DOCUMENT RESUME

ED 222 559 TM 820 714

AUTHOR Romberg, Thomas A.; And Others
TITLE The Development of Mathematical Problem-Solving Superitems. A Report on the NIE/ECS Item-Development Project.
INSTITUTION Education Commission of the States, Denver, Colo. National Assessment of Educational Progress.; Wisconsin Center for Education Research, Madison.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
PUB DATE Jan 82
NOTE 98p.

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS Elementary Secondary Education; Evaluation Methods; *Mathematics Achievement; Models; *Problem Solving; *Quantitative Tests; *Test Construction; Test Interpretation; *Test Items
IDENTIFIERS National Assessment of Educational Progress; *NIE ECS NAEP Item Development Project

ABSTRACT

The purpose of this report is to describe the development of a pool of mathematical problem-solving situations and a set of items for each situation which provides information about students' qualitatively different levels of reasoning ability as applied to that situation. For each problem-solving situation, a set of "structured superitems" was developed. "Superitems" are a set of test items about a common situation or stem. Such item sets have been shown to be valid and useful for assessing mathematical problem solving. For the constructed item pool, a recently developed taxonomy based on the structure of the observed learning outcome (SOLO) was used as the basis for the development of the superitems. Four questions were written for each problem situation to assess five levels of reasoning concerning the situation. The superitems were administered in separate group tests to 300 students of 9, 11, 13, and 17 years of age. The 38 items in the final batteries were mathematically correct and were related to the levels of reasoning in the SOLO Taxonomy. This document focuses on the details of the preparation of the items prior to their administration. Instructions for judging items, manuals, and sample tests are appended. (Author/PN)

## A REPORT ON THE NIE/ECS ITEM DEVELOPMENT PROJECT

# The Development of Mathematical Problem-Solving Superitems

by Thomas A. Romberg, Kevin F. Collis, Brian F. Donovan, Anne E. Buchanan, and Martha N. Romberg



January 1982

Wisconsin Center for Education Research
*an institute for the study of diversity in schooling*

A report on the NIE/ECS Item Development Project

# THE DEVELOPMENT OF MATHEMATICAL PROBLEM-SOLVING SUPERITEMS

by Thomas A. Romberg, Kevin F. Collis,
Brian F. Donovan, Anne E. Buchanan, and Martha N. Romberg

Thomas A. Romberg
Principal Investigator

Wisconsin Center for Education Research
The University of Wisconsin
Madison, Wisconsin

January 1982

This material is based upon work supported by the National Institute of Education and the Education Commission of the States under Contract No. 02-81-20321 with the Wisconsin Center for Education Research. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of ECS, NIE, or the Department of Education.

## TABLE OF CONTENTS

| | Page |
|---|---|
| List of Tables | v |
| List of Figures | v |
| Abstract | vii |
| Introduction | 1 |
| Basis of the Methodology | 3 |
| Problem Situations | 3 |
| Superitems | 3 |
| Problem Situation Development | 6 |
| Basic Validity Check | 8 |
| The Use of SOLO Taxonomy Approach in Developing Superitems | 11 |
| The SOLO Categories | 11 |
| Criteria for Construction of Superitems | 15 |
| Second Validation | 15 |
| Third Validation | 15 |
| Final Revisions for Test Administrations | 20 |
| Summary | 21 |
| Appendix A: Instructions to Wausau Teachers for Judging Items | 27 |
| Appendix B: Wausau Teacher's Judgments (Tables B1-B6) | 41 |
| Appendix C: Instructions to Graduate Students for Judging Items | 47 |
| Appendix D: Test Administrators' Manuals and Sample Test Booklet Format and Instructions to Students | 61 |

List of Tables

| Table | | Page |
|---|---|---|
| 1 | Number of Item Stems by Content Categories and Level of Directness . . . . . . . . . . . . . . . . . | 7 |
| 2 | Frequency of Judgments into Six Content Categories by All 20 Teachers . . . . . . . . . . . . . . . . . | 9 |
| 3 | Frequency of Judgments into Three Cognitive Objective Categories by All 20 Teachers . . . . . . . . . . . . | 10 |
| 4 | Frequency of Judgments as to Appropriateness by All 20 Teachers . . . . . . . . . . . . . . . . . . . . . | 12 |
| 5 | Frequency of Judgments as to Appropriateness by Teachers at Grades 12, 7, and 4 . . . . . . . . . . . | 12 |
| 6 | Summary of Response Modes . . . . . . . . . . . . . . | 14 |
| 7 | Judgments of Graduate Students of the Content of Superitems and the Level of Reasoning . . . . . . . | 19 |
| 8 | Order of Superitems by Form for 17 Year Olds . . . . | 22 |
| 9 | Order of Superitems by Form for 9, 11, and 13 Year Olds . . . . . . . . . . . . . . . . . . . . . . . . | 23 |

List of Figure

| Figure | | |
|---|---|---|
| 1 | Levels of the SOLO Taxonomy and Response Description | 13 |

## Abstract

The purpose of this report is to describe the development of a pool of mathematical problem-solving situations and a set of items for each situation which provides information about students' qualitatively different levels of reasoning ability as applied to that situation.

The strategy being used is that, for each problem-solving situation, a set of "structured superitems" was developed. "Superitems" are a set of test items about a common situation or stem. Such item sets have been shown to be valid and useful for assessing mathematical problem solving. Four questions were written for each problem situation. The questions attempt to assess five levels of reasoning about the situation. The superitems then were administered to a trial population. This reports documents the details of the preparation of the items prior to their administration. A second report will describe the results of the test administration.

## Introduction

The purpose of this report is to describe the steps that were followed to develop a pool of mathematical problem-solving situations and a set of items for each situation which were designed to provide information about students' qualitatively different levels of reasoning ability.

The strategy followed was to develop a set of "structured super-items" for each of a set of problem-solving situations. The method for creating a pool of situations and questions was based on Cureton's (1965) notion of "superitems" (a set of test questions based on a common situation or stem). The structure for the superitems was based on Collis and Biggs' (1979) SOLO taxonomy used to classify the structure of observed learning outcomes. The items were prepared to be administered to students of 9, 11, 13, and 17 years of age. This report describes how the items were developed. A second project report will examine the validity of the superitems and the utility of the procedure for large scale assessments. If the goals of this study are attained, a more useful assessment procedure for this critical aspect of mathematics will be available for large scale assessments.

The project was funded by the Education Commission of the States (with funds supplied by the National Institute of Education). Ostensibly the resulting items would be useful in future National Assessment of Education Progress (NAEP) studies in mathematics.

To accomplish the goals of this study, a seven-stage project was designed.

Stage 1. December to March 1981--Problem Situation Development.

For the student populations, a set of problem situations was developed.

Stage 2. March to May 1981--Basic Validity Check.

Each problem situation was examined by classroom teachers at the respective grade levels to check on the appropriateness of the concepts and prerequisite skills for students of those ages.

Stage 3. April to July 1981--Superitem Development.

At this stage, sets of items for each situation were written, reviewed, and tried out with a small sample of students under the direction of Professor Collis. The items were again reviewed by graduate students to check the items for their mathematical appropriateness and their fit to the SOLO taxonomy. This tryout was done to ensure that students could read the items and follow directions and to see if there were any procedural problems.

Stage 4. July to September 1981--Preparation of Trial Materials.

At this stage, the set of situations and superitems appropriate for the target population was organized into batteries for administration to a large population of students.

Stage 5. September 1981--Administration of Batteries.

Early in the school year the batteries were administered to a population of students.

Stage 6. October through December 1981--Data Analysis.

All test booklets and questionnaires were scored and analysis of the data was carried out at this stage.

Stage 7. December through January 1982--Report Preparation.

In this report only information about Stages 1 to 4 are included. The administration of the items and the analysis will be reported in a second technical report.

## Basis of the Methodology

### Problem Situations

Practical problem solving has been consistently and broadly criticized, most notably by Kline (1973). Kline has been critical of the application problems appearing in mathematics texts as purporting to represent "real life" situations but having little in common with what constitutes real life. Nelson and Kirkpatrick (1975) also have emphasized the importance of "real life" situations. What is important is that the problems be drawn from situations susceptible to mathematical analysis (NACOME, 1975). A guide for the construction of appropriate situations might be found in Hilbert's (1906) comment:

> A mathematical problem should be difficult in order to entice us, yet not completely inaccessible, lest it mock at our efforts. It should be to us a guidepost on the hazy paths to hidden truths and ultimately a reminder of our pleasure in the successful solution. (p. 59)

### Superitems

In some tests, items come in groups, e.g., paragraph-reading tests with several questions on each paragraph, or table-reading tests with several items on each table. The problem situations or stems, the paragraphs and tables in the example, contain considerable information. The sets of questions with the stem are called superitems (Cureton, 1965), a term chosen to emphasize that differences among respondents in comprehension of a stem may produce correlated errors of measurement between items for the same stem. While such sets of items have long been used in some tests, the problem of correlated errors was not considered serious as long as the answer to any one question did not depend upon one's response to another question related to the stem. Cureton's interest in superitems was methodological, i.e., how to estimate the reliability of tests comprising such items. However, he argued that the specific questions within a set are not truly independent since they all depend on a basic understanding of the paragraph or table. In fact, if the questions within superitems were constructed carefully, the structure of responses within the set potentially could be revealing. The unresolved question is upon what basis structured superitems should be constructed.

It is this question which is being addressed in this study. That is, can a set of superitems be constructed so that responses will reflect "level of reasoning" used by the subject with respect to mathematical situations?

During the 1970s, Wearne and Romberg (1977) developed several versions of a superitem test of mathematical problem-solving for elementary school children. Those tests were designed to produce three scores: a comprehension score, an application score, and a problem-solving score. Thus each superitem contained a comprehension question, an application question, and a problem-solving question. The comprehension question assessed the child's understanding of the information contained either implicitly or explicitly in the item stem. The application question assessed the child's mastery of a prerequisite concept or skill of the problem-solving question (this was a fairly straightforward application of some rule or concept). The third question in the set was the problem-solving question.

These tests were then used in several studies carried out under Romberg's direction (Meyer, 1975; Meyer, 1976; Wearne, 1976; Whitaker, 1976). In all, these tests proved to be useful in providing information about a student's mastery of the prerequisites for the problem-solving question. The tests not only yielded more information about the student but also provided a "better" measure of problem-solving ability by making it possible to consider only those problems for which the student demonstrated preparedness.

However, the problem-solving items in the Romberg-Wearne tests do not yield information about the level of a student's reasoning with regard to each problem situation. The "structure" used as a basis for the development of those tests was derived from Bloom's (1956) Taxomony. The three questions for each stem were developed to reflect Bloom's categories of comprehension, application, and analysis, respectively.

For the pool of items constructed in this study, a recently developed taxonomy based on the "structure of the observed learning outcome" (SOLO) (Collis & Biggs, 1979) was used as the blueprint for the development of superitems with four questions. There were two reasons for using this taxonomy. First, the levels of reasoning reflect basic operational stages of thought related to the Piagetian stages of development. Thus, rather than being used to simply categorize errors and infer how reasoning was carried out, the responses can be aggregated with respect to this framework against which the appropriateness of learning activities eventually can be considered. Second, individual items which reflect these levels of reasoning had previously been written for the "ACER Mathematics Profile Series" (Australian Council of Educational Research, 1977).

## Problem Situation Development

Initially problem stems were written for six content categories and for direct or indirect responses. The content categories were numbers and numeration; variables and relationships; size, shape, and position; measurement; statistics and probability; and unfamiliar. These categories correspond to the five NAEP content designations and an additional area termed "unfamiliar." The levels of directness, labeled direct or indirect, evolved from consideration of research on moral dilemmas. A dilemma presents a solution requiring a choice between equally unfavorable, or disagreeable, alternatives. For Kohlberg (1963) responses to dilemmas were not seen as correct or incorrect but rather as indicative of differences in stages in moral reasoning. In mathematical problem-solving, students may be placed in a setting which requires them to examine patterns of incorrect reasoning and generate responses consistent with that pattern. In responding the student is expected to consider the reasoning process described in an item stem. Our intention was to focus on the reasoning of the student rather than the correctness of a response. Hence, students were asked to explain their answers for many items. A direct response was one which usually followed from a calculation but did not require explanation. Other direct response forms involved brief written statements or the drawing of diagrams.

The initial 40 stems were based on ideas and information from a variety of sources: mathematics journals, items developed for the National Longitudinal Study of Mathematical Abilities (Romberg & Wilson, 1969), and textual material. The distribution of the items by content and level of directness was posited, prior to the initial validation check, to be that shown in Table 1. Then for each item stem, three to five questions were written which reflected the comprehension, application, and analysis categories used by Wearne and Romberg (1977).

Table 1

Number of Item Stems

by Content Categories and Level of Directness

| Content Categories | | Levels of Directness | |
|---|---|---|---|
| | | Direct | Indirect |
| A | Numbers and Numeration | 4 | 5 |
| B | Variables and Relationships | 4 | 3 |
| C | Size, Shape, and Position | 2 | 2 |
| D | Measurement | 4 | 2 |
| E | Statistics and Probability | 4 | 3 |
| F | Unfamiliar | 3 | 4 |

## Basic Validity Check

At this stage, 20 classroom teachers (8 twelfth-grade teachers, 6 seventh-grade teachers, and 6 fourth-grade teachers) were recruited to judge the superitems on three dimensions. The teachers were all from the Wausau, Wisconsin, school district. They were recruited by DuWayne Kleinschmidt, the district's mathematics coordinator. The dimensions teachers were to consider were content--to see if the teachers thought the item stems fit the six NAEP content categories; reasoning levels--to see if teachers judged each question in a superitem to fit one of three objective categories; and appropriateness--to see if teachers thought the questions in each superitem were appropriate for students at their grade level.

The questions for each stem were randomly ordered and the superitems also were randomly ordered and then the set of superitems for each teacher was sent to Wausau for judging. The judging was done during March 1981. The instructions for judging appear in Appendix A.

The results of the judging have been summarized by presenting the data in a set of contingency tables. First, the judgment of teachers about the content of the stems is shown in Table 2. With the exception of the seven "unfamiliar" stems, content agreement by the teachers appears to be fairly consistent with the content categories for which the items were written. Given the apparent agreement both an index of nominal agreement ($P_o$) for the total judgments (Frick & Semmel, 1978) and Light's (1971) coefficient $\kappa_{p_i}$ for agreement on each category were calculated. For these data, $P_o \geq .61$ and $\kappa_{p_i}$ = .68, .48, .57, .51, .82, and .19 for categories A to F respectively with an average $\kappa_{p_i}$ = .54. These indices

Table 2

Frequency of Judgments into Six Content Categories by All 20 Teachers

| Teacher's Judgment | Type of Item A (9 items) | B (7 items) | C (4 items) | D (6 items) | E (7 items) | F (7 items) |
|---|---|---|---|---|---|---|
| *A Numbers and Numeration | 137 | 34 | 1 | 18 | 5 | 24 |
| B Variables and Relationships | 18 | 79 | 0 | 2 | 6 | 24 |
| C Size, Shape and Position | 2 | 0 | 50 | 21 | 0 | 33 |
| D Measurement | 4 | 6 | 23 | 69 | 6 | 5 |
| E Statistics and Probability | 14 | 21 | 0 | 7 | 121 | 21 |
| F Unfamiliar | 4 | 0 | 6 | 3 | 2 | 33 |

*Problem A-3 omitted from one teacher's review.

are all significantly different from zero at $p < .01$, and only $\kappa_{p_i}$ for the "unfamiliar" is practically low. Since numbers or relationships between numbers are a part of most of the "unfamiliar content" stems, this result was not surprising. If one omits judgments on this category, then the nominal agreement $P_o = .69$ and the average $\kappa_{p_i} = .61$. The judgments on content by the teachers at each grade level are very similar. Tables for teachers at each grade level appear in Appendix B.

The teacher's judgments about cognitive objectives for each question within the superitems is shown in Table 3. Clearly, overall agreement of teacher judgments with the intended cognitive level for each question was good. For these data the index of nominal agreement $P_o = .62$ and Light's indices are $\kappa_{p_i} = .49$, .33, and .47 for comprehension, application, and analysis, respectively, with an average $\kappa_{p_i} = .43$. Again all indices are statistically significant. The judgments about cognitive level for teachers at each grade level also appear in Appendix B.

Table 3

Frequency of Judgments into Three Cognitive Objective Categories by All 20 Teachers

| Teacher's Judgments | Item Level of Cognitive Objective | | |
|---|---|---|---|
| | Comprehension | Application | Analysis |
| Comprehension | 506 | 131 | 46 |
| Application | 227 | 480 | 250 |
| Analysis | 64 | 169 | 509 |
| Other | 3 | 20 | 5 |

The teacher's judgments about appropriateness for each question are shown in Table 4. Overall 74.5% of the questions were considered appropriate. However, the judgments by teachers at different grades were considerably different. This information by grade level appears in Table 5. Almost all of the questions (94.3%) were considered appropriate by the twelfth grade teachers while 73.5% and 49% of the questions were considered appropriate by seventh or fourth grade teachers, respectively. Obviously, many questions were considered to be too difficult for fourth-grade children or on content they had not covered.

## The Use of SOLO Taxonomy Approach in Developing Superitems

This information from teachers provided us a bank upon which revisions and rewriting could be done. Beginning April 1981 when Professor Collis arrived in the U.S., the questions for each item were rewritten according to the SOLO taxonomy (Collis & Briggs, 1979). Levels of directness were discontinued as an aspect in the development of items.

The SOLO taxonomy was designed as a response model, the basic idea being that the child is given information or data and asked a question which can be answered by reference to that information. The child's response is classified as belonging to one of five levels according to the way in which the response is structured.

## The SOLO Categories

There are five basic categories into which responses may be classified using this method. They are set out diagrammatically in Figure 1. The figure is meant to cover the general case. For a particular content area,

Table 4

Frequency of Judgments as to Appropriateness

by All 20 Teachers

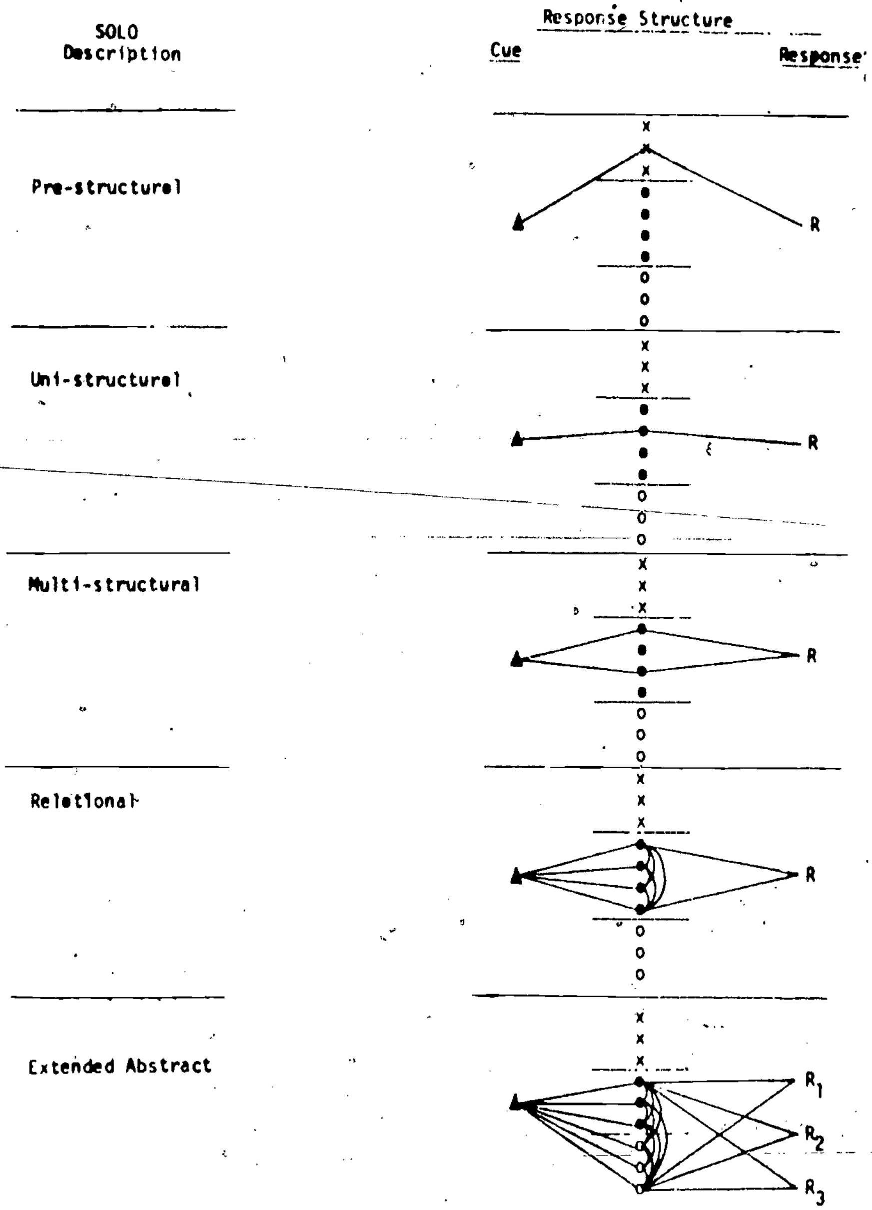
| Teacher's Judgment | f | % |
|---|---|---|
| Appropriate | 1787 | 74.5 |
| Inappropriate | 532 | 22.3 |
| Undecided | 64 | 2.7 |
| Other | 17 | .71 |

Table 5

Frequency of Judgments as to Appropriateness

by Teachers at Grades 12, 7, and 4

| Teacher's Judgment | Grade 12 | | Grade 7 | | Grade 4 | |
|---|---|---|---|---|---|---|
| | f | % | f | % | f | % |
| Appropriate | 905 | 94.3 | 529 | 73.5 | 353 | 49.0 |
| Inappropriate | 35 | 3.6 | 163 | 22.6 | 334 | 46.4 |
| Undecided | 8 | .80 | 26 | 3.6 | 30 | 4.2 |
| Other | 12 | 1.3 | 2 | .28 | 3 | .42 |




KEY: Kinds of data used:

x = Irrelevant or inappropriate

● = Related and given in display

o = Related and hypothetical, not given.

Figure 1. Levels of the SOLO taxonomy and response description (Collis & Biggs, 1979, p. 16).

certain idiosyncracies peculiar to that area need to be taken into account. For mathematics, the information in Table 6 can serve as a highly condensed summary of response models which are meaningful within the content of school-based mathematical material.

Table 6

Summary of Response Modes for Mathematics

| Responses | Response Modes |
|---|---|
| UNI-STRUCTURAL | Makes a single direct relationship to concretely (either physically or iconically) available criteria |
| MULTI-STRUCTURAL | Handles multiple operations with small numbers by a series of meaningful closures, for instance, may be seen as analogous to using a sequence of given propositions to support a particular judgment in other content areas |
| RELATIONAL | Relates elements within the immediately available concrete system and forms generalizations on this basis |
| EXTENDED ABSTRACT | Acceptance of lack of closure, use of the reciprocal operation, and ability to work with multiple interacting and abstract systems involve a comprehensive use of the given data together with related hypothetical constructs |

It was hypothesized that, by using the SOLO technique in reverse as it were, one ought to be able to design items such that a series of questions on the stem would require more and more sophisticated use of the information in the stem to obtain a correct result. This increase in sophistication should parallel the increasing complexity of structure noted in the SOLO categories.

## Criteria for Construction of Superitems

Clearly, the construction of these items consists of two parts: first, the writing of the stem and, second, the construction and writing of questions to reflect the SOLO levels. The former is concerned with content validity and will not be dealt with here. The latter requires that suitable criteria be set up to enable four questions to be asked that will not only require a knowledge of the information in the stem but will also, if correctly responded to, indicate an ability to respond to that information at least at the level reflected in the SOLO structure of the particular question. To achieve this last, the following criteria were set up for designing the questions:

1. Uni-structural: Use of one obvious piece of information coming directly from the stem.
2. Multi-structural: Use of two or more discrete closures directly related to separate pieces of information contained in the stem.
3. Relational: Use of two or more closures directly related to an integrated understanding of the information in the stem.
4. Extended Abstract: Use of an abstract general principle or hypothesis which is derived from or suggested by the information in the stem.

In each superitem, the correct achievement of question 1 would indicate an ability to respond to the problem concerned at at least the uni-structural level. Likewise success on question 2 corresponds to an ability to respond at multi-structural level and so on.

### Second Validation

The items were extensively revised following their validation check by Wausau teachers and restructured according to the Collis' SOLO taxonomy.

Some stems were eliminated and new ones written, and some "unfamiliar" stems were revised and categorized under other content headings. Selected items were administered to 6 children from Shawno elementary and middle schools, 30 from Cottage Grove elementary school, and 6 from Monona Grove High School. The Shawno children were ages 8 to 14 and distributed two per grade from grades 3, ,, and 8. Eleven children from Cottage Grove were drawn from grade 2, 9 from grade 5, and 10 from grade 8. The 6 Monona Grove students had successfully completed a grade 12 statistics class. At each grade level, the children involved in the test program were close to the average age and were considered by the grade teacher to represent from the middle to upper achievement levels in mathematics.

Children from Shawno and Cottage Grove were individually administered four and eight items, respectively. Each child was administered common items with the remaining items assigned from the pool of 40 items so that each item was administered at least once per grade level. The six students at Monona Grove were each given different starting points in sets of items in the same order and instructed to use the one hour test administration period to work as many items as possible. Both the individual interviews and the group administered tests had the purpose of identifying the SOLO level used by each child in each item. Account was also taken of language and conceptual difficulties children experienced.

Results indicated a great deal of consistency in the SOLO levels recorded for each child and also for children at the same grade level. The grade 2 children were generally operating at a uni-structural level, the grade 5 children at a multi-structural level, the grade 8 children at the transition from multi-structural to relational levels, and those from grade

12 at the transition from relational to extended abstract level. Variance in levels for each child was almost wholly within one response category of the level of reasoning generally observed for the grade.

Based on this information, all items were reviewed and many revisions were made. At this stage the content categorization of items was not altered. Some items, however, were adjusted to reduce calculation work. The wording of some items was modified to abbreviate the length of the stem, reduce complexity, or simplify technical language. Some students had difficulty following the information given in the stems of some of the geometry items. The stems were rewritten to present the information in both diagrammatic and verbal forms. Finally, some questions were rewritten to elicit responses at the desired level of reasoning.

## Third Validation

In June 1981, six graduate students in mathematics at the University of Wisconsin-Madison responded to the pool of 40 items which had been revised after the tryouts at Shawno, Cottage Grove, and Monona Grove. The graduate students were instructed to work each item, classify each as being primarily in one of the five NAEP content categories or most likely to be outside the domain of the mathematics regularly covered in school. In addition, the students were to identify, for each question in the items, the level of reasoning likely to be employed. Both the items and the questions within each item were randomly ordered for the validity check. Prior to being given the items, the students were instructed in the theory and development of the SOLO taxonomy. Sample items were provided and discussed for levels of reasoning. Instructions to the graduate students appear in Appendix C.

The results shown in Table 7 indicate a generally high level of agreement for both content and level of reasoning categorizations. The nominal index of agreement for the content of judgments was high, $P_o = .68$; and Light's index for content categories was $\kappa_{p_i} = .70, .53, .83, .62, .67$, and .22 for categories A to F, respectively, with an average of .60. Again while all were statistically significant, only the $\kappa_{p_i}$ for the "unfamiliar" stems was practically low. For the level of reasoning judgments, all the indices were very high: $P_o = .86$, $\kappa_{p_i} = .85, .76, .79$, and .83 for levels 1 to 4, respectively, with an average of .81. In classifying for content items which the authors considered would be "unfamiliar," there was a wide response spread. This phenomenon was also noted earlier in the responses of the Wausau teachers. During the debriefing session with graduate students, discrepancies were discussed, in particular, reasons for the low agreeement on the "unfamiliar stems" level. In general, it was felt that while the item content may not have been commonly a part of school mathematics, in many cases there was disagreement and thus the item was categorized into one of the other content categories. Again if judgments on the unfamiliar items are omitted, the overall indices on content judgments are much higher: $P_o = .73$ and the average $\kappa_{p_i} = .67$.

Since these indices were quite high for judgments about content and particularly for judgments on level of reasoning, we felt content or face validity of the superitems had been demonstrated.

Table 7

Judgments of Graduate Students of the Content of Superitems and the Level of Reasoning

| Items | Content Category A | B | C | D | E | F | Level of Reasoning 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|---|---|---|---|---|
| A (1-8) | 37 | 5 | - | - | 2 | 4 | | | | |
| a | | | | | | | 41 | 6 | 1 | - |
| b | | | | | | | 6 | 37 | 4 | 1 |
| c | | | | | | | - | 5 | 36 | 7 |
| d | | | | | | | - | - | 7 | 41 |
| B (1-8) | 8 | 30 | 5 | 2 | 1 | 2 | | | | |
| a | | | | | | | 46 | 2 | - | - |
| b | | | | | | | 2 | 44 | 2 | - |
| c | | | | | | | - | 9 | 39 | - |
| d | | | | | | | - | 3 | 7 | 38 |
| C (1-6) | - | - | 3 | 1 | - | 4 | | | | |
| a | | | | | | | 30 | 4 | 2 | - |
| b | | | | | | | 4 | 27 | 5 | - |
| c | | | | | | | 1 | 3 | 31 | 1 |
| d | | | | | | | - | - | 2 | 34 |
| D (1-6) | 1 | 1 | 5 | 24 | 1 | 4 | | | | |
| a | | | | | | | 28 | 8 | - | - |
| b | | | | | | | 6 | 26 | 4 | - |
| c | | | | | | | - | 9 | 26 | 1 |
| d | | | | | | | - | - | 5 | 31 |
| E (1-8) | 5 | 6 | - | 1 | 35 | 1 | | | | |
| a | | | | | | | 46 | 2 | - | - |
| b | | | | | | | 1 | 42 | 5 | - |
| c | | | | | | | - | 4 | 39 | 5 |
| d | | | | | | | - | - | 6 | 42 |
| F (1-4) | 7 | 5 | 2 | 1 | 2 | 7 | | | | |
| a | | | | | | | 22 | 2 | - | - |
| b | | | | | | | 1 | 22 | 1 | - |
| c | | | | | | | - | 1 | 21 | 2 |
| d | | | | | | | - | - | 1 | 23 |

Final Revisions for Test Administrations

A final technical review of all items was carried out by the project staff under the direction of the test development specialist. This review was in part editorial; for example, wording was simplified, tenses were checked, and agreement in terminology and symbols among the stem and all questions for each item was inspected. Further, the appropriateness of vocabulary both in terms of the grade levels to be tested and general familiarity to students was reexamined. Art work was reviewed to insure that content was consistent with the narrative, drawings were accurate and to scale, and labeling was adequate.

Item and test format as a whole were also reexamined at this time. Such considerations as sufficient space for student responses, standard size and terms for unknowns, and possible confusion between labels for an item and information within the item itself were checked. All items were also worked once again as a final verification of expected responses.

Separate group-administered test batteries were then prepared for 17 year olds, and for 9, 11, and 13 year olds. Separate batteries were necessary because the test items for 17 year olds included the stem and questions for the four levels of reasoning whereas the tests for ages 9, 11, and 13 did not include the extended abstract question. From the pool of 40 items, one B item which had been particularly enjoyable for students and had discriminated well among levels of reasoning was chosen for the sample item. One of the least successful items in the validity checks was discarded. It was decided that three items judged to be quite difficult for 17 year olds should be administered at that age level only; and three items which were particularly easy for 17 year olds were assigned only to the tests for ages 9, 11, and 13. Thus, there were 38 items total and 35 items in each of the two batteries.

Five test forms of seven items each were created for the two test batteries by randomly assigning items with the restriction that each content category (except unfamiliar) be represented at least once but no more than twice per form. The assignment was adjusted so that items in the same content category were not contiguous within each form. The assignment of items is outlined in Tables 8 and 9.

Test administrator's manuals and directions for students were prepared for the two batteries. Copies appear in Appendix D.

## Summary

The procedures described above are typical in any test development effort. From the care in preparing items and the various tryouts, we were reasonably assured that the 38 items in the final batteries were mathematically correct and were related to the levels of reasoning in the SOLO taxonomy.

Table 8

Order of Superitems by Form for 17 Year Olds

| Question Number | Form | | | | |
|---|---|---|---|---|---|
| | S1 | S2 | S3 | S4 | S5 |
| 1 | C6 | C5 | F3[a] | C3 | B7 |
| 2 | B2 | D6 | E7 | B6 | D1 |
| 3 | E8 | A3 | C2 | D3 | E3[a] |
| 4 | F1[a] | D2 | D4 | B4 | F2 |
| 5 | D5 | E6 | B5 | E5 | C1 |
| 6 | B8 | B3 | C4 | A8 | A1 |
| 7 | A4 | E1 | A6 | F4 | E4 |

[a]Item not included in tests for 9, 11, and 13 year olds.

Table 9

Order of Superitems by Form for 9, 11, and 13 Year Olds

| Question Number | Form | | | | |
|---|---|---|---|---|---|
| | UMR1 | UMR2 | UMR3 | UMR4 | UMR5 |
| 1 | A3 | B4 | C2 | F4 | C5 |
| 2 | B3 | F2 | B5 | A6 | A2[a] |
| 3 | D2 | E7 | D4 | B7 | E5 |
| 4 | E6 | B8 | E4 | A7[a] | B2 |
| 5 | C6 | A1 | A5[a] | E1 | A4 |
| 6 | D1 | C4 | E8 | D3 | D5 |
| 7 | A8 | D6 | B6 | C3 | C1 |

[a]Item not included in tests for 17 year olds.

References

Australian Council for Educational Research Limited. Operations Test Teachers Handbook, ACER Mathematics Profile Series, 1977.

Bloom, B. (Ed.) Taxonomy of educational objectives, handbook I: Cognitive domain. New York: David McKay Co., 1956.

Collis, K.F., & Biggs, J.B. Classroom examples of cognitive development phenomena: The SOLO taxonomy. Report prepared at conclusion of an Educational Research and Development Committee funded project, University of Tasmania, 1979.

Cureton, E.E. Reliability and validity: Basic assumptions and experimental designs. Educational and Psychological Measurement, 1965, 25, 326-346.

Frick, T., & Semmel, M.I. Observer agreement and reliabilities of classroom observational measures. Review of Educational Research, 1978, 48, 157-184.

Hilbert, D. Mathematical problems. Archives of Mathematics and Physics, 1906, 111-119.

Kline, M. Why Johnny can't add: The failure of the new math. New York: Vintage Books, 1973.

Kohlberg, L. The development of children's orientations toward a moral order: 1. Sequence in the development of moral thought. Vita Ilumina, 1963, 6, 11-33.

Light, R.J. Measures of response agreement for qualitative data: Some generalizations and alternatives. Psychological Bulletin, 1971, 76, 365-377.

Meyer, R.A. A study of the relationship of mathematical problem-solving performance and intellectual abilities of fourth-grade children. Unpublished doctoral dissertation, University of Wisconsin, 1975.

Meyer, R.A. A study of the relationship of mathematical problem-solving performance and intellectual abilities of fourth-grade boys and girls (Working Paper No. 160). Madison: Wisconsin Research and Development Center, 1976.

National Advisory Committee on Mathematical Education. Overview and analysis of school mathematics K-12. Washington, D.C.: Authors, 1975.

Nelson, L.D., & Kilpatrick, J. Problem solving. In J. Payne (Ed.), Mathematics learning in early childhood. Reston, VA: National Council of Teachers of Mathematics, 1975.

Romberg, T.A., & Wilson, J.W. The development of tests. In J.W. Wilson, L.S. Cahen, and E.G. Begle (Eds.), NLSMA Reports Volume 7. Stanford, CA: School Mathematics Study Group, 1969.

Wearne, D.C. Development of a test of mathematical problem-solving which yields comprehension, application, and problem-solving scores. Unpublished doctoral dissertation, University of Wisconsin, 1976.

Wearne, D.C., & Romberg, T.A. DMP accountability tests (Working Paper No. 217). Madison: Wisconsin Research and Development Center, 1977.

Whitaker, D. A study of the relationships between selected non-cognitive factors and the problem-solving performance of fourth-grade children (Technical Report No. 396). Madison: Wisconsin Research and Development Center, 1976.

# Appendix A

## INSTRUCTIONS TO WAUSAU TEACHERS FOR JUDGING ITEMS

# INSTRUCTIONS FOR SORTING TEST SUPERITEMS

## Materials

You should have

(1) a set of 40 superitems
(2) 1 large envelope
(3) a description of the project
(4) Special Sorting Instructions
(5) Background Information Form

## Tasks

(1) Read the description of the project.
(2) You are to find the answers to each question for each superitem.
(3,4,5) You are then to sort the items three times into sets according to the special sorting instructions.
(6) Fill out the background information form.
(7) Place all superitems and the background information in the large envelope and return it to Duane Kleinschmidt.

## Caution

(1) The sorting of superitems must be done independently.
(2) Carefully follow the steps outlined in the instructions.

## Frame of Reference for Judging

You are to judge the items on the basis of how a typical "B" student who has just completed fourth, seventh, or eleventh grade (the grade prior to when you teach them) would react to the problem.

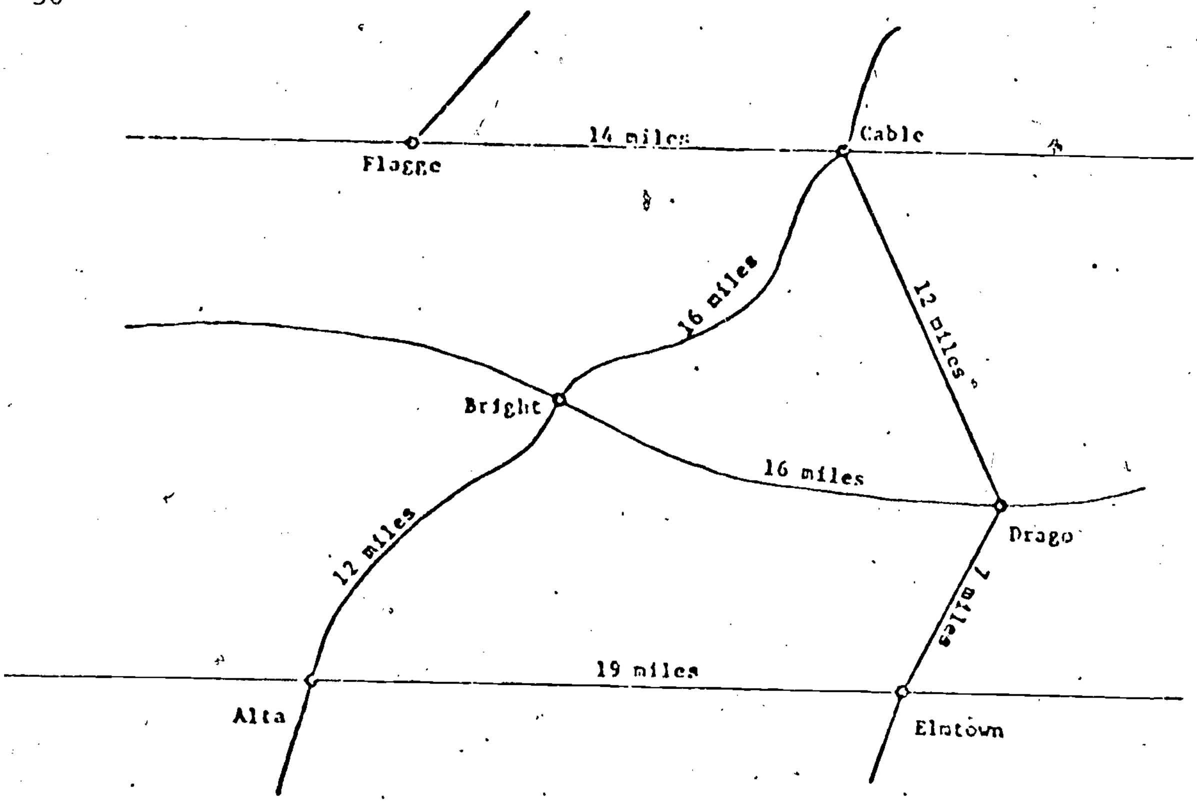
Task 1. Read the following information.

Description of the Project

The purpose of this study is to develop a pool of mathematical problem solving situations and a set of items for each situation which provide information about the student's qualitatively different levels of reasoning ability applied to that situation. The strategy being used is that for each problem solving situation, a set of "structured superitems" will be developed. Thus, a superitem contains a problem situation or stem and a set of questions about the information presented in that stem. For example, the problem in Figure 1 is a map with lots of information about towns, roads, and distances. Three questions follow which can be answered based on that information. However, the questions differ in level of reasoning required to find the answer. The structure for deriving the levels of reasoning in the superitems is based on a five-level taxonomy used to classify the structure of observed learning outcomes developed by Kevin Collis.

Assuming that a satisfactory pool of items can be developed, a sample of about 300 students from a population of 9, 13, and in-school 17 year olds in the Wausau schools will be given sets of the items next September. In addition, they will be asked to complete a background questionnaire and an attitude instrument. From this data we expect to demonstrate that this procedure for assessing mathematical reasoning will be a viable method of gathering, analyzing, and reporting data for large-scale assessment projects such as the National Assessment of Educational Progress (NAEP) who has funded this project.

The information derived should enable educators to identify levels of learned outcomes which should be useful in examining differential performance by diverse groups of students.




The distance from Alta to Bright is:
- 7 miles
- 12 miles
- 16 miles
- 19 miles

The shortest distance from Alta to Drago is:
- through Bright
- through Cable
- through Elmtown
- through Flagge

The sign

| | |
|---|---|
| BRIGHT | 16 |
| ELMTOWN | 19 |

should be placed:
- in Drago
- in Alta
- in Flagge
- in Cable

Figure 1. An example of a superitem (Wearne, 1976).

Purpose of Your Work

The development of new tests takes a lot of time and effort. After preparing a set of items one needs to know if the items are valid and appropriate. We are asking you to help us make judgments about the validity and appropriateness of an initial set of superitems. After working the items (Task 2), you are to sort the items three times (Tasks 3, 4, and 5) following the special sorting instructions. The sortings will allow us to judge both validity and appropriateness.

Task 2. Open the set of superitems. Find answers to each of the questions for each superitem.

This task is to familiarize you with the superitems to facilitate the sorting which follows. (We are not going to score you on number of items you got correct.) However, please jot down any comments on each page about the items, wording, etc, as you do them. Also, for one of the sorting procedures, the set of questions associated with each stem have been put in random order. That would <u>not</u> be the case when superitems are given to students. For them, the sets of questions would be ordered from least complex to most complex.

Special Sorting Instructions

Task 3. Content Sorting

In this task you are to sort the set of 36 superitems into six content categories. The purpose of this sorting is to establish the content validity of the items. The term, validity, refers to judging whether the items measure what they were intended to measure.

Step 1. Read the following information about the content categories.

There are six content categories used in this assessment:

A. Numbers and Numeration

B. Variables and Relationships

C. Size, Shape and Position

D. Measurement

E Statistics and Probability

F. Familiar

These content categories organize the domain, but are not intended to be represented equally in the assessment.

A. Numbers and Numeration

This category deals with the ways numbers are used, processed or written. Number properties and order relations are also included.

B. Variables and Relationships

The questions in this content category deal with facts, definitions and symbols of algebra; the use of variables in equations and inequalities; the use of variables to represent elements of a number system; functions and formulas; and coordinate systems.

C. Shape, Size and Position

The questions in this content category measure objectives related to school geometry. The emphasis is not on geometry as a formal deductive system. The questions concern plane and solid shapes, congruence, similarity, properties of triangles, properties of quadrilaterals, constructions, sections of solids, other basic theorems and relationships, rotations and symmetry.

D. Measurement

These items cover appropriate units; equivalence relations; instrument reading; length, weight, capacity, time and temperature; perimeter, area and volume; non-standard units; and precision and interpolation.

E. Statistics and Probability

This content area, statistics and probability, assesses data collection; organizing data with tables, charts, graphs; finding patterns in data, interpreting and analyzing data; drawing inferences; making generalizations and predicting outcomes.

F. Unfamiliar

This content area contains information which is not a part of a mathematics program. While in each of the other categories, some prior information about the mathematical ideas can be assumed to have been taught in schools, in this category that is not the case.

Step 2: Sorting. Starting with the first superitem in the set (each superitem is on a separate page) study the stem to determine what it tests. Little attention should be given to such characteristics as difficulty, relative quality of distractors (incorrect alternatives), or to length of the item stem. Instead, attention should be focused on the content which a student should be familiar with to find the correct answer. Once you have decided what content it tests, put it down.

Study the second stem in the same fashion. If you decide that the first and second superitems test the same content, place it with the first stem. Otherwise, form a second category.

Consider subsequent items similarly and in each case decide the content the superitem tests.

Step 3: Resorting. Review the six categories. Check to see that all items within a category are--to your satisfaction--similar with respect to what content they test. When they are not, sort inappropriate items into other categories.

Step 4: Labeling. When you are satisfied the items are in appropriate categories, label each item in the content box with the letter A, B, C, D, E or F associated with each content category.

Task 4. Judgment About the Level of Reasoning

In this task you are to judge which of three levels of reasoning is required by each of the questions in each superitem. The purpose of this judgement task is to establish the construct validity of the questions in each superitem.

Step 1. Read the following information about the levels of reasoning.

There are three levels of reasoning used in this assessment.

A. Comprehension

B. Application

C. Analysis

A. Comprehension

Comprehension represents the lowest level of understanding. It refers to a type of understanding or apprehension such that the individual knows what is being communicated and can make use of the material or idea being communicated without necessarily relating it to other material or seeing its fullest implications. If a question only requires a student to understand what is being communicated in order to answer it, it is to be labeled as comprehension. For example, question (a) in Figure 1 requires only that the student comprehends how distance is communicated on a map.

B. Application

Application concerns the use of rules or procedures in particular and concrete situations. If a question requires a student to use a standard rule(s) or procedure(s) to find the answer, it is to be labeled, application. For example, in Figure 1 question (b) requires adding indicated distances and comparing sums.

C. Analysis

Analysis involves the breakdown of a communication into its constituent elements or parts such that the relative hierarchy of ideas is made clear and/or the relations between the ideas expressed are made explicit. For example in Figure 1 question (c) requires a student to break down the information in the question in order to determine the connection between the elements. Questions such as that are to be labeled, analysis.

Caution: Before you start deciding on level of reasoning re-read Frame of Reference for Judging.

Step 2. Judging. Starting with the first superitem in the set, study question (a) and determine the level of reasoning that is needed. Once you have decided what level of reasoning is appropriate, write A, B, or C in the box in front of question (a).

Study question (b) in the same superitem in the same fashion and follow the same procedures. Consider subsequent questions for each superitem similarly.

Task 5. Judgment about Appropriateness

In this task you are to judge whether each question in a superitem is appropriate for students at the grade level you teach. (Again, refer to the Frame of Reference for Judging.)

Step 1. Read the following.

Three levels of appropriateness are to be considered:

A. Appropriate

B. Inappropriate

C. Undecided.

A. Appropriate

A question is to be judged appropriate if the terminology, symbols, operations, rules or procedures are likely to be known by a student at that grade level.

B. Inappropriate

A question is to be judged inappropriate if the terminology, etc. would not be commonly known by a student at that grade level.

C. Undecided. If you have doubt, code C and comment on your doubt.

Step 2. Judging. Starting with the first superitem in the set, study the stem and then question (a) and determine if it is appropriate. Once you have decided write A, B, or C in the appropriateness box in back of question a.

Study question (b) in the same superitem in the same manner. Consider subsequent superitems and questions similarly.

Task 6. Fill out the following questionnaire.

Name ______________________________

School ______________________________

Grade level you teach ________

Number of years teaching ______

Sex______ Ethnicity ______

Task 7. Now put all of the superitems and this questionnaire in the big envelope and return it to Duane Kleinschmidt.

Thank you for completing this set of tasks. You will receive a check for $50 for your effort within the next few weeks.

Appendix B

WAUSAU TEACHER'S JUDGMENTS

Tables B-1 to B-6

Table B-1

Frequency of Judgments into Six Content Categories by Eight Grade 12 Teachers

| Teacher's Judgment | Type of Item<br>A (9 items) | B (7 items) | C (4 items) | D (6 items) | E (7 items) | F (7 items) |
|---|---|---|---|---|---|---|
| *A Numbers and Numeration | 54 | 21 | 0 | 7 | 1 | 17 |
| B Variables and Relationships | 8 | 26 | 0 | 1 | 4 | 8 |
| C Size, Shape and Position | 1 | 0 | 19 | 5 | 0 | 14 |
| D Measurement | 1 | 2 | 13 | 30 | 4 | 3 |
| E Statistics and Probability | 7 | 7 | 0 | 4 | 47 | 10 |
| F Unfamiliar | 0 | 0 | 0 | 1 | 0 | 4 |

*Problem A-3 was omitted from one teacher's review.

Table B-2

Frequency of Judgments into Six Content Categories by Six Grade 7 Teachers

| Teachers's Judgment | Type of Item | | | | | |
|---|---|---|---|---|---|---|
| | A (9 items) | B (7 items) | C (4 items) | D (6 items) | E (7 items) | F (7 items) |
| A Numbers and Numeration | 41 | 11 | 1 | 5 | 3 | 16 |
| B Variables and Relationships | 4 | 23 | 0 | 1 | 1 | 5 |
| C Size, Shape and Position | 1 | 0 | 17 | 9 | 0 | 11 |
| D Measurement | 1 | 2 | 4 | 19 | 2 | 2 |
| E Statistics and Probability | 4 | 6 | 0 | 0 | 34 | 0 |
| F Unfamiliar | 3 | 0 | 2 | 2 | 2 | 8 |

Table B-3

Frequency of Judgments into Six Content Categories by Six Grade 4 Teachers

| Teacher's Judgment | Type of Item: A (9 items) | B (7 items) | C (4 items) | D (6 items) | E (7 items) | F (7 items) |
|---|---|---|---|---|---|---|
| A Numbers and Numeration | 42 | 2 | 0 | 6 | 1 | 6 |
| B Variables and Relationships | 6 | 30 | 0 | 0 | 1 | 11 |
| C Size, Shape and Position | 0 | 0 | 14 | 7 | 0 | 7 |
| D Measurement | 2 | 2 | 6 | 20 | 0 | 1 |
| E Statistics and Probability | 2 | 2 | 0 | 3 | 40 | 11 |
| F Unfamiliar | 1 | 0 | 4 | 0 | 0 | 6 |

Table B-4

Frequency of Judgments into Three Cognitive

Objective Categories by Grade 12 Teachers

| Teacher's Judgment | Item Level of Cognitive Objective | | |
|---|---|---|---|
| | Comprehension | Application | Analysis |
| Comprehension | 208 | 61 | 22 |
| Application | 92 | 201 | 116 |
| Analysis | 17 | 45 | 177 |
| Other | 3 | 13 | 5 |

Table B-5

Frequency of Judgments into Three Cognitive

Objective Categories by Grade 7 Teachers

| | Item Level of Cognitive Objective | | |
|---|---|---|---|
| | Comprehension | Application | Analysis |
| Comprehension | 169 | 37 | 9 |
| Application | 57 | 157 | 69 |
| Analysis | 17 | 45 | 177 |
| Other | 3 | 13 | 5 |

Table B-6

Frequency of Judgments into Three Cognitive Objective Categories by Grade 4 Teachers

| Teachers' Judgments | Item Level of Cognitive Objective | | |
|---|---|---|---|
| | Comprehension | Application | Analysis |
| Comprehension | 129 | 33 | 15 |
| Application | 78 | 122 | 65 |
| Analysis | 33 | 82 | 160 |
| Other | 0 | 3 | 0 |

# Appendix C

## INSTRUCTIONS TO GRADUATE STUDENTS FOR JUDGING ITEMS

# INSTRUCTIONS FOR SORTING TEST SUPERITEMS

## Materials

You should have

(1) a set of 40 superitems
(2) 1 large envelope
(3) a description of levels of reasoning
(4) Special Sorting Instructions

## Tasks

(1) Find the answers to each question for each superitem.
(2) You are then to sort the items into content areas according to the special sorting instructions.
(3) You are then to decide the level of reasoning required by each question.
(4) Place all superitems in the large envelope and return it to Brian Donovan.

Description of the Project

The purpose of this study is to develop a pool of mathematical problem solving situations and a set of items for each situation which provide information about the student's qualitatively different levels of reasoning ability applied to that situation. The strategy being used is that for each problem solving situation, a set of "structured superitems" will be developed. Thus, a superitem contains a problem situation or stem and a set of questions about the information presented in that stem. Four or five questions follow which can be answered based on that information. However, the questions differ in level of reasoning required to find the answer. The structure for deriving the levels of reasoning in the superitems is based on a five-level taxonomy used to classify the structure of observed learning outcomes developed by Kevin Collis.

Assuming that a satisfactory pool of items can be developed, a sample of about 300 students from a population of 9, 13, and in-school 17 year olds will be given sets of the items next September. In addition, they will be asked to complete a background questionnaire and an attitude instrument. From this data we expect to demonstrate that this procedure for assessing mathematical reasoning will be a viable method of gathering, analyzing, and reporting data for large-scale assessment projects such as the National Assessment of Educational Progress (NAEP) who has funded this project.

The information derived should enable educators to identify levels of learned outcomes which should be useful in examining differential performance by diverse groups of students.

Purpose of Your Work

The development of new tests takes a lot of time and effort. After preparing a set of items, one needs to know if the items are valid. We are asking you to help us make judgments about the validity of a set of superitems. After working the items (Task 1), you are to sort the items twice (Tasks 2 and 3) following the special sorting instructions. The sortings will allow us to judge both validity and appropriateness.

Task 1. Open the set of superitems. Find answers to each of the questions for each superitem.

This task is to familiarize you with the superitems to facilitate the sorting which follows. (We are not going to score you on number of items you got correct.) However, please jot down any comments on each page about the items, wording, etc, as you do them. Also, for one of the tasks, the set of questions associated with each stem have been put in random order. That would not be the case when superitems are given to students. For them, the sets of questions would be ordered from least complex to most complex.

Special Sorting Instructions

Task 2. Content Sorting

In this task you are to sort the set of 40 superitems into six content categories. The purpose of this sorting is to establish the content validity of the items. The term, validity, refers to judging whether the items measure what they were intended to measure.

Step 1. Read the following information about the content categories.

There are six content categories used in this assessment:

A. Numbers and Numeration

B. Variables and Relationships

C. Size, Shape and Position

D. Measurement

E. Statistics and Probability

F. Unfamiliar

These content categories organize the domain, but are not intended to be represented equally in the assessment.

A. Numbers and Numeration

This category deals with the ways numbers are used, processed or written. Number properties and order relations are also included.

B. Variables and Relationships

The questions in this content category deal with facts, definitions and symbols of algebra; the use of variables in equations and inequalities; the use of variables to represent elements of a number system; functions and formulas; and coordinate systems.

C. Shape, Size and Position

The questions in this content category measure objectives related to school geometry. The emphasis is not on geometry as a formal deductive system. The questions concern plane and solid shapes, congruence, similarity, properties of triangles, properties of quadrilaterals, constructions, sections of solids, other basic theorems and relationships, rotations and symmetry.

D. Measurement

These items cover appropriate units; equivalence relations; instrument reading; length, weight, capacity, time and temperature; perimeter, area and volume; non-standard units; and precision and interpolation.

E. Statistics and Probability

This content area, statistics and probability, assesses data collection; organizing data with tables, charts, graphs; finding patterns in data, interpreting and analyzing data; drawing inferences; making generalizations and predicting outcomes.

F. Unfamiliar

This content area contains information which is not a part of a mathematics program. While in each of the other categories, some prior information about the mathematical ideas can be assumed to have been taught in schools, in this category that is not the case.

Step 2: Sorting. Starting with the first superitem in the set (each superitem is on a separate page) study the stem to determine what it tests. Little attention should be given to such characteristics as difficulty, relative quality of distractors (incorrect alternatives), or to length of the item stem. Instead, attention should be focused on the content which a student should be familiar with to find the correct answer. Once you have decided what content it tests, put it down.

Study the second stem in the same fashion. If you decide that the first and second superitems test the same content, place it with the first stem. Otherwise, form a second category.

Consider subsequent items similarly and in each case decide the content the superitem tests.

Step 3: Resorting. Review the six categories. Check to see that all items within a category are--to your satisfaction—similar with respect to what content they test. When they are not, sort inappropriate items into other categories.

Step 4: Labeling. When you are satisfied the items are in appropriate categories, label each item in the upper left-hand corner with the letter A, B, C, D, E, or F associated with each content category. Please use the red marking pen to make this label.

Task 4. Judgment About the Level of Reasoning

In this task you are to judge which of four levels of reasoning is required by each of the questions in each superitem. The purpose of this judgment task is to establish the construct validity of the questions in each superitem.

Step 1. Read the following information about the levels of reasoning.

A. Comprehension

B. Application

C. Analysis

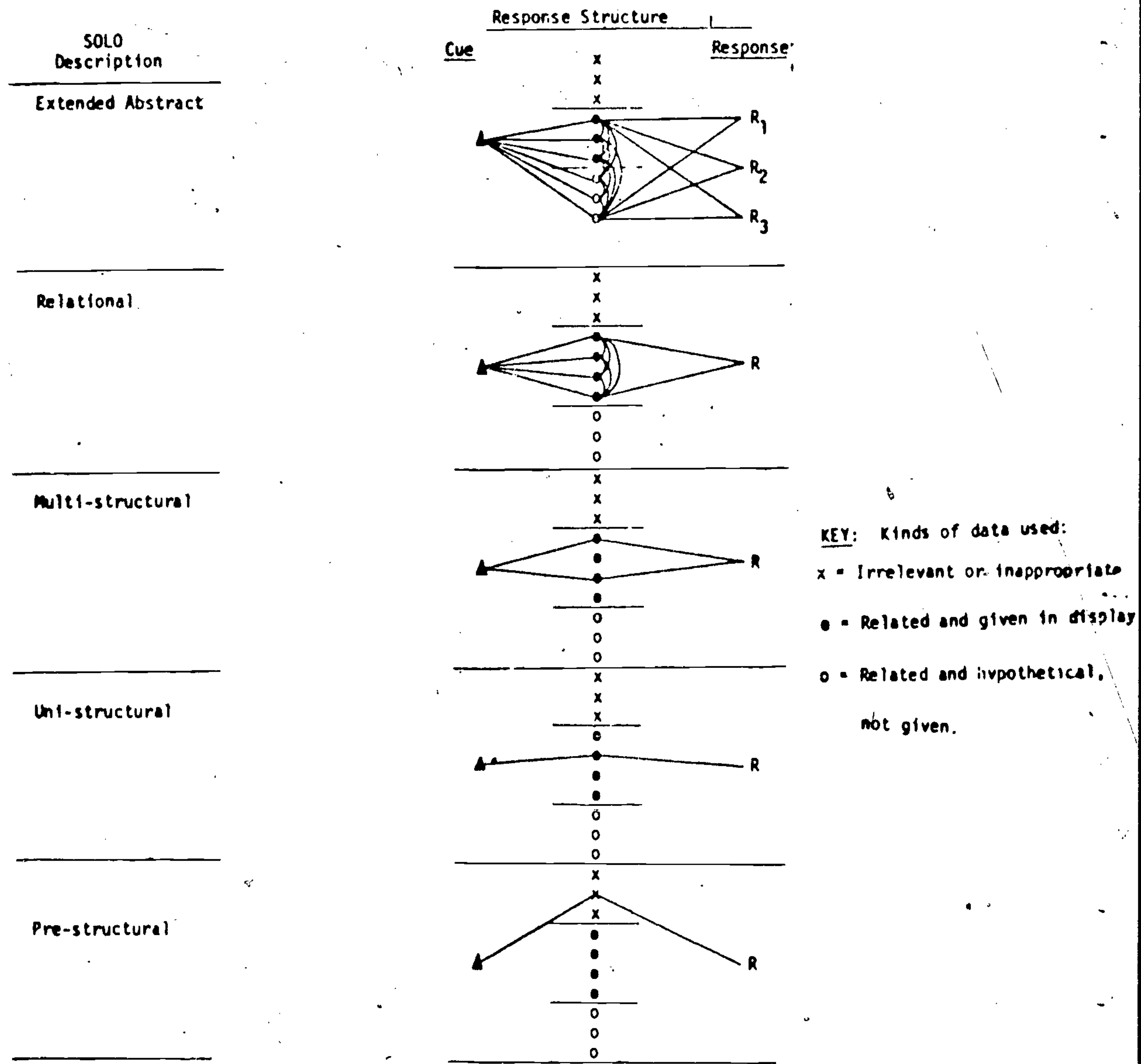
Introduction: A Response Model

The SOLO taxonomy was designed as a response model. The basic idea being that the child is given information or data and asked a question, which can be answered by reference to the information provided. The child's response is classified as belonging to one of the five levels to the way in which the response is structured.

The Categories

There are five basic categories into which responses may be classified using this method. They are set out diagrammatically below:

SOLO and Response Description




The above diagram is meant to cover the general case. For particular content areas, certain idiosyncracies peculiar to that area need to be taken into account. For mathematics the following can serve as a highly condensed summary of response models which are meaningful within the content of school based mathematical material.

Summary of Response Modes

UNI-STRUCTURAL RESPONSES

Marked by a single direct relationship to concretely (either physically or iconically) available criteria

MULTI-STRUCTURAL RESPONSES

The ability to handle multiple operations with small numbers by a series of meaningful closures, for instance, may be seen as analogous to using a sequence of given propositions to support a particular judgment in other content areas

RELATIONAL RESPONSES

The individual relates elements within the immediately available concrete system and forms generalizations on this basis

EXTENDED ABSTRACT RESPONSES

Acceptance of lack of closure, use of the reciprocal operation and ability to work with multiple interacting and abstract systems involve a comprehensive use of the given data together with related hypothetical constructs

Superitems

Cureton (1965) seems to have been responsible for coining this term to describe sets of questions which were asked about a particular problem situation. Typically the problem situation would be described in the stem which would consist of a paragraph describing the problem and the items would consist of a series of questions which could be answered by reference to the information in the stem. Cureton's basic interest was methodological but others (e.g., Wearne & Romberg, 1976) have used the notion since to develop tests of mathematical problem solving.

This latter work showed that the tests were useful because they provided more information about the student as well as a more refined measure of the child's problem-solving ability. However, they did not give information about the level of a child's reasoning in respect of each problem situation. It is this latter aspect that this project is designed to shed light on.

It is hypothesized that, by using the SOLO technique in reverse as it were, one ought to be able to design items such that a series of questions on the stem would require a more and more sophisticated use of the information in the stem in order to obtain a correct result. This increase in sophistication should parallel the increasing complexity of structure noted in the SOLO categories.

Criteria for Construction of Superitems

Clearly the construction of these items consists of two parts, the writing of the stem and the construction and writing of questions to reflect the SOLO levels. The latter requires that suitable criteria be set up to enable four questions to be asked which would not only require a knowledge of the information in the stem but would also be such that a correct response to each question would be indicative of an ability to respond to the information in the stem at least at the level reflected in the SOLO structure of the particular question. To achieve this last, the following criteria were set up for designing the questions:

1. Uni-structural: Use of one obvious piece of information coming directly from the stem.

2. Multi-structural: Use of two or more discrete closures directly related to separate pieces of information contained in the stem.

3. Relational: Use of two or more closures directly related to an integrated understanding of the information in the stem.

4. Extended Abstract: Use of an abstract general principal or hypothesis which is derived from or suggested by the information in the stem.

In each superitem the correct achievement of the first question would indicate an ability to respond to the problem concerned at at least the uni-structural level. Likewise success on a second question corresponds to an ability to respond at multi-structural level and so on.

Step 2. Judging. Starting with the first superitem in the set, study a question and determine the level of reasoning that is needed. Once you have decided what level of reasoning is appropriate, write 1, 2, 3, or 4 in red in front of the question.

Study the next question in the same superitem in the same fashion and follow the same procedures. Consider subsequent questions for each superitem.

Appendix D

TEST ADMINISTRATORS' MANUALS AND SAMPLE TEST BOOKLET FORMAT AND INSTRUCTIONS TO STUDENTS

(Note: The manuals pertain to two test formats, Booklet 1 and Booklet 2. Booklet 1 contains the superitems in the format, which is discussed in this report. The content and format of Booklet 2 will be examined in the second report in this series.)

## Grades 4, 6, 8

# Test Administrator's Manual

## Mathematics Problem-Solving Test

## Booklets 1 & 2

Forms UMR1, UMR2, UMR3, UMR4, UMR5 (Booklet 1)

Forms U, M, R, UM, UR, MR (Booklet 2)

Developed by the Wisconsin Research and Development Center for Individualized Schooling, University of Wisconsin, pursuant to a grant from the National Institute of Education, Department of Education. Grant No. FCS-02-81-20321.

# GENERAL DIRECTIONS

## Introduction

The purpose of this testing program in mathematical problem-solving is to examine the validity, reliability, usability, and item structure of 35 test problems called superitems prior to their possible use in the National Assessment of Educational Progress. Measures of related factors such as attitude toward mathematics, student background, and general verbal ability are also included.

## Materials

Each student will need:

- 1 test envelope (contains both Booklets 1 and 2)
- 2 sharpened pencils with erasers
- library book or other schoolwork for use if finished early (optional)

Each test administrator will need:

- watch or clock (stopwatch NOT necessary)
- 1 test manual
- 2 sample student booklets, Booklet 1 and Booklet 2

NOTE: There are five forms of Booklet 1 and six forms of Booklet 2.[1] The format and directions as well as both questionnaires and the Similar Words Test are identical in all forms of a booklet; only the mathematics problems differ. It is not necessary for an administrator to have all forms.

## Time Schedule

These materials are developmental; therefore, the times given below are estimates based on trial administrations. Some students will finish early. Others may need a few minutes longer.

**NOTE** It is important that the information for each student be as complete as possible. You may adjust the time schedule accordingly. (EXCEPTION: The 6-minute limit for the Similar Words Test may not be changed.)

| Booklet 1 (sitting 1) | Time Estimate |
|---|---|
| · cover, directions, sample problem | 10 minutes |
| · 7 mathematics problems (each with 3 parts) | 40 minutes |
| · mathematics questionnaire | 15 minutes |

[1] Booklet 1: Forms UMR1, UMR2, UMR3, UMR4, UMR5
Booklet 2: Forms U, M, R, UM, UR, MR

| | Time Estimate |
|---|---|
| Break | optional length |
| Booklet 2 (sitting 2) | |
| · cover, directions | 5 minutes |
| · Similar Words Test | 6 minutes (Timed) |
| · 10 mathematics problems (each with 1 or 2 parts) | 40 minutes |
| · student questionnaire | 10 minutes |

Monitoring the Testing

- mathematics problems (sitting 1 and 2)

  Some students may find all or some parts of the mathematics problems difficult. Reassure them, as necessary, indicating that they may not yet have learned how to do these kinds of problems.

  Encourage them to try every part of every problem but not to spend too long on any part. (See specific directions below.)

- mathematics test (sitting 2 only)

  For sitting 2, due to the technical demands of this test program, the six forms of the test vary in length and difficulty. You may find that some students finish considerably sooner than others. Therefore, you may wish to have students prepared with a library book, or other schoolwork, for sitting 2.

  Certain problems from sitting 1 are repeated in sitting 2. It is important that students redo these problems and not be allowed to refer back to sitting 1 (Booklet 1).

Distribution of Test Envelopes

The test booklets have been packed and the envelopes boxed in random order. Do not remove the tests from the envelopes and keep the envelopes in the order they have been boxed (as much as possible). Distribute one envelope to each student, after they have been seated (or place the envelopes on the seats prior to students' entering the room). If this procedure is followed, students sitting near each other will not have the same test forms, except by chance.

Return of Test Materials

Since these test items may be used for the National Assessment of Educational Progress, it is important that test security be maintained. All test booklets, manuals, extra envelopes, etc., must be returned to your central office.

SPECIFIC DIRECTIONS (Booklet 1/Sitting 1)

| |
|---|
| Material enclosed in boxes below is reprinted directly from student test booklets. |

SAY (or paraphrase[1]): Please do not open your envelope until I ask you to do so.

DO: If you plan to collect the envelopes between sittings 1 and 2, direct students to write their names on the envelopes. Otherwise, this is not necessary.

SAY: There are two test booklets in your envelope. Take out Booklet 1-- it has a green cover. Leave Booklet 2 in the envelope.

Complete all the information on the cover of Booklet 1. Make an X in the box next to "Boy" or "Girl." Stop when you see the STOP sign.

DO: If students do not know their year of birth, please ascertain this from the school office after the testing. Make sure students give their current age.

SAY: Open your booklet to the directions on page 2.

Read the directions silently while I read them aloud.

READ:

> DIRECTIONS
>
> This booklet contains:
>
> 1) 7 mathematics problems
> 2) a mathematics questionnaire
>
> You will not be graded on any of this work. But it is important that you answer as accurately and carefully as you can because your test results will be used for the National Assessment of Educational Progress.
>
> Each of the seven mathematics problems has Parts A, B and C. Use the extra space to do your work. Then write your answer on the line. If you don't know an answer, leave the line blank.
>
> Now please do the example. Stop when you see the Stop sign.

[1]You may paraphrase all directions as required to better communicate with your students.

DO: Allow about 3 minutes working time for the example. (The example is somewhat easier than most problems, for which about 5 minutes is suggested.)

Move about the room and see that students go right on from Part A to Parts B and C, and that they put their answers on the lines.

Do not assist them with the example.

SAY: Stop working. Turn to the next page, page 4.

READ:

> **Here are the answers for the example. Read and compare them with your answers.**

SAY: You may look back at page 3 at your answers, but do not change them or write anything else on page 3.

DO: Pause briefly while the students flip back and forth from the answers to their work.

SAY: Now look at page 4 while I explain the answers.

DO: Read through the answers and explanations <u>briskly</u>. Do not discuss the example further.

READ:

> **A. If 14 is put out, what number was put in?**
>
> **ANSWER** 4

SAY: The answer for Part A is 4 because the information in the problem said that when 14 is put out, 4 was put in.

READ:

> **B. If we put in a 5, what number will the machine put out?**
>
> $5+5+5+2=17$
>
> **OR**
>
> $(3\times5)+2=17$
>
> **ANSWER** 17

SAY: The answer for Part B is 17 because 5 added three times is 15, plus 2 more is 17. Another way to do Part B is 3 times 5 is 15 plus 2 more is 17.

READ:

> **C. If we got out a 41, what number was put in?**
>
> $41 - 2 = 39$
>
> $39 \div 3 = 13$.
>
> ANSWER 13

SAY: The answer for Part C is 13. First we subtract the 2 that was added. So, 41-2 is 39. That tells us that some number added 3 times gives 39. If we divide 39 by 3, we get 13. So the answer is 13.

Now look at the top of the next page, page 5. Read it silently as I read it aloud.

READ:

> The test begins on the next page. You will have 40 minutes to complete the seven problems. Try every part for each problem but don't spend too much time on one part. If you have time, you may go back and try any part you could not do at first.

SAY: Turn to page 6 and find problem #1.

Remember, you may use the extra space to work. Then write your answer on the line. Keep working until you see the STOP sign. Begin.

DO: Do not assist the students with the test; this includes not helping them read the problems. Keep your responses to any questions noncommital.

Allow about 40 minutes working time (estimate). IMPORTANT: Slightly over 5 minutes has been suggested for each problem. Therefore, after 5 minutes have passed, most students should have begun problem #2; after 10 minutes, most should have begun problem #3; etc. Move about the room to insure that students are working at this pace, encouraging individual students to move on as necessary. (GRADE 8 students may be instructed to pace themselves--that is, to give themselves about 5 minutes per problem. Do not ask this of GRADE 4 & 6 students.)

SAY: Stop working. Now turn to the mathematics questionnaire on page 20.

For this questionnaire you will read sentences that describe feelings about mathematics. If the sentence tells how you feel, make an X in the box under true. If the sentence does not tell how you feel, make an X in the box under false. The sentence might be mostly true for you or mostly false. Or, you might make an X under "I don't know." Mark only one box for each sentence.

I will read the questionnaire to you. Look at the top of page 20. It says . . .

DO: After the first sentence, make sure students have marked only <u>one</u> box. Check again after the second sentence. Then go on, allowing about 30 seconds per sentence (total reading and response time).

READ:

MATHEMATICS QUESTIONNAIRE

I. <u>WHAT DO YOU FEEL WHEN YOU TRY TO SOLVE A MATH PROBLEM?</u>

For each sentence please put an X in the box that best describes your feelings.

| | false | mostly false | I don't know | mostly true | true |
|---|---|---|---|---|---|
| 1. I feel there is something that keeps me from getting at the problem, a sort of fence I can't get across. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 2. I feel like I am inventing something when I am solving a problem. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 3. When I start a problem, I feel completely in the dark. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 4. If I can't find the answer, I feel defeated. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 5. When I discover a way to do a problem, I feel better. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 6. When I see a problem, I want to give up right away. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 7. If I find the answer right away, I feel satisfied. | ☐ | ☐ | ☐ | ☐ | ☐ |

II. <u>WHAT DOES DOING MATHEMATICS MEAN TO YOU?</u>

For each sentence please put an X in the box that best describes your feelings.

| | false | mostly false | I don't know | mostly true | true |
|---|---|---|---|---|---|
| 8. It means doing something basic which is the key to everything else. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 9. It does not mean anything, it is nonsense. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 10. It is doing something that you are told to do and that you have to keep doing over and over like a machine. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 11. It is doing something which I think I just can't do. | ☐ | ☐ | ☐ | ☐ | ☐ |
| 12. It is constantly discovering something new. | ☐ | ☐ | ☐ | ☐ | ☐ |

20

II. (continued)

| | false | mostly false | I don't know | mostly true | true |
|---|---|---|---|---|---|
| 13. It is doing something required, something you have to do. | | | | | |
| 14. It is a way of training my mind. | | | | | |
| 15. It is trying to find connections between different things. | | | | | |

III. WHAT DO YOU THINK OF THE FOLLOWING SENTENCES?

For each sentence please put an X in the box that best describes your feelings.

| | false | mostly false | I don't know | mostly true | true |
|---|---|---|---|---|---|
| 16. When you do mathematics, there is no room for your personality. Everything you do has been done before, everything has been planned. | | | | | |
| 17. Mathematics builds a strong personality. | | | | | |
| 18. Mathematics may sometimes cause destruction; just think of the atomic bomb. | | | | | |
| 19. Mathematics gives you the pleasure of creating something. | | | | | |
| 20. Mathematics means another world in which I feel at home. | | | | | |
| 21. Working with mathematics helps you build a well-balanced personaltiy. | | | | | |
| 22. Mathematics helps you develop good reasoning. | | | | | |
| 23. When you are deeply involved in math work, you have trouble quitting. Therefore you must not get too involved in it. | | | | | |
| 24. Doing mathematics simplifies everything too much; it takes away the beauty of things. | | | | | |
| 25. Those who do too much mathematics risk losing touch with the real world. | | | | | |

21

SAY: This is the end of the first part of the test. Please put your booklet back in the envelope.

DO: Give the students a break. (If you collect the envelopes, for your convenience make sure student names are on them.)

SPECIFIC DIRECTIONS (Booklet 2/Sitting 2)

SAY: Take out Booklet 2--it has a pink cover.

Write your name on the line. Then open your booklet to the directions on page 3.

Read page 3 silently as I read it aloud.

READ:

DIRECTIONS

This booklet contains:

1) a Similar Words Test
2) 10 mathematics problems
3) a student questionnaire

Read the directions below for the similar words test. Stop when you see the (Stop) sign.

SIMILAR WORDS TEST

In the test on the next page there are rows of words. ONE word in each row does not belong. Here is an example:

EXAMPLE. yellow green white (wet) red

The word "wet" does not belong with the others because all the other words are the names of colors. So "wet" is circled.

You will have 6 minutes to work. Do as many rows as you can. If you are not sure about an answer, circle the word you think it is.

SAY: Now turn to page 4. Do not begin working until I ask you to do so.

Remember, circle the one word in each row which does not belong. Please try every row. Begin working.

DO: This is a timed test. Allow <u>exactly</u> 6 minutes. Move about the room, making sure students try each row but do not dwell on one row.

SAY: Stop working. Look at the top of page 5. Read it silently as I read it aloud.

READ:

> There are ten mathematics problems in the next test. They are like the problems you did in Booklet 1. But for this test some students will do only Part A for every problem. Some students will do only Part B or Part C. Others will do Parts A <u>and</u> C, and so on.
>
> You might have some problems from Booklet 1 to do again; <u>please do the work for them again</u>. Do NOT look back at Booklet 1.
>
> You will have ~~30~~ 40 minutes to complete the 10 problems. Try every problem. If you have time, go back and try any part you could not do.

SAY: Now turn to the next page, page 6; find problem #1 and begin working.

(OPTIONAL: If you finish this test early, you may take out other work, or, read a library book.)

DO: Allow about 40 minutes working time (<u>estimate</u>).

As noted above, for sitting 2 the mathematics tests vary in length and difficulty. Some students, for example, have only one easy part per problem; others have two hard parts. Therefore, students may move through the ten problems at quite different speeds.

On the average, however, about 4 minutes per problem is suggested (whether 1 or 2 parts are included). After 4 minutes, most students should have begun problem #2; after 8 minutes most should be on problem 3; etc.

Some students may have problems from sitting 1 <u>repeated</u> in sitting 2; they must redo these problems without referring back to Booklet 1.

SAY: Stop working. Turn to the Student Questionnaire on page 26.

DO: GRADES 4 AND 6--read the questionnaire to the students.
GRADE 8--students may complete the questionnaire independently.

You may assist the students as necessary with this questionnaire.

READ:

STUDENT QUESTIONNAIRE

For each sentence put an X in ONE box, like this: ☒

1. Does your family get a newspaper regularly?

☐ Yes ☐ No ☐ I don't know.

2. Does your family get any magazines regularly?

☐ Yes ☐ No ☐ I don't know.

3. Are there more than 25 books in your home?

☐ Yes ☐ No ☐ I don't know.

4. Is there an encyclopedia in your home?

☐ Yes ☐ No ☐ I don't know.

5. How much school did your father complete?
(X IN THE ONE BOX which best shows how much school your father completed.)

☐ Did not complete the 8th grade
☐ Completed the 8th grade, but did not go to high school
☐ Went to high school, but did not graduate from high school
☐ Graduated from high school
☐ Some education after graduation from high school

☐ I don't know.

6. Did your father graduate from a college or university?

☐ Yes ☐ No ☐ I don't know.

26

Go right on to the next page.

7. How much school did your mother complete?
(X IN THE ONE BOX which best shows how much school your mother completed.)

- [ ] Did not complete the 8th grade
- [ ] Completed the 8th grade, but did not go to high school
- [ ] Went to high school, but did not graduate from high school
- [ ] Graduated from high school
- [ ] Some education after graduation from high school
- [ ] I don't know.

8. Did your mother graduate from a college or university?

[ ] Yes [ ] No [ ] I don't know.

9. The metric system uses units like centimeters, liters, and kilograms. Have you used the metric system of measurement?

[ ] Yes [ ] No [ ] I don't know.

10. How often have you used the metric system in mathematics?

[ ] Often [ ] Sometimes [ ] Never [ ] I don't know.

11. How often have you used a hand calculator?

[ ] Often [ ] Sometimes [ ] Never [ ] I don't know.

12. Do you or your family own a hand calculator?

[ ] Yes [ ] No [ ] I don't know.

13. Does your school have hand calculators that you can use in mathematics class?

[ ] Yes [ ] No [ ] I don't know.

Please put your booklet back in the envelope.
Thank you for your help. 27

SAY: This is the end of the test. Put your booklet back in the envelope.

FORM UMR1

# Booklet 1 Mathematics Problem-Solving Test

Please fill in the information below.

NAME ____________________ ____________________
first last

AGE __________

DATE OF BIRTH __________ __________ __________
month day year

GRADE __________

SCHOOL ________________________________________

BOY ☐ GIRL ☐

STOP Do not open the booklet until told to do so.

Developed by the Wisconsin Research and Development Center for Individualized Schooling, University of Wisconsin, pursuant to a grant from the National Institute of Education, Department of Education. Grant No. ECS-02-81-20321.

DIRECTIONS

This booklet contains:

1) 7 mathematics problems
2) a mathematics questionnaire

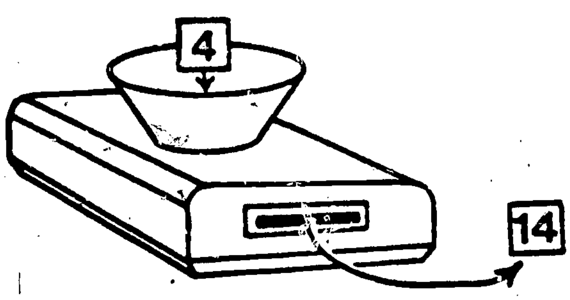
You will not be graded on any of this work. But it is important that you answer as accurately and carefully as you can because your test results will be used for the National Assessment of Educational Progress.

Each of the seven mathematics problems has Parts A, B and C. Use the extra space to do your work. Then write your answer on the line. If you don't know an answer, leave the line blank.

Now please do the example. Stop when you see the Stop sign.

EXAMPLE

**EX. This is a machine that changes numbers. It adds the number you put in three times and then adds 2 more. So, if you put in 4, it puts out 14.**




**A. If 14 is put out, what number was put in?**

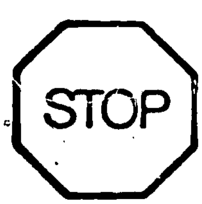
**ANSWER** ____________

B. If we put in a 5, what number will the machine put out?

ANSWER ____________________

C. If we got out a 41, what number was put in?

ANSWER ____________________




Do not go on to the next page until told to do so.

Here are the answers for the example. Read and compare them with your answers.

A. If 14 is put out, what number was put in?

ANSWER 4

B. If we put in a 5, what number will the machine put out?

$5+5+5+2=17$

OR

$(3\times5)+2=17$

ANSWER 17

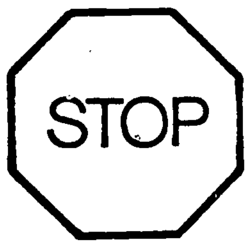
C. If we got out a 41, what number was put in?

$41-2=39$

$39\div3=13$

ANSWER 13

The test begins on the next page. You will have 40 minutes to complete the seven problems. Try every part for each problem but don't spend too much time on one part. If you have time, you may go back and try any part you could not do at first.




Do not go on to the next page until told to do so.

# Grade 12

# Test Administrator's Manual

## Mathematics Problem-Solving Test

## Booklets 1 & 2

Forms S1, S2, S3, S4, S5 (Booklet 1)

Forms M, R, E, MR, ME, RE (Booklet 2)

Developed by the Wisconsin Research and Development Center for Individualized Schooling, University of Wisconsin, pursuant to a grant from the National Institute of Education, Department of Education. Grant No. ECS-02-81-20321.

# GENERAL DIRECTIONS

## Introduction

The purpose of this testing program in mathematical problem-solving is to examine the validity, reliability, usability, and item structure of 35 test problems called superitems prior to their possible use in the National Assessment of Educational Progress. Measures of related factors such as attitude toward mathematics, student background, and general verbal ability are also included.

## Materials

Each student will need:

- 1 test envelope (contains both Booklets 1 and 2 and one sharpened pencil)
- library book or other schoolwork for use if finished early (optional)

Each test administrator will need:

- watch or clock (stopwatch NOT necessary
- 1 test manual
- 2 sample student booklets, Booklet 1 and Booklet 2

NOTE: There are five forms of Booklet 1 and six forms of Booklet 2.[1] The format and directions as well as both questionnaires and the Similar Words Test are identical in all forms of a booklet; only the mathematics problems differ. It is not necessary for an administrator to have all forms.

## Time Schedule

These materials are developmental; therefore, the times given below are estimates based on trial administrations. Some students will finish early. Others may need a few minutes longer.

**NOTE** It is important that the information for each student be as complete as possible. You may adjust the time schedule accordingly. (EXCEPTION: The 6-minute limit for the Similar Words Test may not be changed.)

| Booklet 1 (sitting 1) | Time Estimate |
|---|---|
| · cover, directions, sample problem | 10 minutes |
| · 7 mathematics problems (each with 3 parts) | 40 minutes |
| · mathematics questionnaire | 10 minutes |

[1]Booklet 1: Forms S1, S2, S3, S4, S5
Booklet 2: Forms M, R, E, MR, ME, RE

| | Time Estimate |
|---|---|
| Break | optional length |
| Booklet 2 (sitting 2) | |
| · cover, directions | 5 minutes |
| · Similar Words Test | 6 minutes (Timed) |
| · 10 mathematics problems (each with 1 or 2 parts) | 40 minutes |
| · student questionnaire | 10 minutes |

Distribution of Test Envelopes

The test booklets have been packed and the envelopes boxed in random order. Do not remove the tests from the envelopes and keep the envelopes in the order they have been boxed (as much as possible). Distribute one envelope to each student, after they have been seated (or place the envelopes on the seats prior to students' entering the room). If this procedure is followed, students sitting near each other will not have the same test forms, except by chance.

Return of Test Materials

Since these test items may be used for the National Assessment of Educational Progress, it is important that test security be maintained. All test booklets, manuals, extra envelopes, etc., must be returned to your central office.

## SPECIFIC DIRECTIONS (Booklet 1/Sitting 1)

SAY (or paraphrase[1]): Please do not open your envelope until asked to do so.

DO: Distribute the envelopes.

If you need to collect the envelopes between sittings 1 and 2, direct the students to write their names on the envelopes. Otherwise, this is not necessary.

| · cover, directions, sample problem | 10 minutes |
|---|---|

SAY: There are two test booklets in your envelope. Take out Booklet 1 (blue cover). Leave Booklet 2 (yellow cover) in the envelope.

(Pause.)

Please complete all the information on the cover of Booklet 1.

(Pause.)

Now read the directions on page 2 and complete the sample problem on pages 2 and 3. Then go right on to pages 4 and 5.

DO: Allow about 5 minutes. Do not assist students with the example.

| · 7 mathematics problems | 40 minutes |
|---|---|

SAY: You have 40 minutes for the seven problems, so you should pace yourself and spend no more than 5 or 6 minutes per problem. Remember, use the extra space to do your calculations, and then put your answer on the line. Begin working.

DO: Allow about 40 minutes working time (estimate). It is more important that complete information for every problem and every student be attained than it is to adhere strictly to this time estimate.

NOTE: Slightly over 5 minutes have been suggested for each problem. Therefore, after 5 minutes have passed, most students should have begun problem #2; after 10 minutes, most should have begun problem #3; etc. Monitor the test to insure that students are working at this pace; encourage individual students to move on as necessary.

Do not assist the students with the test; this includes not helping them read the problems. Keep your responses to any questions noncommital.

[1] You may paraphrase all directions as required to better communicate with your students.

SAY: Stop working.

| | |
|---|---|
| · mathematics questionnaire | 10 minutes |

SAY: Now turn to the mathematics questionnaire on pages 20-21.

You have about 10 minutes to complete this questionnaire which asks you to describe your feelings about mathematics. Please answer how _you_ feel. Mark only _one_ box for each sentence. Begin working.

DO: Allow about 10 minutes for the questionnaire.

SAY: This is the end of the first part of the test. Please put Booklet 1 back in the envelope.

DO. Give students a break. (If you collect the envelopes between sittings, make sure names are on them.)

## SPECIFIC DIRECTIONS (Booklet 2/Sitting 2)

| · cover, directions | 5 minutes |
|---|---|

SAY: Take out Booklet 2 (yellow cover). Leave Booklet 1 in the envelope.

(Pause.)

Write your name on the line. Then read the directions on page 3. Please do not go on to page 4.

(Pause.)

| · Similar Words Test | 6 minutes (Timed) |
|---|---|

SAY: Now turn to page 4 and begin working.

DO: This is a timed test. Allow exactly 6 minutes.

SAY: Stop working.

| · 10 mathematics problems | 40 minutes |
|---|---|

SAY: Read the directions on page 5.

(Pause.)

You will be asked to do some parts of some problems from Booklet 1 again. These repeated problems are a special part of this testing program. Do them again and do not refer back to Booklet 1.

There is an error in your booklet, which says you have 30 minutes. You have about 40 minutes, so you should spend no more than 4 minutes on one problem. If you finish early, go right on to the questionnaire.

DO: Allow about 40 minutes working time (estimate).

Students will have some problems from sitting 1 repeated in sitting 2; they must redo these problems without referring back to Booklet 1.

For sitting 2 the 6 forms of the mathematics test vary in length and difficulty. Some students, for example, have only one easy part per problem; others have two hard parts. Therefore, students may move through the ten problems at quite different speeds.

On the average, however, about 4 minutes per problem is suggested (whether 1 or 2 parts are included). After 4 minutes, most students should have begun problem #2; after 8 minutes most should be on problem #3; etc.

SAY: Stop working.

| | |
|---|---|
| • student questionnaire | 10 minutes |

SAY: If you have not begun the questionnaire on page 26, please do so.

DO: Allow about 10 minutes.

SAY: This is the end of the test. Be sure you put both booklets back in the envelope. You may keep the pencil if you wish.

26

FORM S1

# Booklet 1
# Mathematics
# Problem-Solving Test

Please fill in the information below.

NAME ______________________ ______________________
first last

AGE ________ ________ ________

DATE OF BIRTH ______________ ______________ ______________
month day year

GRADE ________

SCHOOL ______________________________________

MALE ☐ FEMALE ☐

STOP Do not open the booklet until told to do so.



Developed by the Wisconsin Research and Development Center for Individualized Schooling, University of Wisconsin, pursuant to a grant from the National Institute of Education, Department of Education. Grant No. ECS-02-81-20321.

DIRECTIONS

This booklet contains:

1) 7 mathematics problems
2) a mathematics questionnaire

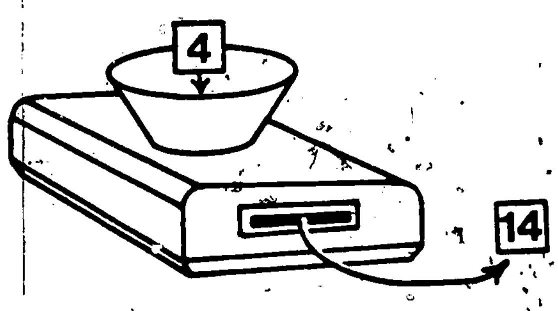
You will not be graded on any of this work. But it is important that you answer as accurately and carefully as you can because your test results will be used for the National Assessment of Educational Progress.

Each of the seven mathematics problems has Parts A, B, C and D. Use the extra space to do your work. Then write your answer on the line. If you don't know an answer, leave the line blank.

Now please do the example. Stop when you see the (Stop) sign.

EXAMPLE

EX. **This is a machine that changes numbers. It adds the number you put in three times and then adds 2 more. So, if you put in 4, it puts out 14.**




**A. If 14 is put out, what number was put in?**

**ANSWER** ____________

B. If we put in a 5, what number will the machine put out?

ANSWER ____________

C. If we got out a 41, what number was put in?

ANSWER ____________

D. If x is the number that comes out of the machine, when the number y is put in, write down a formula which will give us the value of y whatever the value of x.

ANSWER ____________



Go right on to the next page.

Here are the answers for the example. Read and compare them with your answers.

**A. If 14 is put out, what number was put in?**

**ANSWER** 4

B. If we put in a 5, what number will the machine put out?

$5+5+5+2=17$

OR

$(3\times5)+2=17$

ANSWER 17

C. If we got out a 41, what number was put in?

$41-2=39$

$39\div3=13$

ANSWER 13

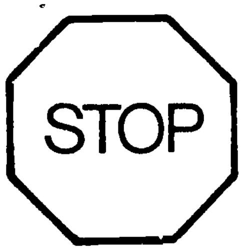
D. If x is the number that comes out of the machine, when the number y is put in, write down a formula which will give us the value of y whatever the value of x.

$$3y+2=X$$
$$3y=X-2$$
$$Y=\frac{X-2}{3}$$

ANSWER $Y=\frac{X-2}{3}$

Note: The formula x = 3y + 2 is <u>not</u> a correct answer because you were asked to give a formula for y in terms of x.

The test begins on the next page. You will have 40 minutes to complete the seven problems. Try every part for each problem but don't spend too much time on one part. If you have time, you may go back and try any part you could not do at first.




Do not go on to the next page until told to do so.