DOCUMENT RESUME

ED 435 625 TM 030 223

AUTHOR	Suzuki, Kyoko
TITLE	Cognitive Constructs Measured in Word Problems: A Comparison of Students' Responses in Performance-Based Tasks and Multiple-Choice Tasks for Reasoning.
PUB DATE	1997-03-00
NOTE	23p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, March 24-28, 1997).
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Algebra; Cognitive Processes; *High School Students; High Schools; Holistic Evaluation; Multiple Choice Tests; Performance Based Assessment; Responses; *Scoring; *Word Problems (Mathematics)
IDENTIFIERS	Scoring Rubrics

ABSTRACT

This study investigated cognitive constructs to be measured by word problems in algebra. One performance-based assessment was administered to 290 high school students. Students' responses were scored by three scoring systems: the correct/incorrect criterion (0/1); a holistic scoring rubric (0-4); and an analytical scoring rubric for measuring maturity levels of mathematical reasoning (0-4). The results demonstrate that cognitive constructs in a performance-based format were different from those in a multiple-choice format. Outcome cognitive constructs in students' responses were not the same as the planned ones in tasks. Bloom's Taxonomy was not sufficient for classifying mathematical reasoning. The correct/incorrect scores could not distinguish different levels of mathematical reasoning. The combined use of a holistic scoring rubric and an analytical one for reasoning were informative. (Contains 1 table, 8 figures, and 13 references.) (Author/SLD)

Running head: A COMPARISON OF COGNITIVE CONSTRUCTS

Cognitive Constructs Measured in Word Problems: A Comparison of Students' Responses in Performance-based Tasks and Multiple-choice Tasks for Reasoning

Kyoko Suzuki

Department of Educational Psychology

University of Illinois at Urbana-Champaign

230 Education Building

1310 South Sixth Street

Champaign, IL 61820

phone: (217) 333 - 2245

fax: (217) 244 - 7620

email : k-suzuki@uiuc.edu

Paper presented at the 1997 annual meeting of the American Educational Research Association, March, 24 - 28, Chicago, IL.

Abstract

This study investigates cognitive constructs to be measured in word problems in algebra. One performance-based assessment was administered to 290 high school students. Students' responses were scored by three scoring systems: the correct/incorrect criterion (0/1), a holistic scoring rubric (0-4), and an analytical scoring rubric for measuring maturity levels of mathematical reasoning (0-4). The results demonstrated that cognitive constructs in a performance-based format were different from those in a multiple-choice format. Outcome cognitive constructs in students' responses were not the same as planned ones in tasks. Bloom's Taxonomy was not sufficient for classifying mathematical reasoning. The correct/incorrect scores could not distinguish different levels of mathematical reasoning . The combined use of a holistic scoring rubric and an analytical one for reasoning were informative.

Cognitive Constructs Measured in Word Problems: A Comparison of Students' Responses in Performance-based Tasks and Multiple-choice Tasks for Reasoning

Recent studies of cognitive psychology have suggested the need for change in achievement testing. How can we assess students' thinking processes and reasoning? How can we infer the levels of students' understanding? What cognitive constructs are measured in different task formats using different scoring criteria? This study investigates cognitive constructs to be measured in solving word problems in algebra.

Because conventional achievement tests are based on the psychological theory of behaviorism, they assess students' observable behaviors that can be reliably recorded as either present or absent (Bloom, Hastings, and Madaus, 1971). However, recent research in cognitive psychology has a changed view of learning. The differences between a novice and an expert are not the amount of knowledge, but the ways of viewing aspects and of structuring problems. "Learning should be a qualitative change in a person's conception of a certain phenomenon or of a certain aspect of reality (Johansson., Marton, & Svensson., 1985)." Therefore, the purpose of assessment is not to establish the presence or absence of specific behaviors, but to infer the nature of students' understandings of particular circumstances (Masters and Mislevy, 1993; Mislevy, 1995).

The National Council of Teachers of Mathematics (NCTM) has stressed fostering problem solving, reasoning, and communication in mathematics education. Assessment should seek evidence of reasoning processes in solving problems. Communication is the vehicle by which students appreciate mathematics as involving the processes of problem-solving and reasoning (NCTM, 1991, p. 96). The format of tasks used in assessment is an important factor affecting students' performances. Open-ended questions are more language-dependent than are multiple-choice questions, both in the statement of the problem and in students' responses; however, open-ended questions can provide for greater diversity in solution strategies and students' understanding. Open-ended questions can offer more insight into students' thought than multiple-choice tests (NCTM, 1995. pp. 58 - 59). The important question is, therefore, how we can assess reasoning process involved in solving problems in paper-pencil type, large-scale testing.

Some studies showed a high correlation between students' performances on multiple-choice and performance-based tasks (e.g. Wolf, 1994). When we consider performance-based tasks, however, there are two methods for scoring responses: (a) scoring final answers using the correct/incorrect criterion and (b) scoring reasoning processes using scoring rubrics. The results would be different based on what scoring system is used. Wolf has pointed out that the more items are structured in terms of the task to be performed and the specification of acceptable responses, the easier such items are to administer and score. On the other hand, this task structure and scoring system rely on observable behaviors rather than inferences of understanding. Open-ended questions or less-structured tasks that have multiple solution strategies and/or multiple answers can offer rich insights into students' thinking processes. Nevertheless, the scoring criteria may be difficult to determine. Furthermore, the scoring system that is used in an assessment greatly affect the results that are obtained.

This study investigates cognitive constructs in tasks (planned constructs) and in students' responses (outcome constructs). These cognitive constructs are compared in different scoring systems. Since performance-based, open-ended tasks can reveal a wide variety of insights of thinking processes in students' written responses, the scoring system should reflect the variety of cognitive constructs in thinking processes. Scoring appropriateness is a key issue for measuring reasoning processes. First, the task structure in multiple-choice format and in performance-based format is examined by both mathematically and cognitively. Next, students' responses are classified by Bloom's Taxonomy. Finally, cognitive constructs of students' responses are compared using different scoring criteria.

Three questions are discussed in this paper:

(1) What cognitive constructs are we measuring in solving word problems?

(2) Are the same cognitive constructs measured in a multiple-choice format and a performance-based format?

(3) How can we score different cognitive levels of reasoning processes?

Method

Items

Eight items in algebra were chosen from similar content areas and were grouped into 4 forms having 5 items each. The reasons for using four different types of forms were: (1) feasibility in a classroom hour (approximately 45 minutes), and (2) detecting which items mostly likely have higher generalizability. The time constraints must be considered for administering a performance-based, achievement test in a classroom. For minimizing the effect of speededness, five items were chosen for each form. Five of the eight items were modified from publicly released SAT multiple-choice type tasks. One task was administered with a slight change in presentations and conditions in different sites to examine cognitive constructs in students' responses measured in different task structure. For full description of the items, please refer to Suzuki & Harnisch, 1996a.

Scoring Rubric

The scoring rubric in this study was adopted from the QUASAR[1](Quantitative Understanding: Amplifying Student Achievement and Reasoning) project. The scale ranges from 0 to 4, and a single score was assigned to a response with holistic perspectives considering three components; mathematical conceptual and procedural knowledge, strategic knowledge, and communication. The description of this rubric was shown in Appendix (Illinois State Board of Education, 1995, Lane 1993).

The MARS (Maturity of Algebraic Reasoning and Strategies) scale was developed by the author for this study and utilized for the measuring maturity levels of mathematical reasoning. The scale which ranges from 0 to 4 intends to measure mathematical achievement levels of solution strategies in problem-solving. The score level was determined for each problem by classifying students' solutions. Detailed construction of the MARS scale was described in the paper of Suzuki and Harnisch (1996 b). The cognitive constructs at each score level are to be discussed compared to Bloom's Taxonomy.

[1] QUASAR (Quantitative Understanding: Amplifying Student Achievement and Reasoning) is a national project that seeks instructional programs in the middle-school grades that promote the acquisition of thinking and reasoning skills in mathematics (Silver, 1991). The project is directed at students attending schools in economically disadvantaged communities.

Samples

The four forms were randomly assigned in one class period (approximately 45 minutes) to 142 Algebra II students in two high schools in Midwestern cities in the U.S. and 148 eleventh graders in one high school in a suburb in Japan. Data were collected during the period of November, 1994 to January, 1996.

Scoring Students' Responses

Students' responses were scored by trained raters using the QUASAR holistic scoring rubric previously described. The inter-rater reliability exceeded .9 in this study. The main reason for this high rate was that the raters were working together for two years in the same project using the same scoring rubric. The significant feature of the scoring procedure was assessing the reasoning and communication skills for finding their answer, rather than the final answer itself. Emphasis was placed on the processes of finding answer and to communicating solution strategies with others in written format. Therefore, a response could receive a "4" (the highest score) if the strategy and process were correctly specified, even though the final answer was not sufficient or even was incorrect. On the other hand, a response could be scored a "2" when a solution process was not provided or was poorly specified, although the final answer was correct.

Classification of Cognitive Constructs in Responses

Wilson's "Table of Specifications for Secondary School Mathematics" (WTS) was used to determine cognitive constructs for tasks and students' responses. WTS was developed to classify tasks in secondary school mathematics using the Bloom's

Taxonomy (Wilson, 1971). The mathematical achievement levels are measured by WTS with two dimensions: categories of mathematical contents and levels of cognitive behaviors.

The content area includes number systems, algebra, and geometry. Number systems includes: (1.1) whole number, (1.2) integers, (1.3) rational numbers, (1.4) real numbers, (1.5) complex numbers, (1.6) finite number systems, (1.7) matrices and determinants, (1.8) probability, and (1.9) numeration systems. Algebra includes: (2.1) algebraic expressions, (2.2) algebraic sentences and their solutions, and (2.3) relations and functions. Geometry includes: (3.1) measurement, (3.2) geometric phenomena, (3.3) formal reasoning, and (3.4) coordinate systems and graphs.

Cognitive behaviors have four levels: computation, comprehension, application, and analysis. The computation-level behaviors include: (A.1) knowledge of specific facts, (A.2) knowledge of terminology, (A.3) ability to carry out algorithms. The comprehension-level behaviors contain six sub-categories: (B.1) knowledge of concepts, (B.2) Knowledge of principles, rules, and generalizations, (B.3) knowledge of mathematical structure, (B.4) ability to transform problem elements from one mode to another, (B.5) ability to follow a line of reasoning, and (B.6) ability to read and interpret a problem. The application-level behaviors involve a sequence of responses by a student: (C.1) ability to solve routine problems, (C.2) ability to make comparisons, (C.3) ability to analyze data), and (C.4) ability to recognize patterns, isomorphisms, and symmetries. The analysis-level behaviors are so called "doing mathematics" level, that is, those where we ask a student to go beyond what he/she has done during previous instruction. Five sub-categories are involved at this level: (D.1) ability to solve nonroutine problems, (D.2) ability to discover relationships, (D.3) ability to construct proofs, (D.4) ability to criticize proofs, and (D.5) ability to formulate and validate generalizations.

Results and Discussion

Task Structure

The task structure of a single item (Item 1) and its variations (Item 2 and Item 3) were discussed to illustrate how cognitive constructs measured in the task changed as some conditions were changed. Item 1, shown in Figure 1, was from publicly released SAT multiple-choice tasks. The item was classified [2] based on the WTS as (C.1) the application level, to solve routine problems and (1.1) whole number in number systems.

Item 2, shown in Figure 2, was a performance-based task used in this test. The use of horizontal sums instead of vertical sums changed the problem structure , although the representations of both sums were equivalent mathematically. In vertical sums, the notation of "AB" may be interpreted as "10A + B" because of A and B being odd digits. Once horizontal sums are used under the same conditions, the notation of "XY" can be interpreted by either "10X + Y" or "X multiplied by Y." For example, when X = 3 and Y = 5 (X and Y are odd digits), XY = (X)(Y) = 15 (XY is a two digit number). In this case, the condition of "odd digit" is interpreted as "odd single positive integer." The horizontal sums are interpreted as "3XY = YZ," which becomes an algebraic problem, no longer a number sense problem. Cognitively, vertical sums presentation emphasizes the condition

[2] This classification is based on an expert opinion (ref. Dr. Kenneth Travers, University of Illinois at Urbana-Champaign).

of "a digit" for X, Y, and Z, to read XY = 10X + Y. Horizontal presentation reduces one's attention to the condition; therefore, the word "digit" is more likely interpreted as "a single positive integer" and hence XY is more likely interpreted as a multiplication of X and Y. Because there are two ways to interpret XY, the problem has multiple strategies and multiple answers.

This task (Item 2) was classified in Analysis level, because it was no longer a routine problem. Furthermore, a performance-based format almost always requires a description of reasoning; therefore, the task was classified based on the WTS as (D.3) ability to construct proofs in the analysis-level behavior and (2.3) relations and functions in algebra. The content area of this task shifted to algebraic relations from number systems. The task became less structured because multiple answers were possible. However, the difference between Item 1 and Item 2 involved only mathematically equivalent representations. What does this imply? The following paragraph discusses an interpretation about this phenomenon.

The original SAT item contains a "hidden" assumption, directed by the school curriculum, which is not explicitly stated. In other words, the item is measuring abilities different from mathematical understanding. That is "school curriculum convention." If students do not share the same curriculum convention, the task is a biased item when it is used in a multiple-choice format. This is because a student may interpret the vertical sums as 3AB = BC. A student could interpret like this if the task is given right after he/she learned algebraic expressions. This interpretation could occur based on a student's background, but not by the mathematical understanding. From this perspective, the performance-based format has an advantage over the multiple-choice format because it can reveal how a student interprets the problem.

Item 3, shown in Figure 3, used the same horizontal sums presentation as Item 2, but the condition of X, Y, and Z was changed from "different odd digits" to "different odd integers." The variables were no longer required to be one-digit integers . Moreover, it was clearly stated that XY and YZ were two-digit integers, instead of "correctly worked sum of three two-digit numbers." This task reduced some of the conditions possessed by Item 2 . Thus Item 3 was less structured than Item 2. Therefore, the range of correct answers was increased [3]. Although some of the mathematical assumptions were changed, the task classification based on the WTS stayed at the same level as Item 2: (D.3) ability to construct proofs at the analysis level and (2.3) relations and functions in algebra.

When a task format is changed, the task structure is also changed. Generally speaking, all performance-based items require students to justify solution processes and reasoning in their own words. So items are usually classified in (D.3) level based on the WTS. When a task is a routine problem or is familiar to students, it is possible to classify as (C.1) ability to solve routine problems at the application-level. In either way, there are only two possible classifications, (D.3) or (C.1). The WTS may not be sufficient to describe cognitive constructs for performance-based tasks.

[3] Mathematically, we can determine a set of correct answers which satisfies sufficient and necessary conditions. However, for this purpose of assessment, we do not expect students to determine the perfect answer. Therefore, we accept a sub set of the perfect answer.

Analyses of Students' Responses in Performance-based Tasks

Students' responses were classified according to the WTS to determine cognitive constructs in problem-solving. These constructs (outcome constructs) were compared to those for tasks (planned constructs) if they were matched. The WTS was originally developed for task classification in secondary school mathematics. Although students' backgrounds and experiences affect the way the students solve a task, the classification of the task is determined by the "average" or most students' experiences on the task for a particular grade level. Here, the description of cognitive processes in the table was used to classify students' products.

Students' responses for Item 2 were shown in Figures 4 through 6. Student 1 rewrote the sum notation into vertical from horizontal format as shown in Figure 4. Cognitively, vertical sums presentation may help students to find the answer intuitively. We could infer that Student 1 justified the answer to be correct after finding it, because no evidence was provided about how to find X = 1, and/or why X needs to be 1.

Student 2 (see Figure 5) justified why X needs to be 1. This student demonstrated a higher level of mathematical reasoning than Student 1. Student 3 (see Figure 5) found the correct answer intuitively, and the reasoning was not mathematical. The difference between Student 1 and Student 3 was that Student 1 could justify the intuitive solution mathematically but Student 3 could not. Therefore, the responses of Student 1 and Student 2 were classified based on the WTS as (D.3) ability to construct proofs in the analysis-level, which was the same as the level of the task classification.

The response of Student 3 was classified as (C.1) ability to solve routine problems at the application-level, which was the same as the level of Item 1, the multiple-choice format. The response of Student 3 could even be classified as either (B.3) knowledge of mathematical structure at the comprehension-level or (B.6) ability to read and interpret a problem at the content category of (1.1) whole numbers in the number systems. It may be expected that students use this intuitive solution for Item 1, a multiple-choice format, because the original SAT item is one of 25 questions for a 30-minute test. On the other hand, the performance-based task is one of five questions for a 45-minute test. Based on this time allocation for solving the problem, the multiple-choice format requires a more intuitive solution rather than mathematical reasoning ability.

Item 3 was the least structured task among the three, and the task promoted a variety of reasoning among students as shown in Figure 7 (Student 4) and Figure 8 (Student 5). Although neither of them showed perfect mathematical reasoning, we could infer the achievement levels of mathematical reasoning from their responses. Student 4 could understand that X needs to be 1 or 3, but did not provide any reason of why X needs to be 1 or 3. Student 5 demonstrated a deeper understanding than Student 4, although an insufficient reasoning process was involved. The response of Student 4 was classified based on the WTS as (D.1) ability to solve nonroutine problems because verbal justification was not provided. The response of Student 5 was classified as (D.3) ability to construct proofs in the analysis-level. It should be noted that the outcome cognitive constructs that were determined in students' responses were not the same as the planned cognitive constructs which were classified for tasks.

Cognitive Constructs Represented by Scoring Systems

Cognitive constructs were compared by different scoring systems to demonstrate how scores assigned to a response could represent cognitive constructs measured in a task. The correct/incorrect scores, the QUASAR holistic scores, and the MARS scores were compared.

All five students (Students 1 through 5) received credit based on either a multiple-choice format or the correct/incorrect scoring criterion. Therefore, we concluded that the correct/incorrect criterion measured the same performance task ability as the multiple-choice task. However, the cognitive constructs measured in the multiple-choice format were different from those in the performance-based format. As we discussed previously, the multiple-choice format required a more intuitive solution rather than reasoning because of time allocation. Consequently, the correct/incorrect scoring criterion may not be appropriate for performance-based tasks. Moreover, the criterion did not distinguish differences among outcome cognitive constructs in students' responses.

When utilizing the QUASAR holistic scoring rubric, Student 1 and Student 2 scored a "4," whereas Student 3 scored a "2" because the explanation was not mathematically justified. Student 4 scored a "3" because of a poor verbal communication. Student 5 scored a "4," although the response involved insufficient reasoning and an incorrect answer. The QUASAR holistic scores could represent outcome cognitive constructs shown in students' responses. The score, however, did not distinguish different achievement levels of mathematical reasoning demonstrated by Student 1 and Student 2.

The MARS scale was developed by the author to classify levels of students' reasoning. The scores involve 5 levels ranged from 0 to 4: 0 for no understanding or no response, 1 for limited understanding with major conceptual errors, 2 for intuitive solutions without mathematical reasoning, 3 for mathematical reasoning which concerns sufficient conditions only (for example, no consideration of why X needs to be 1), and 4 for mathematical reasoning which considers sufficient and necessary conditions. Both a "3" and a "4" level of the MARS scale in mathematical reasoning represent "the evaluation stage" in terms of Bloom's Taxonomy. However, the "4" level represents a higher ability of mathematical reasoning than the "3" level. Classification of cognitive constructs according to Bloom's Taxonomy may not be sufficient to distinguish mathematical reasoning.

Based on the MARS scale, Student 2 and Student 5 received a "4," Student 1 and Student 4 received a "3," and Student 3 received a "2." These scores could distinguish the difference between Student 1 and Student 2, whereas the QUASAR scores could not. In order to assess mathematical reasoning in performance-based tasks, levels of mathematical reasoning need to be represented by an assigned score.

Conclusions and Implications

Cognitive constructs measured in an item change as test formats and task structure are changed. A performance-based format can measure different cognitive constructs from a multiple-choice format. However, when final answers of performance-based tasks are the target of scoring using the correct/incorrect criterion, the scores are the same as a multiple-choice format. Although the scoring is easy and stable, performance-based assessments with such a scoring criterion do not have any advantage over multiple-choice tests because they seek the same evidence.

Performance-based tasks can reveal varieties of mathematical reasoning which cannot be identified in multiple-choice tasks. Because planned cognitive constructs for tasks are not always the same as outcome cognitive constructs in students' responses, performance-based tasks have some advantage in assessing students' cognitive stages over multiple-choice tasks. In addition, we should rethink the fact that a multiple-choice format is well-structured. We may measure a use of "intuitive" solution rather than mathematical reasoning in a multiple-choice format. We may even measure something else such as "curriculum convention" by a well-structured multiple-choice item rather than measuring mathematical ability. Reasoning processes in students' responses in performance-based tasks can clarify the ways students are thinking. It might be a good chance to reconsider the distinction of well-structured items and less-structured or ill-structured items.

The WTS based on Bloom's Taxonomy may not be sufficient for classifying performance-based tasks for mathematical reasoning. Even Bloom's Taxonomy does not distinguish achievement levels of mathematical reasoning: a level of reasoning which considers both sufficient and necessary conditions and a level of reasoning which considers only sufficient conditions. How can we describe these different abilities in mathematical reasoning psychologically?

Scoring systems and criteria are key issues to assess a variability of reasoning processes. A variety of reasoning processes are revealed by less-structured tasks with multiple strategies and/or multiple answers. The QUASAR holistic scores can measure some cognitive constructs; however, they do not distinguish the different ability levels in reasoning. The MARS scale is designed to measure the maturity levels of reasoning and solution strategies. The combined use of both scales may be informative to assess mathematical reasoning in problem-solving.

Wilson stated that there is no evidence to support the assumption that performances at one cognitive level require the mastery of related content at lower levels. Accordingly, performances at all cognitive levels should be expected for all students (Wilson, 1971. p650). Performance-based assessments for mathematical reasoning can be a powerful tool to improve instructions and students' reasoning ability.

Fostering reasoning and communication skills in mathematics education is not an easy process for either students or teachers. Many students tend to believe that finding a correct answer is the goal of solving math problems. Assessments for reasoning and communication could assist students in correcting their misconception of math learning.

References

Bloom, B. S., Hastings, J. T., & Madaus, G. F. (1971). Handbook on formative and summative evaluation of student learning. New York: McGraw-Hill.

Illinois State Board of Education (1995). Performance assessment in mathematics: Approaches to open-ended problems. IL: Author.

Johansson, B., Marton, F., & Svensson, L. (1985). An approach to describing learning as change between qualitatively different conceptions. In L. H. West & L. A. Pines (Eds.), Cognitive structure and conceptual change (pp. 233-257). Orlando, FL: Academic Press.

Lane, S. (1993). The conceptual framework for the development of a mathematics performance assessment instruments. Educational Measurement: Issues and Practice, 12, 16 - 23.

Masters, G., N., & Mislevy, R., J. (1993). New views of student learning: Implications for educational measurement. In N. Frederiksen, R. J. Mislevy, & I. I. Bejar (Eds.), Test Theory for a new generation of tests (pp. 219-242). Hillsdale, NJ: LEA

Mislevy, R. J. (1995). Evidence and inference in educational assessment (Research Rep. No. RR-95-8-ONR). Princeton, NJ: Educational Testing Service.

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

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

Silver, E. A. (1991). Quantitative Understanding: Amplifying Student Achievement and Reasoning. Pittsburgh, PA: Learning Research and Development Center.

Suzuki, K., & Harnisch, D., L. (1996a, April). An investigation on the generalizability of performance-based assessment in mathematics. Paper presented at the annual meeting of AERA, New York, NY.

Suzuki, K., & Harnisch, D., L. (1996b, April). Measuring levels of mathematical thinking in algebra: Do test formats affect students' thinking? Paper presented at the annual meeting of AERA, New York, NY.

Wilson, J. W. (1971). Evaluation of learning in secondary school mathematics. In B. S. Bloom, J. T. Hastings, & G. F Madaus (Eds.), Handbook on formative and summative evaluation of student learning (pp. 643 - 696). New York: McGraw-Hill

Wolf, R. M. (1994). Performance assessment in IEA studies. International Journal of Educational Research, 21(3), 239 - 245.

Figure 1. Item 1: Multiple-choice Format

$$\begin{array}{r} AB \\ AB \\ +\ AB \\ \hline BC \end{array}$$

If A, B, and C are different odd digits in the correctly worked sum of three two-digit numbers shown above, what is the value of B?

(A) 9 (D) 3
(B) 7 (E) 1
(C) 5

WTS:(1.1) whole number/ number sense

(C.1) solve a routine problem/ application

Figure 2. Item 2: Performance-based Format

If X, Y, and Z are different odd digits in the correctly worked sum of three two-digit numbers shown below, find the value of Y.

XY + XY + XY = YZ

Show all your work and explain in words how you found your answer.

WTS: (2.3) relations/algebra

(D.3) construct a proof/ analysis

Figure 3. Item3: Less-structured Task

Assume that X, Y, and Z are different odd integers, and XY and YZ are two-digit integers. When X, Y, and Z have the relation shown below, find the value of Y.

XY + XY + XY = YZ

Show all your work and explain in words how you found your answer.

WTS: (2.3) relations/algebra

(D.3) construct a proof/ analysis

Figure 4. Student 1

$$\begin{array}{r} XY \\ +\ XY \\ +\ XY \\ \hline YZ \end{array} \qquad \begin{array}{r} 13 \\ +\ 13 \\ +\ 13 \\ \hline 39 \end{array}$$

If X = 1, then X + X + X = 3, which could be used for the value of Y. Y + Y + Y which is 9 could then be 9. All three numbers are odd and therefore these three digits are the solution.

X = 1, Y = 3, Z = 9

WTS: (1.1) whole number/ number sense
(D.3) construct a proof/ analysis

C/I : C QUASAR: 4 MARS: 3

Figure 5. Student 2

XY + XY + XY = YZ
13 + 13 + 13 = 39
Y = 3

Know X can only be a 1 or a 3, because any other X value tripled would provide a 3-digit answer. Use one as X, because 1 + 1 + 1 = 3. 3 + 3 + 3 = 9, which is also an odd #. Y has to be the same for Y on both sides of the equation, so 13 + 13 + 13 = 39 is the only choice by guess & check method under the criteria.

WTS: (1.1) whole number/ number sense
(D.3) construct a proof/ analysis

C/I : C QUASAR: 4 MARS: 4

Figure 6. Student 3

13 + 13 + 13 = 39, Y = 3

I guessed what the numbers would be and then worked it out on my calculator.

WTS: (1.1)whole number/ number sense

(C.1)solve a routine problem/ application

* (B.3)knowledge of mathematical structure
(B.6) ability to read and interpret a problem

C/I : C QUASAR: 2 MARS: 2

Figure 7. Student 4

XY + XY + XY = YZ

3XY = YZ

Y(3X - Z)= 0

(i) When X = 1 and Z = 3, Y = 5, 7, 9

(ii) When X = 3,and Z = 9, Y = 1, 5, 7.

WTS: (2.3) relations/ algebra
(D.1) solve a nonroutine problem/ application

C/I : C QUASAR: 3 MARS: 3

Figure 8. Student 5

X, Y, Z: odd integers, XY, YZ: 2-digit integers,

3XY = YZ, XY ≥ 11, YZ ≥ 33.

Then, XY and YZ have the following ranges.

33 ≥ XY ≥ 11, 99 ≥ YZ ≥ 33.

X	11	11	5	3	3	3
Y	1	3	5	7	9	11
Z=3X	33	33	15	9	9	9

The answers are: 1, 3, 5, 7, 9, 11

WTS: (2.3) relations/ algebra
(D.3) construct a proof/ analysis

C/I : C QUASAR: 4 MARS: 4

Appendix The QUASAR holistic scoring rubrics (adapted from Illinois State Board of Education, 1995)

MATHEMATICS SCORING RUBRIC: A GUIDE TO SCORING OPEN-ENDED ITEMS

SCORE LEVEL	MATHEMATICAL KNOWLEDGE Knowledge of mathematical principles and concepts which result in a correct solution to a problem	STRATEGIC KNOWLEDGE Identification of important elements of the problem and the use of models, diagrams and symbols to systematically represent and integrate concepts	COMMUNICATION Written explanation and rationale for the solution process
4	* shows complete understanding of the problem's mathematical concepts & principles * uses appropriate mathematical terminology & notation (e.g: labels answer as appropriate)[1] * executes algorithms completely and correctly	* identifies all the important elements of the problem and shows complete understanding of the relationships among elements * reflects an appropriate and systematic strategy for solving the problem * gives clear evidence of a complete and systematic solution process	* gives a complete written explanation of the solution process employed; explanation addresses what was done, and <u>why</u> it was done * if a diagram is appropriate, there is a complete explanation of all the elements in the diagram
3	* shows nearly complete understanding of the problem's mathematical concepts and principles * uses nearly correct mathematical terminology and notations * executes algorithms completely; computations are generally correct but may contain minor errors	* identifies most of the important elements of the problem and shows general understanding of the relationships among them * reflects an appropriate strategy for solving the problem * solution process is nearly complete	* gives a nearly complete written explanation of the solution process employed; may contain some minor gaps * may include a diagram with most of the elements explained
2	* shows some understanding of the problem's mathematical concepts and principles * may contain major computational errors	* identifies some important elements of the problem but shows only limited understanding of the relationships among them * appears to reflect an appropriate strategy but application of strategy is unclear * gives some evidence of a solution process	* gives some explanation of the solution process employed, but communication is vague or difficult to interpret * may include a diagram with some of the elements explained
1	* shows limited to no understanding of the problem's mathematical concepts and principles * may misuse or fail to use mathematical terms * may contain major computational errors	* fails to identify important elements or places too much emphasis on unimportant elements * may reflect an inappropriate strategy for solving the problem * gives minimal evidence of a solution process; process may be difficult to identify * may attempt to use irrelevant outside information	* provides minimal explanation of solution process; may fail to explain or may omit significant parts of the problem * explanation does not match presented solution process * may include minimal discussion of elements in diagram; explanation of significant elements is unclear
0	* no answer attempted	* no apparent strategy	* no written explanation of the solution process is provided

BEST COPY AVAILABLE

[1] "As appropriate" or "if appropriate" relate to whether or not the specific element is called for in the stem of the item.

Adapted from Lane (1993)

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC®

TM030223

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Cognitive Constructs Measured in Word Problems: A Comparison of Students' Responses in Performance-based Tasks and Multiple-choice Tasks for Reasoning	
Author(s): Kyoko Suzuki	
Corporate Source: University of Illinois at Urbana-Champaign	Publication Date: 3/26/97

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
Level 1	Level 2A	Level 2B
☑	☐	☐
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproductio n by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here,→ please

Signature: Kyoko Suzuki	Printed Name/Position/Title: Kyoko Suzuki /postdoctoral fellow	
Organization/Address: ETS/ MS02-T, Educational Testing Service Princeton, NJ 08541	Telephone: 609-734-5079	FAX: 609-734-5420
	E-Mail Address: ksuzuki@ets.org	Date: 9/18/99

(over)

BEST COPY AVAILABLE

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

The Catholic University of America
ERIC Clearinghouse on Assessment and Evaluation
210 O'Boyle Hall
Washington, DC 20064
Attn: Acquisitions

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
1100 West Street, 2nd Floor
Laurel, Maryland 20707-3598

Telephone: 301-497-4080
Toll Free: 800-799-3742
FAX: 301-953-0263
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com

(Rev. 9/97)