriculum resources. Have the results of the assessment given a specific direction to improve how to teach and learn mathematics? Has the constructed-response assessment conducted in accordance with its original purpose in the field? Are

teachers actually committed to improving the linkage between the assessment and the teaching-learning process utilizing the assessment results? Those questions should be raised while conducting an in-depth research. Furthermore, the results of the research should be applied to mathematics education in order to enhance expertise on assessment.

Constructed-response assessment should be used to evaluate a simple concept and principle. In addition, the

relevance between an analytic rubric used as an assessment framework and the holistic scoring method should be more specifically considered. Then, systems that can be complementary to gaps from the two results should be reinforced as well. Furthermore, in-depth analysis of educational significance and effects of the constructed-response assessment will be needed. The constructed-response assessment should be based on open-ended problems, which induce the enhancement of mathematical thinking, creativity and problem-solving skills; further, it is an alternative for in-depth qualitative assessment.

Education policy has sought to expand the constructed-response assessment. However, the policy has not been welcomed due to practical reasons, such as heavy workload of teachers and lack of assessment data. In this regard, it is necessary to meet teachers' continuing needs for the development of questions and evaluation criteria as well as pursue improvement for the quality of the constructed-response assessment. For this, prestigious institutions should take more responsibility for developing and applying questions and assessment criteria in order to enhance their reliability, validity and objectivity. Moreover, individual teacher's commitment is also needed. Appropriate measures and follow-up teaching methods will be able to be put in place through the constructed-response assessment as a formative assessment. Further, information regarding the results of the questions to which students respond, the learners' thinking process, representation and communication skills can be figured out by the formative assessment in order to identify the depth of learners' understanding.

Similar to the results of the research by Kim et al., (2008) and Noh et al., (2008), the problem about and the solution for implementing the constructed-response assessment have turned out to develop and promote items and evaluation criteria for the assessment. Multiple institutions, including Seoul Education Research & Information Institute (<http://www.serii.re.kr/>), have worked on the expansion of the constructed-response assessment using their websites and brochures distributed to schools. The websites and brochures include assessment items regarding the performance assessment and constructed-response assessment, whose evaluation levels vary depending on the grade and subject. However, teachers are still demanding for the development of more items and evaluation criteria. Thus, institutions must be more committed to expanding the constructed-response assessment and suggesting recommendations for implementation in order to make the assessment more reliably established within school education.

## Acknowledgements

This work was supported by the National Research Foundation of Korea Grant funded by the Korean government (NRF-2011-32A-B00216). Parts of this paper were presented at the International Congress on Mathematical Education in Seoul, Korea in July 2012 (Kim, Cho, Hong, Joo, & Lee, 2012).

## REFERENCES

- Baek, S. G. (2000). *Principles of performance assessment*. Seoul: Education and Science Press.
- 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*. Reston, VA: NCTM.
- Black, P., Hariison, C., Lee, C., Marashall, B., & Wiliam, D. (2003). *Assessment for learning: putting it into practice*. Berkshire: Open University Press.
- Brookhart, S. M. (2005). Research on formative classroom assessment. In Paper presented in the symposium, Formative classroom assessment: research, theory, and practice, at the annual meeting of the American Educational Research Association Montreal, Canada.
- Charles, R., Lester, F., & O'Daffer, P. (1987). *How to evaluate progress in problem solving*. Reston, VA: National Council of Teachers of Mathematics.
- Chi, E. L. (2000). Comparing scoring methods for essay type performance assessment. *The Journal of Educational Research*, 16, 235-246.
- Chicago Public Schools (n.d.). Retrieved from Office of Performance website, <http://research.cps.k12.il.us/cps/accountweb/Assessment/Resources/IdeasandRubrics/RubricBank/>
- 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.
- Cho, J. M., Kim, S. J., Lee, S. H., Kim, K. Y., Ok, H. J., Rim, H. M., Park, Y. B., Lee, M. H., Han, H. J., & Son, S. K. (2011). *The trends in international mathematics and science study (TIMSS 2011): A technical report of the main survey in Korea*. Seoul: Korea Institute of Curriculum & Evaluation.
- Diane, H. (1994). *Authentic assessment: a handbook for educators*. Menlo Park, CA: Addison-Wesley Publishing Company.
- 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 and Development.
- Hur, K. C., Baek, S. G., Park, K. M., Choi, M. S., Yang, G. S., & Kim, G. J. (1999). *Status and recommendations for performance assessment policy*. Seoul: Korea Institute of Curriculum & Evaluation.
- Kim, K. H., Si, K. J., Kim, M. Y., Ok, H. J., Rim, H. M., Kim, S. H., Jung, S., Jung, J. Y., & Park, H. J. (2010). *The programme for international students assessment (PISA 2009)*

- results. Seoul: Korea Institute of Curriculum & Evaluation.
- 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. *Journal of Korea Society of Educational Studies in Mathematics*, 10(3), 401-422.
- Kim, K. H., Si, K. J., Kim, M. Y., Ok, H. J., Rim, H. M., Kim, S. H., Jung, S., Jung, J. Y., & Park, H. J. (2010). The programme for international students assessment (PISA 2009) results. Seoul: Korea Institute of Curriculum & Evaluation.
- Kim, M. K., Cho, M. K., Hong, J. Y., Joo, Y. R., & Lee, J. (2012, July). Development and implication of descriptive assessment in elementary mathematics classroom in Korea. Paper presented at the meeting of the 12<sup>th</sup> International Congress on Mathematics Education, Seoul, Korea.
- Kim, M. K., Cho, M. K., & Joo, Y. R. (2012). A survey of perception and status about descriptive assessment. *Journal of Elementary Mathematics Education in Korea*, 16(1), 63-95.
- Kim, M. K., Cho, M. K., Kim, R. Y., Kim, G., & Noh, S. (2012). A comparative analysis of system and problems of descriptive assessment in elementary school mathematics between Korea and U.S.A. *Journal of the Korean School Mathematics Society*, 15(2), 239-258.
- Kim, M. K., & Noh, S. (2010). Alternative mathematics assessment: A case study of the development of descriptive problems for elementary school in Korea. *Eurasia Journal of Mathematics, Science & Technology Education*, 6(3), 173-186.
- Lappan, G., Phillips, E. D., & Fey, J. T. (2007). The case of Connected Mathematics. In C. R. Hirsch (Ed.), *Perspectives on the design and development of school mathematics curricula* (pp.67-80). Reston, VA: NCTM.
- Leahy, S., Lyon, C., Thompson, M., & Wiliam, D. (2005). Classroom assessment: Minute by minute, day by day. *Educational Leadership*, 63(3), 18-24.
- Martin, O. M., Mullis, I.V.S., & Foy, P. (2013). TIMSS 2015 assessment design. In I.V.S. Mullis, & M. O. Martin (Eds.). *TIMSS 2015 assessment frameworks* (pp. 85-99). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Ministry of Education, Science and Technology (2009). *National school curriculum for mathematics*. Seoul: Ministry of Education, Science and Technology of Korea.
- Mullis, I. V. S., & Martin, M. O. (2011). *TIMSS 2011 item writing guidelines*. Chestnut Hill, MA: International Study Center, Lynch School of Education, Boston College.
- Mullis, I.V.S., Martin, M.O., Smith, T.A., Graden, 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: International Study Center, Lynch School of Education, Boston College.
- 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.
- National Council of Teachers of Mathematics (2001). *Mathematics assessment: Cases and discussion questions for grades K-5*. Reston, VA: Author.
- National Council of Teachers of Mathematics (2011a). *Developing essential understanding of addition & subtraction: Pre-K-Grade 2*. Reston, VA: Author.
- National Council of Teachers of Mathematics (2011b). *Using classroom assessment to improve student learning*. Reston, VA: Author.
- Noh, S., Kim, M. K., Cho, S. M., Jeong, Y. S., & Jeong, Y. A. (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.
- Park, G. R. & Pang, J. S. (2008). A survey of the comprehension of basic knowledge of mathematics of 6th graders in elementary school by essay test. *The Mathematical Education*, 47(2), 181-195.
- 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 (2010). *Assessment change plan for the development of Creativity*. Retrieved from <http://www.sen.go.kr>
- Seoul Metropolitan Office of Education (2011). *The major plan for 2011*. Retrieved from <http://www.sen.go.kr>.
- Slavin, R. E. (2009). Assessing student learning. In R. E. Slavin, *Educational psychology: theory and practice* (9th edition) (pp. 404-453). Upper Saddle River, NJ: Pearson Education, Inc.
- Wiggins, G., & McTighe, J. (2005). *Understanding by design* (2nd edition). Alexandria, VA: Association for Supervision and Curriculum Development.
- Wiliam, D. (2007). Keeping learning on track: classroom assessment and the regulation of learning. In F. K. Lester, *Second handbook of research on mathematics teaching and learning* (pp. 1053-1098). Reston, VA: National Council of Teachers of Mathematics.

