DOCUMENT RESUME

ED 135 616 SE 021 985

AUTHOR	Ruud, Orville George
TITLE	The Construction of an Instrument to Measure Proportional Reasoning Ability of Junior High Pupils.
PUB DATE	Dec 76
NOTE	280p.; Ph.D. Dissertation, University of Minnesota; Not available in hard copy due to marginal legibility of original document
EDRS PRICE	MF-$0.83 Plus Postage. HC Not Available from EDRS.
DESCRIPTORS	*Cognitive Development; Developmental Tasks; Doctoral Theses; *Educational Research; Learning Theories; Measurement Instruments; *Physical Sciences; Science Education; Secondary Education; *Secondary School Science; *Tests
IDENTIFIERS	*Piaget (Jean); Research Reports

ABSTRACT

The purpose of this study was to develop a paper-pencil test of Piagetian levels of proportional thinking for junior high school students in the context of physical science. Two thousand twenty-seven students were tested to develop the instrument and the description of its characteristics. The final form consisted of 24 items with four subtests each of six items for Piagetian levels: Concrete Operational I, Concrete Operational II, Formal Operational I, and Formal Operational II. Piagetian task interviews were also given to a group of students, and the paper-pencil test results correlated positively with the task results of the students who took both tests. Content, concurrent construct, divergent, and convergent validity measurements showed the paper-pencil test to be valid. The test was also shown to have a high reliability and good item discrimination between proportional reasoning levels. (MH)

THE CONSTRUCTION OF AN INSTRUMENT TO MEASURE PROPORTIONAL REASONING ABILITY OF JUNIOR HIGH PUPILS

A Thesis

Submitted to the Faculty of the Graduate School

of the University of Minnesota

by

Orville George Ruud

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

December, 1976

CONTENTS

CHAPTER Page

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

APPENDIX

CHAPTER 1

THE PROBLEM

Introduction

The purpose of this study was to develop a paper-pencil test of Piagetian levels of proportional thinking of junior high school pupils in the context of physical science. This seemed to be a desirable goal for several reasons:

1. The junior high pupil's proportional reasoning ability is of special interest. The age of thirteen, as Inhelder and Piaget (1958) showed, is the common age for transition to formal thought levels in proportional reasoning.

2. Present science curricula in the junior high school include such content as density, quantitative relationships of chemical reactions, genetic ratios and the dynamic relationships between force, mass and acceleration. The establishment of the level of proportional reasoning ability of a class of pupils would provide a basis for the selection of appropriate curriculum content.

3. Instructional materials and instructional strategies used by junior high science teachers are intended to develop, among other outcomes, cognitive reasoning. Pre- and post-measures of proportional reasoning levels would direct the choice and design of appropriate materials and strategies of instruction.

4. Existing paper-pencil tests do not measure the level of proportional reasoning attained by the subjects. Mathematics tests whose subtests purport to measure competency in using ratio and proportion do so through seeking one correct answer. The other answers available for selection do not have a logical basis and make no contribution to determining the subject's level of proportional reasoning in the Piagetian sense.

5. Task interviews provide an intensive measure of a limited population and are important as research tools. A typical interview requires about 20 to 40 minutes and establishes a proportional reasoning level for one person in one type of content. They are not, therefore, practically applicable for use with the large numbers of pupils with whom teachers meet.

6. Experience and techniques used in designing a paper-pencil test from task interviews in proportional reasoning should be applicable to other such test design. Rigorous application of the principles of criterion-referenced test design has not been frequently accomplished.

Statement of the Problem

Hypothesis and Task of Study

It was hypothesized in this study that proportional reasoning in physical science may be measured by appropriate criterion-referenced paper-pencil testing and that these criterion-referenced paper-pencil tests would provide the same kind and amount of information that could be obtained through the use of other modes of examination.

The task of this study was to develop a set of paper-pencil items to assess the Piagetian proportional reasoning level of pupils. The test to be developed should have these characteristics: 1) Require a 30-minute testing session. 2) Allow for the measurement of large numbers of persons. 3) Use items with different science content. 4) Have the reliability offered by several measures of the same person. 5) Require no expertise of the test administrator. 6) Be usable as a source of information for determining the numbers of pupils at the various proportional reasoning levels and which pupils are at each of these levels.

Definitions

Proportions, for the purpose of this study, are "two ratios that are equivalent" (Copeland, 1974, p. 160).

Proportional reasoning levels, for the purpose of this study, were the levels used by Inhelder and Piaget (1958). They are listed here in ascending order of complexity, with a description of the kind of proportional reasoning pupils might use.

Preoperational — Subject guesses or makes no ordered connection between things which change.

Concrete I Operational — Subject compensates in some qualitative way and may match direct ordered relations.

$$\begin{array}{ccccccc} A & < & B & < & C & < & D \\ \cdot & & \cdot & & \cdot & & \cdot \\ J & < & K & < & L & < & M \end{array}$$

Concrete II Operational — Subject uses a rule, usually addition, to calculate increase or decrease and may order corresponding relations with inverse.

$$\begin{array}{ccccccc} & < & B & < & C & < & D \\ \cdot & & \cdot & & \cdot & & \cdot \\ J & > & K & > & L & > & \end{array}$$

Formal I Operational — Subject calculates by multiplying or using simple ratios, contrasts ratios and can order them. $5/25 > 2/25$

Formal II Operational — Subject uses proportions and recognizes the appropriate proportion to be used. $A/B = C/D$ or $A/B = C/D = E/F$. Subject will seek and refer to a general rule linking the relationship.

Criterion-referenced testing, for the purpose of this study, is a testing referenced to the criteria of the discrete levels of proportional thinking. Item design and item selection techniques are those of good criterion testing technique.

Performance criteria, for the purpose of this study, is the level of performance which identified the behavior characteristics of a person achieving the level, a master, from a person not achieving the level, a non-master. Potential masters and potential non-masters were identified by reason of maturity or measurement. Grade 11 science pupils were supposed, generally, to be masters of formal proportional reasoning while grade 5 pupils were supposed, generally, to be non-masters. Piaget and others in the field suggest that most pupils would achieve formal proportional reasoning only after reaching age thirteen. The performance criteria of each proportional reasoning level for task interview performance were derived from Piaget's descriptions. Performance criteria for paper-pencil performance were set at success on two-thirds of the items for that level as discussed in Chapter 5.

Basic Design

This study was conducted in three steps or phases: an initial trial or pilot phase, an intensive task testing phase with 40 pupils to produce an initial item design, and an extensive paper-pencil testing phase with groups that in some cases exceeded 300 pupils from which the final item set was written.

Phase I - Pilot Study

In the pilot study the writer sought to assess whether it might be possible to identify proportional reasoning levels in the pupils and to measure them with paper-pencil items.

Individual interview tasks were administered to a group of pupils and different proportional reasoning levels were discerned among the pupils. Paper and pencil items derived from the tasks were later administered to the same pupils. It was found to be possible with tasks to identify the different levels of proportional reasoning to which the pupils had developed. These proportional reasoning levels were found to be measurable with paper-pencil items.

Phase II - Task Interview Testing

In this phase the writer sought to measure proportional reasoning levels of a sample of pupils by interview tasks and to use this measure to validate and select an initial set of paper-pencil items.

Forty pupils were selected by stratifying all the grade eight pupils of a school according to their Lorge-Thorndike total score and [illegible] choosing pupils randomly within IQ score levels to ensure a range of proportional reasoning ability. Extensive individual task testing on this sample was carried out with rigorously defined tasks. Paper-pencil items were carefully derived from the original tasks, written to four levels of proportional thinking, and administered to the pupils. From the results of this paper-pencil testing an initial set of items was chosen for use in Phase III.

Phase III - Paper-Pencil Testing

In the final phase the writer sought to produce a paper-pencil test with an administration time of approximately 30 minutes that would measure proportional reasoning levels of eighth grade pupils.

The initial item set was used with large populations of grade eight pupils. The item responses were analyzed for their ability to discriminate between proportional reasoning levels. Items were revised or replaced and the test was administered again. Populations of masters, senior high science pupils, and of non-masters, grade five students, were also used. Ten versions of the test were used. The validity and reliability of the final version were measured.

CHAPTER 2

SURVEY OF RELATED RESEARCH LITERATURE

Because this study was concerned with the development of an instrument for large scale measure of proportional reasoning ability [illegible] high pupils, three types of literature were pertinent to the study: 1) studies of the formal stages of intellectual growth of pupils, 2) studies of proportional thinking, and 3) studies of measurement with criterion referenced testing.

There is general discourse concerning Piaget's research and there are scholarly statements of explanation like those of Darley and Anderson (1951), Jensen (1973), Wood (1974), Beistel (1975), Herron (1975), and Mallon (1976) where postulates, guidelines and suggested instructional strategies are proposed for general science teaching and where the problems of proportional reasoning are discussed. Such discourse and statements are not reviewed in this chapter because of their lack of research information. Expert statements and procedural recommendations in the literature on criterion testing are reviewed because of their interest to criterion test design.

Proportional thinking was classified by Inhelder and Piaget (1958) as a formal operational level ability. The studies of formal operational stages are thus of concern. A proportion is defined by Mandell (1974) as "a statement of equality of two ratios." Studies of pupil operations with ratios as well as with proportions are reviewed. A criterion-referenced test as viewed by Glaser and Nitko (1971) is a test that is deliberately constructed to yield measurements that are interpreted in terms of performance standards. Criterion-referenced testing is concerned with the measurement of individual and group performance in relationship to established criteria. Professional statements and studies here dealing with the design of criterion-referenced tests are important to the study.

Studies of Formal Operations

Original Studies

The description of formal operational thought originated with Piaget (1926). Specific attention to proportional reasoning appeared later.

In The Growth of Logical Thinking, Inhelder and Piaget (1958) described the study of intellectual stages of growth of persons from five to fifteen years in age. The subjects were individually given task interviews. Fifteen such separate investigations were conducted. Discernible levels of concrete and of formal thought were reported for each investigation. Piaget (1972) noted that individuals performing different tasks do exhibit different levels of thought. He suggested that the formal operation tasks should be such that for subjects the situations should involve equal aptitudes or comparable interests.

Piaget and Inhelder (1969) identified the emergence of proportional reasoning with the ages of eleven or twelve. Piaget (1972) described the formal stage as being related to verbal capacities and characterized the formal stage as a stage where the capacity to reason in terms of verbally stated hypothesis appeared. Piaget (1972) described the stages as resulting in a certain number of overall structures which became necessary with development. An important problem he noted was the time lag between solution of problems in different areas. He reported that at certain ages changing the material or situation used in testing gave different test results. Piaget (1964) identified maturation, experience, social transmission and equilibration as factors which explain the person's development from one set of structures to another. Such development he saw as interaction with things. Knowing an object meant acting on it, modifying it and transforming it. It also involves interaction with thought. This thought interaction is the essence of equilibration. Smeslund (1964) explained that the difference between learning and equilibration is the difference between the interaction of thought with things and the interaction of thought with itself.

In summary, Piaget and his colleagues identified a formal stage of proportional reasoning ability emerging in early adolescence. This stage should be discernible in the child's ability to deal with spatial proportions, inertial speeds, probabilities and related concepts in a verbal manner. Performance of the early adolescent in proportional reasoning should depend upon the content of the problem and the child's experience.

Replications of Original Studies

Lovell (1961) repeated ten of the experiments described by Inhelder and Piaget (1958) with 200 British pupils between the ages of eight and eighteen. Lovell found that his results confirmed the main stages in the development of logical thinking proposed by Inhelder and Piaget. Lovell suggested that few junior high pupils reach the level of formal thought. He reported that the least able students remain at a low level of thought. Some fifteen-year-olds were found not to be at the first level of formal thought.

Elkind (1961, 1962) used junior high, senior high and college pupils respectively in a series of replication task interviews in the conservation of volume, mass and density. Elkind confirmed Piaget's finding of a regular age-related order in the conservation of mass, weight and volume, but did not agree on acquisition of an abstract concept of volume by eleven- or twelve-year-olds. He found only about 60 per cent of college freshmen tested believed that the volume of a ball of clay remained constant when the clay was rolled out into a sausage form.

Jackson (1965) studied logical thinking in normal and subnormal children. He used six of the experiments of Inhelder and Piaget with 48 British children with an IQ range 90 to 100, and 40 British children with an IQ range 60 to 80. Jackson reported that the subnormal children showed only limited increase in intellectual development beyond age nine, while the normal ones displayed levels of thinking which generally confirmed the age level statements of Piaget.

DeVries (1973b) used Piagetian tasks to compare the performance of children classed as bright, average and retarded. She asked two questions: with children of the same chronological age, do higher IQ children perform better and with children of the same mental age, do higher IQ pupils perform better? She reasoned that if the answer to both questions is yes, then Piaget tasks measure some type of intelligence. In the results, higher IQ pupils outperformed others of the same chronological age but older children (lower IQ) outperformed others of the same mental age.

Dale (1970) replicated Inhelder and Piaget's first chemistry experiments using 200 Australian children from six to sixteen years old. His findings did support the basic structure of Piaget's theory of development of logical thinking with age and more specifically, the development of combinatorial thinking with age.

Towler and Wheatley (1971) replicated Piaget and Elkind conservation tasks with college pupils. In the 71 female subjects studied at Purdue University, Towler and Wheatley found nearly identical, 61 per cent versus 58 per cent, acceptable responses.

Holloway (1967) reported that the child's conception of geometry was realted to his/her intellectual development level. He noted that at the formal operational stage the logic principle A = B, B = C therefore A = C appears.

Keasy (1971) studied formal operational thinking using three age groups: sixth grade girls, college women and fifty-year-old women. Five of the experiments described by Inhelder and Piaget (1958) were used. Results showed the girls to be at the lowest level, fifty-year-old women were intermediate and the college women at the top. Consistency between age groups was reported. Very few attained the formal operational level.

Bart (1971), Lovell and Butterworth (1966), and Lovell and Shields (1967), using Piaget tasks, substantiated that formal operational skills have a large general factor. All researchers used a principal components analysis to analyze the task performance of pupils. Bart, in his study, administered four Piagetian formal thought tests, three formal operational reasoning tests and a test of verbal intelligence to 90 scholastically above average pupils. He also established that formal thought, as measured by Piaget's tasks, has a substantial verbal intelligence component as well as a nonverbal intelligence component.

McKinnon and Renner (1971), using adaptations of Piaget tasks, found that 50 per cent of college freshmen tested were functioning completely at Piaget's concrete operational level and only 25 per cent of their sample could be considered fully formal in their thought.

Every replication study cited supported Piaget's model of an ordinal sequence of development. Generally, replication study results showed the stages of development came at later ages than those reported by Piaget and Inhelder. This observation was also that of Howe (1974) who reviewed the literature to determine the extent of evidence to support the concept of formal thought. She found the bulk of the evidence seemed to support that there is a qualitative change in cognitive structure or reasoning ability beyond the level of concrete operations, no dependence on the use of all the binary operations of propositional logic in the new structure and more than one process involved in the development of logical thinking beyond the concrete level.

Related Studies

Studies reported here are related to Piaget's work with formal operational thought. However, these studies are different in that they used different techniques for measurement, used batteries of several tasks or investigated relationships between task performance and other pupil characteristics. The general studies of cognitive development which were reviewed produced results that confirmed Piaget levels of development with different testing techniques. Linn and Thier (1975) used a filmed testing sequence to measure logical thinking.

Open questioning was the strategy used by Laurendeau and Pinard (1962). In such questioning, the wording of the question was changed when necessary using terms more familiar to the child, but with care never to suggest more than was included in the instructions.

Karplus and Karplus (1970) used a group presentation with elementary school pupils, junior high school pupils, senior high school pupils, science teachers and physicists of an Islands Puzzle and including introduction of new topics in concrete terms, pupil evaluation of an unsatisfactory hypothesis and creation of discrepant events, requiring reasoning by contradiction. This strategy could be described as midway between the individual task and the group paper-pencil tests. An oral description of the task was given. The subjects responded in writing.

Batteries of Tasks

The use of batteries of several tasks showed that different tasks gave different results (Osiki, 1974; D. R. Phillips, 1974; Karplus, Karplus and Wollman, 1974; Lawson, Nordland and DeVito, 1975). High correlations between tasks were rarely reported. Lawson, Nordland and DeVito (1975) found intercorrelations ranging from .02 to .55. Almy (1970) reported .32 as the highest intercorrelation among a set of tasks. The composite score of such a set of tasks was seen as the best predictor by Sayre and Ball (1975) and Lawson, Nordland and DeVito (1975). In some cases one or two of the tasks alone were found to be better predictors than the entire battery (Lawson and Renner, 1975).

Wohlwill (1960) used a scalogram analysis of Green (1956) to determine the scalability and homogeneity of a set of measured tasks. He determined that tasks had varying difficulties.

Correlational Studies

The studies of Wohlwill (1960), Osiki (1974), D. R. Phillips (1974), Lawson, et al. (1975) and Sayre and Ball (1975) previously described as studies using task batteries were also invest'gacions of the relationships between task performance and other pupil characteristics.

Ball and Sayre (1972) investigated the relationship between pupil Piagetian cognitive development and achievement in science. They contrasted the grades 419 science pupils received with their level of cognitive development as measured by five abstract tasks, and concluded that pupils are being penalized, by receiving lower grades, for not being able to think at the formal operational level.

Higgins and Gaite (1971) studied adolescent mode of thinking on Elkind (1961) conservation tasks in contrast with thinking on a task simulating a familiar real life situation. They found that in the 162 pupils, ages thirteen,to eighteen, successful completion of the conservation tasks and the situation task were independent. A significant positive correlation was established between the mean age of the group and the number who used abstract thinking. No significant positive correlation was found between mean age and successful completion of the Elkind task.

Raven (1972), in a study of concept development in 160 kindergarten, grade one, grade two and grade three pupils, found that task performance was dependent upon the: 1) inference pattern of the task, 2) goal objects of the task, and 3) percepts of the task.

The generalization that Piagetian cognitive level is positively related to achievement was supported by correlational studies. Concrete and formal levels as measured by tasks correlated with the abstract performance level in tests of dogmatism (D. G. Phillips, 1974), achievement in science (Ball and Sayre, 1972; Bridgham, 1969; Sayre and Ball, 1975), achievement on commonly used achievement examinations (Lawson, Nordland and DeVito, 1975; Osiki, 1974), learning of formal concepts in science (Lawson, 1973).

Developmental Studies

A developmental sequence of levels and their scalability was established directly by Wohlwill (1960) who used a scalogram analysis to analyze a set of measured tasks. Studies not utilizing Piaget tasks or adaptations of them have also supported the developmental sequence of levels postulated by Piaget. Nisbet (1964) reported that those adolescents in England who had attained puberty scored higher on intellectual and academic achievement tests than those youngsters who were still at the puberty stage of development. Carpenter, et al. (1975a) reported that in the National Assessment of Educational Progress only 44 per cent of nine-year-olds correctly identified that a 2x8 rectangle had the same area as a 4x4 square. Almost as many of them chose a 3x5 rectangle as having the area of the 2x8 rectangle. It would appear that proportional reasoning was required here and that the reported success is comparable to that found by researchers investigating proportional reasoning. Meyers (1970) illustrated in a collection of questions showing the nature of the math content of the SAT test, that an item dealing with proportional measurement would be answered correctly by 32 per cent of the population taking that test. Reichard. Scheiden and Rapaport (1944), using sorting tasks that were not those of Piaget, found three levels of development. At the most concrete level, up to five or six years, children classified objects on the basis of nonessential incidental features. A functional level, where classification was made on the basis of use, extended to the age of eight, and the abstract level was not much used before the age of ten.

Kohlberg and Gilligan (1971), in describing their observations of the moral development of adolescents, suggested that in moral development one stage of formal operations is reached at age ten to thirteen years and the more complete stage at around fifteen to sixteen.

Studies of Proportional Thinking

Original Studies

A special concern of this study was the nature of proportional thinking as one attribute of the formal operational level of thought.

Proportional thinking was described as one attribute of the formal operational level of cognitive development by Inhelder and Piaget (1958). Their task interviews to test proportional thinking included the simple balance, a cart on an inclined plane, the projection of shadows and a spinning disc testing centripetal force. They commented that they were able to repeatedly observe that proportional reasoning was not acquired until pupils were at the formal operational level of cognitive development.

Proportional reasoning had been investigated by Piaget previously in the areas of space, speed and probability in which it was concluded that the age for such proportional reasoning and for formal operational thought was twelve to fourteen years.

Replication of Original Studies

A collection of research studies replicated the original research of Piaget in proportional reasoning. These studies affirmed the existence of stages and the scalability of proportional reasoning tasks, described the schema of proportional reasoning, tested new measurement approaches and explored correlations between proportional reasoning and other pupil characteristics. The studies generally found proportional reasoning being acquired at older ages than Piaget reported.

Lunzer and Pumfrey (1966) used tasks they designed involving such things as matching lengths of cuisenaire rods, pantograph, beam balance and similarity judgments of objects. They reported that they found that proportional reasoning, unaccompanied by physical actions was rarely used by average subjects below the age of fifteen and that younger children solved some of the tasks by successive addition.

Wollman and Karplus (1974) investigated intellectual development beyond elementary school, with 450 seventh and eighth grade pupils in Orinda, California. They studied children's use of ratio in solving beam balance, proportional length, proportionate size of shadows and pulley turning rate tasks. All tasks were designed by the authors. They concluded that to test proportional thinking, tasks would have to be devised that would apply the ratio concept in familiar situations.

As reported by Steffe and Parr (1968), Lunzer (1965) studied the relationships of developmental thinking with logical proportion (verbal analogies) and with mathematical proportion (metric equivalent ratio pairs). Lunzer's measurements of the difficulties of these two types of tasks for subjects from nine to seventeen years confirmed that numerical proportions and verbal analogies did require formal level thinking.

Steffe and Parr (1968) studied the development of the concepts of ratio and fraction in fourth, fifth, and sixth grades of elementary school. IQ measures were used to designate a high, middle and low group of pupils at each grade. An ability-stratified sample of pupils was chosen. Six paper-pencil tests were used, four on a pictorial level and two on a symbolic level. They reported that there was little correlation between the ability of children to perform successfully in proportionality situations at a symbolic level such as $6/15 = \square/5$, and their ability to perform successfully on proportionality situations based on ratio or fractional pictorial data. Also, whenever the pictorial data, which displayed the proportionalities, were not conducive to solution by visual inspection, the proportionalities were difficult for fourth, fifth, and sixth grade children to solve.

Shepler (1969) studied teachability of probability understandings. The subjects were pupils chosen from a population of 67 sixth grade pupils. All were volunteers and were above average ability. In a pretest task post test approach they did acquire probability concepts.

Hensley (1974) studied proportional thinking in children from grades six through twelve. Fifteen female and fifteen male pupils from each of the sixth, eighth, tenth and twelfth grades were tested with four tasks: beads, inclined plane, switches, projection of shadows. Hensley's results generally support the findings of Piaget. He reported a scalability of levels of proportional thinking, a positive relationship between grade level and task scores. No relationship was found, however, between sex and task scores. No correlation between tasks were calculated. No validity or reliability measures of tasks were reported.

Kavanaugh (1974) generally confirmed the theories of Piaget in the development of the concept of speed in children. He used five Piaget type tasks and determined the hierarchy among subconcepts of the concept of speed. Thirty-six pupils, each from grades six, seven and eight, participated. The average age of formal operational thought of the sample was thirteen years and four months. A relationship between IQ and performance on the tasks was established.

Carpenter et al. (1975b) identified two areas of pupil difficulties in the National Assessment of Mathematics which may relate to proportional reasoning. He reported that the concept of fraction was shown to be difficult to understand and use. A consumer problem that would be solvable with proportional reasoning was correctly answered by fewer than 40 per cent of the seventeen-year-olds or young adults.

Raven (1974) reported research studies he and his pupils had performed over the past seven years concerned with facilitating logical operations in elementary school and junior high school children. He saw the period of formal operations occurring between the eleventh and fourteenth years and proportional thinking, probability thinking, and correlational operations appearing during this stage.

Holloway (1967) reported that pupils at the formal operations level were able to double an area and that a transitional age for this was about twelve years.

Novak (1974), in a review of science education research of 1972, summarized cognitive development research as supporting Piaget's theory. He further saw the general need for established validity in tests that were being used and overall the need of setting research in appropriate learning theory.

[illegible] Components of Proportional Reasoning

Probing into the nature of proportional reasoning, Lovell and Butterworth (1966) made a principal component factor analysis of a set of twenty tasks as performed by 60 pupils of average to above average ability, from nine to fifteen years old. They found that the schema of proportions depends on some central intellective ability which is behind performance on all tasks involving proportion, yet specific abilities contribute to the ability to use proportionality in particular tasks. Also, tasks involving ratio depend less on the control intellective ability than tasks involving proportion. Further, they stated this proportional reasoning ability was found to appear at fourteen years of age in some pupils, while at even fifteen years of age some 50 per cent of the sample might not use proportional reasoning.

This distinction between ratio and proportion was further corroborated by the results of the Minnesota State Assessment of Mathematics. In the Minnesota Assessment of Statewide Performance in Mathematics, no objective specifically dealt with proportional reasoning yet as reported by Adams, et al. (1975). Two items testing proportion IIH3 and IIJ1 state per cent correct was respectively 16.1 and 21.2, while an item involving ratio, VB-1, was answered correctly by 61.2 per cent.

Learning Theory Implications of Some Studies

Lovell (1970) described two types of proportion, metric proportions involving the recognition of the equivalence to two ratios and the schema of proportions such as thermal capacity. This schema of proportions involves second order operations, which are operations on operations. Margenau (1950) saw something like these levels of complexity of Lovell's. Margenau postulated that concepts of physical reality should be classified by the method through which they are attained and the distance they are removed from reality.

Rosskopf, et al. (1970), as a result of observations, stated that the Piagetian proportionality schema is a general structure of actions or operations that can be applied to analogous situations. This suggests a general knowing with some different performances depending upon content but not proficiency in one and zero in another.

Renner and Lawson (1973), in reflecting on their research, suggested that mental structures represent a more or less highly organized mental system to guide behavior. Structures, in their understanding, actually represent our knowledge.

Studies Using Group and Paper-Pencil Tests

A collection of research by Robert Karplus and his colleagues has been based on group tests of proportional reasoning. Included in this collection is a survey (Karplus and Peterson, 1970), a longitudinal study (Karplus and Karplus, 1972), an investigation of cognitive style (Karplus, Karplus and Wollman, 1974), and a study of the use of ratio in differing tasks (Wollman and Karplus, 1974).

In each case, subjects in classroom groups were given pages with information and questions by one of the authors or a trained assistant. The experimenter explained each problem and carried out some demonstrations and measurements. The questions asked for some answer and a reason for the answer. Subject's answers were categorized according to these previously designed categories (Karplus and Peterson, 1970, pp. 814-815).

The survey involved 116 fourth and fifth grade suburban pupils, 82 suburban sixth grade pupils, 95 urban sixth grade pupils, 75 eight to tenth grade suburban pupils, 123 eight to tenth grade urban pupils and 153 eleventh and twelfth grade suburban pupils. The survey results (Karplus and Peterson, 1970) showed that the older urban and suburban groups were better able to solve the ratio problem than their younger colleagues.

Interpreted in terms of Piaget levels, measured performance for 75 eighth to tenth grade pupils was Preoperational, 15 per cent; Concrete Operational, 49 per cent; Formal Operational, 36 per cent. These group results substantially compare with those reported for task measures.

In the longitudinal study, Karplus and Karplus (1972) studied the growth of proportional reasoning of a group of 155 sixth, eighth and eleventh grade suburban pupils over two years of time. About one-third of the pupils showed no change in level. The changes that did occur confirmed the hierarchy of proportional reasoning ability as measured by the group test.

The seventh grade in the school had three instructional groups: "slow," "average" and "fast." The three groups performed very differently when measured in eighth grade. The pupils of the "slow" group made virtually no progress. In the "fast" group only three pupils failed to reach the Piaget Formal Reasoning Level. The pupils in the "average" group made some progress, but nothing as dramatic as that of the "fast" group.

Karplus, Karplus and Wollman (1974) studied cognitive style in the personal preference of persons for procedures for solving ratio and proportion problems.

Two forms of ratio tasks were administered to 616 pupils in grades four through nine. Results suggested that persons who do not use proportional reasoning will use strategies that are suggested by the task's presentation. Specifically, when a task involved comparison of two viewed objects, the subject without proportional reasoning often qualitatively compared the two in a manner involving scaling. When a task involved one object and numerical data for comparison, the subject without proportional reasoning often used some additive approach toward solution.

The ratio value itself might have had an effect. The ratio of 3/2, which lies between one and two, tended to increase the percentage of additive responses. A ratio of 2/1 prompted proportional instead of additive reasoning, a ratio of 5/2 caused some pupils to use approximate ratios of two or three, or become confused.

Whether the task itself affects the level of proportional reasoning, was the subject of Wollman and Karplus' (1974) latest study. They investigated the responses of 450 seventh and eighth grade pupils to six problems that required proportional reasoning and represented differing degrees of concreteness. The study suggested that proportional reasoning level was dependent on the content of the task and the type of ratio or proportion involved.

In this study paper-pencil items were used. A contrast of paper-pencil and group interview results demonstrated that group and paper-pencil tests gave substantially the same results.

Grant and Renner (1975) explored the use of written statements of explanation for multiple choice item responses as a means of identifying different levels of reasoning ability. Pupils, from three different biology sections at one large Oklahoma City area high school, were asked to respond to a twenty-minute multiple choice test and give a written explanation for selecting each answer. The same pupils were administered the separation of variables Piaget task. Results from the study were analyzed through chi-square technique and levels of significance were reviewed. Good agreement between task and written measures were established.

Studies and Precepts of Criterion-Referenced Testing

Measurement with criterion-referenced testing is a comparatively new approach in research. A concern of this study is to demonstrate an exemplary approach to criterion-referenced test design. Literature, that contained precepts for good test construction as well as studies of test construction, item design and appropriate statistics as well as examples of criterion-referenced and other paper-pencil test design, was sought to be included in the review.

Original Studies

Tests, dealing specifically with proportional reasoning at the level of junior high, were not numerous in published test collections. Within the 40 citations available in May of 1974 for mathematics tests, grade seven and above in the test collection of Educational Testing Service, no such test was found. Some subtests contain proportional reasoning components. In the Content Evaluation Series: Mathematics Test Form I by Gilbert Ulness c1969, grades seven through nine, Houghton Mifflin, there is a subtest on ratio. In the Iowa Tests of Basic Skills, Levels Edition Forms 5 and 6 by A. N. Hieronymus, c1971, grades three through eight, Houghton Mifflin, there is a subtest, ratio and proportion. Ratio and proportion is one of some twenty topics of the McGraw-Hill Basic Skills System: Mathematics Test by Alton L. Raygor, c1970, grades eleven through fourteen, CTB/McGraw-Hill; no subscores on ratio and proportion are available.

Problems concerning ratio and proportion is one of eight topics of emphasis in the Mathematics Inventory III Basic Skills of Problem Solving, c1970, grades four through twelve, American Testing Company, but no subscores are available.

Test items in ratio and proportion, when available, ask for a single correct answer and do not identify the subject's reason for a response. No items or subtests relate the score obtained to a subject's proportional reasoning level.

Test Design

Glaser (1963) saw achievement test scores as offering primarily two kinds of information. One, the degree to which the pupil has attained criterion performance. Two, the relative ordering of individuals with respect to test performance. Criterion-referenced tests were seen as having an absolute standard and providing explicit information on what individuals can do independent of the performance of others. Norm-referenced tests were seen as having a relative standard in comparison to others and providing no information on the degree of proficiency of an individual. They further differ in their construction in that items within criterion-referenced tests would have similar difficulties while items within norm-referenced tests would have items with a range of difficulties.

Hieronymus (1971) equated criterion-referenced tests with mastery tests and saw their contribution in the monitoring and assessment of instructional strategies and outcomes.

Ebel (1971) saw major limitations of criterion-referenced testing, the fact that as such tests do not tell us all we need to know about achievement, are difficult to develop on any sound basis and are only possible for a small fraction of important educational achievements.

Task Testing Concerns

Chittenden (1974) saw task testing as requiring open ended, exploratory questioning. He felt that questioning children according to the instructions of a standard protocol would force the observer to conclude that they were, by and large, able to conserve. Using a flexible, exploratory method, he found it was easy to probe to find the children were preoperational.

Flavell (1963) saw the need to allow the pupil to identify or select reasons or rationales rather than give totally their explanation.

Item Collections and Scoring

Fremer (1972) suggested that the judgment of achievement of mastery be based on achievement of a proportion of some group of items tied to a single objective. The sampling error associated with the selection of only a single exercise would pose serious problems of interpretation.

Fremer's (1972) statement in generating cutting scores was to use an operational approach. Ratings and scores would be collected for a sample of studies. That level of test performance which best discriminates among pupils judged to be above or below the minimal competency level would be sought. A cutting score on the test could be selected that would lend to the most correct classification in the sample.

Easley (1974) found a conflict between the drive for protocol uniformity to produce reliability and the need for flexibility to allow the necessary depth for probing. He felt that the quest for reliability, which results in rigid formats, is doomed to generate many errors in the identification of cognitive structures because it lacks the flexibility needed for probing.

Rowell and Hoffman (1975) stated that a group measure was needed. The individually administered tests developed by Inhelder and Piaget (1958) were viewed as prohibitively time consuming for use in the normal classroom situation. They saw that a group test, easily administered, readily marked, and yet retaining as many as possible of the attributes of the original Piagetian tasks was needed. They tested 193 pupils with a group chemistry task and 189 pupils with a group pendulum task.

No validation was made of the group task with individual tasks; no reliability was measured. The product moment correlation coefficient between the group measures was reported as r = .56.

Studies, which involved the use of more than one task (Lunzer and Pumfrey, 1966; Hensley, 1974), reported different performances for the different tasks. Some tasks were easier than other tasks and correlations between tasks when reported were in the range .25 to .42.

D. R. Phillips (1974) identified these common errors and misapplications of Piaget found in the literature: 1) training studies in which children are taught verbal responses to specific tasks, 2) interviewing techniques in which the investigator does not ask the child for reasons for his choices and 3) scoring criteria for reasons, when asked, that do not incorporate reversibility or logical necessity.

Goodyear and Renner (1975), in a preliminary study of reasons pupils gave for multiple choice item responses, found guessing to be the highest category after thought that they knew the right answer. Also overall 21.8 per cent of those having wrong answers thought they were able to justify them. The authors from this indication of probable partial knowledge suggested that a test involving pupil reasons for answers would be useful.

Written Tasks

Karplus and Karplus (1974) discussed interview versus written tests. They saw the pupil's school work as more closely similar to the written task situation than to the clinical interview.

Studies Employing Criterion-Referenced Testing

DeAvilla and Struthers (1967) developed a group measure of pupil level with subtests in conservation, causality, relations and logic. A cartoon format based on thirty or so situations from Piaget experiments was used. Test quality was described in terms of homogeneity ratios and reliability coefficients. Tests resulting had limited homogeneity and good reliability. The reliability values, Cronbach's Alpha (1951), were conservation, .694; causality, .550; relations, .001; logic, .227; total test, .717.

The domain referenced assessment of Hively, Patterson and Page (1968) is a process of generating items out of a matrix or grid expressing the contents and behaviors to assess with the assumption that all relevant contents, behaviors and related factors can be defined from a domain or a universe of objectives. Basic item shells would next be constructed to generate items to meet the prespecified criteria. Such prescribed procedures were followed by Bart (1972) and Gray (1970) where items originated from item shell descriptions for their stem and distractors.

DeVries (1973a} through factor analysis, probed the relationships among Piagetian, achievement and intellectual assessments. She concluded that Piagetian measures represent some aspects of intelligence and achievement which are not included in standardized assessments. DeVries (1973b) further reported that psychometric tests and Piagetian tasks seem to reflect two different kinds of intelligence.

Robertson and Richardson (1975) studied the problem of whether the conservation of a derived quantity in physics is dependent upon the conservation of constituent fundamental quantities. A random sample of 25 boys and 25 girls from each of grades seven through ten were participants in the study. This sample stratified for age and sex represented 25 per cent of the pupils in a coeducational high school in an outer Sydney area.

Testing was done using a procedure where the materials and operations were demonstrated clearly to the pupils. A question which was printed on the question paper was repeated. The subjects were required to indicate their response on the paper by circling yes or no. Reliability of the testing was established through test and retest of a random sample drawn from grades seven and eight, individually and group processes were suitable. Testing was completed in two days. Chi-square analysis was applied to identify significant change. The writer established that conservation of constituent fundamental quantities was a determinant in conservation of a derived quantity.

McLeod, Birkheimer, Fyffe and Robison (1975) accomplished the development of a collection of criterion validated test items to measure the science processes of controlling variables, interpreting data, formulating hypothesis and defining operationally. The development proceeded from writing a collection of face validated items which were administered to 56 individual competency measured pupils.

Pearson product moment correlation coefficients between scores on the individual criterion measures and scores on the selected group test items ranged from .535 to .705 and all correlations were significant at the .001 level.

An attempt was made to develop and validate a Piagetian-based written test with successful use of the logic of specific Piagetian tasks defined as the criterion by Gray (1970). Ninety-six randomly selected nine- to sixteen-year-olds, stratified by age, were individually presented the Piagetian tasks of pendulum, balance, and combinations and group administered a thirty-six item logically equivalent written test. Results indicated that a criterion-referenced approach to constructing a Piagetian-based written test of cognitive development is possible and that the average age of change from concrete to formal operations is consistent with previous research.

Analysis Techniques of Validity and Reliability

Lawson and Renner (1975) developed content based reasoning level tests. Face validity was established by six prominent science educators with competence in science and experience in Piagetian theory. Examinations were content validated by the classroom teachers in the respective subject matter areas. Reliability of each subject matter examination was determined by using the Spearman-Brown split half correlation technique. The reliabilities were: biology exam, 0.76; chemistry exam, $r_H = 0.71$; physics exam, $r_H = 0.59$. However, test items had no described theoretical basis or construct validity.

Glaser and Nitko (1971) suggested that criterion-referenced tests may not directly employ classical measures of reliability since many of the item and test statistics employed with norm-referenced tests are dependent on the observed variance of the total test scores. Criterion-referenced tests are expected to have little variance in total test scores.

Hambleton and Novick (1972), in reviewing the definitions for criterion-referenced tests of Glaser and Nitko, Harris, Steward, Bormuth, and Hively, Patterson and Page, stated that common to criterion-referenced tests is the definition of a well specified content domain and the development of procedures for generating appropriate samples of test items. Criterion-referenced tests may often be multidimensional while made up of unidimensional subscales.

Carver (1970) suggested that the reliability of a single form of a criterion-referenced device could be estimated by administering it to two comparable groups. The percentage that met the criteria in one group could be compared to the percentage that met the criterion in the other group. He further suggested that the reliability of a criterion-referenced test should be assessed by comparing the percentage of examinees achieving the criterion on parallel tests.

Zeiky (1974) described a reliability index as an indication of the consistency or stability of a test score. A reliability index, in his description, technically indicates what percentage of the score variance is true score variance.

Livingston (1972) proposed a measure for criterion-referenced test reliability which includes a special case, norm-referenced reliability. Livingston reasoned that the basic difference between norm-referenced and criterion-referenced measurements is that when using norm-referenced measures, one wants to know how far a pupil's score deviates from the group mean and when using criterion-referenced measures one wants to know how far his score deviates from a fixed standard. Therefore, each concept based on deviations from the mean score should be replaced by a corresponding concept based on deviations from the criterion score.

Harris (1972) objected to the Livingston coefficients because it appeared identical to a conventional reliability coefficient, when that coefficient was based on two populations with means equally distant above and below the criterion score. Livingston replied to this objection emphasizing that criterion-referenced test score interpretations do not require that the criterion score be seen as a mean of score distribution.

A test-retest approach to criterion-referenced test reliability was the suggestion of Zeiky (1974). The percentage of cases that shift classification, between successive administrations of the same test or between parallel terms, would be the measure.

Content validity of a criterion-referenced test must be high. Popham and Husek (1969), Kriewall (1969), Carver (1970) and Hambleton and Novick (1972) all state this in some way. Popham and Husek saw this as the primary measure of validity.

Zeiky (1974) discussed the methods of cutting scores. Among these he included the method of empirically using preselected groups which within a school system, particularly at the elementary years, could be the grade levels. Masters could be those pupils who have taken a course or by age have had the experience. Non-masters would be from some lower grade. The criterion-referenced test would be administered to both groups and the distribution of scores obtained. A cutting score then would be selected that best discriminated between the two groups. This idea of cutting scores and empirical examination of levels gives direction to the examination and design of a developmental level test.

Zeiky (1974) applied the ideas of classical test theory to criterion-referenced tests. He felt it should be possible to apply traditional methods if score variance is "built-in" by selecting two pretest samples known by independent means to be split evenly above and below mastery level and pooling them into one group.

Woodson (1974) had similar views and stated that for criterion-referenced tests, item analysis and test development must be done on observations representative of the observations within the range of interest on the characteristic of interest that is above and below the criterion level.

Zeiky (1974), Kriewall (1969) and Ivens (1970) saw that item difficulty measures can be used to improve a set of intended homogeneous items. Ivens suggested that any one of a set of homogeneous items that has a difficulty widely discrepant from others in the set should be treated with caution.

Zeiky summarized the recommendations concerning item discrimination indices use of Popham and Husek (1969) and Nitko and Hsu (1974) that one should consider score variance as well as the index. If normal discrimination indices are low because score variance is low, there is no problem. If score variance exists in reasonable amounts and item discrimination is still low, there is likely to be a problem. If discrimination indices are negative, there is definitely a problem which should be corrected. An index of item quality was suggested by Besel (1973) based on estimates of the probability that a "non-master" will answer an item correctly; the probability that a "master" will have an item wrong. The index identifies with high indices those items with the most information for dividing pupils into masters and non-masters. Estimates of the index can be obtained by administering the item to groups known by independent means to consist of non-masters and masters respectively.

CHAPTER 3

PHASE I - THE PILOT STUDY

Phase I of this study was a probe into the nature of proportional reasoning levels and a trial of the possibility of measuring proportional reasoning levels with a paper-pencil test.

Setting

School Site

The pilot study was conducted in Penn Junior High School in Bloomington, Minnesota. The city of Bloomington had three junior high schools. Penn Junior High School pupils ranked the highest of all junior high schools in the mean composite score on the Iowa Tests of Basic Skills. With regard to socioeconomic status, Penn Junior High School ranked second among the three junior high schools.

Penn Junior High School was chosen because of the interest and cooperation of their science teaching staff. The writer had worked with this staff to review their goals for science teaching. The study had its origin in questions this group had about the problems their eighth grade pupils were having while using proportions in physical science.

Pupils

Classes of two of the four grade eight physical science teachers were used by the writer in conducting Piagetian task interviews with pupils. The teachers of these classes pointed out pupils with low and with high class performances so that the writer might select pupils with some range of ability. The pupils in the sample had completed some three months of the half-year course at the time of task interviewing and had completed all of the course at the time of paper-pencil testing.

Basic Design

Initial Study

The writer had tested four grade eight mathematics classes with the Mr. Tall and Mr. Short ratio problem (Karplus and Karplus, 1970). Pupil answers followed the pattern found by Karplus.

Discussions, with Robert Karplus, with Clarence Boeck and with John Stecklein, encouraged the writer to develop a paper-pencil instrument.

The writer sought in a pilot study to gain some indication of probable tasks to use, task testing experience, and appropriate content for proportional reasoning testing.

Task Interviews

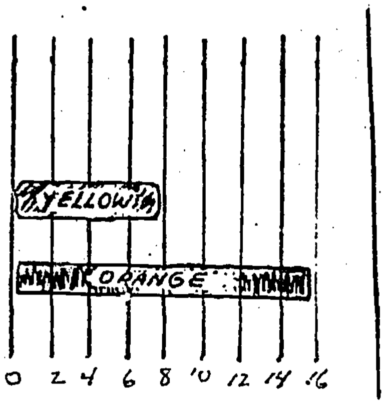
Piagetian task interviews were conducted using a total of 25 tasks with a total of 25 pupils. Each group of five pupils performed a set of five tasks. That is to say: pupils A-E performed tasks 1-5 and pupils F-J performed the next five tasks and so on through the full 25. No pupil performed more than five tasks but each task was performed by five pupils. This is tabled in the Phase I results later in the chapter.

Each task involved physical objects and materials. The pupils observed and handled these objects and materials. The tasks involved physical and geometric proportions. Direct, inverse, direct-as-square and inverse-as-square relations were all included in the interview tasks. Each interview followed a defined question format that was structured after the Chittenden (1974) approach of probing questions culminating in a direct question asking for the student's reasoning.

Task:

The rods are measured for the pupil.

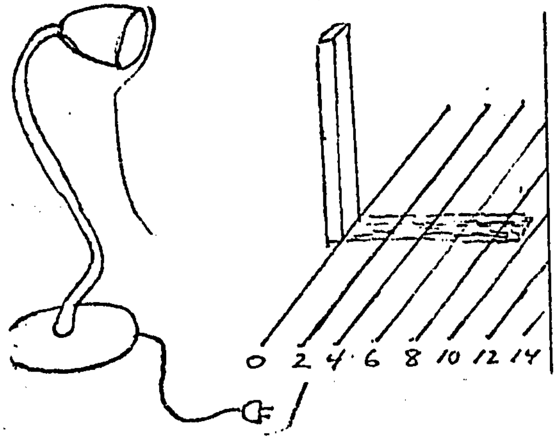
The longer one is set up and its shadow measured.



Materials:

Cuisenaire rods,
8 cm orange and
4 cm yellow

Ruled grid,
Lamp - Hi intensity



Questioning:

Introduction: The orange rod you can see is about 16 units long. The yellow one is about 8. When I set up the orange rod and the lamp, the rod has a shadow 10 units long.

Prediction: The number of units of shadow I would get if I set up the yellow rod in the same way without moving the lamp.

Appendix B includes similar descriptions of the final version of many of these tasks.

Five task interviews were conducted with each pupil. The interview and each pupil's response were recorded on audio tape as well as being recorded in notes. Responses were scored into categories according to the criterion behavior exhibited and given a numerical value. This scoring is described in Table 3.1.

Table 3.1

Task Interview Criteria

Stage	Criterion Behavior and Example	Score
Preoperational	Subject guesses--or makes no connection between how things change and some rule. Pupil example: "I guessed."	0
Concrete I Operational	Subject compensates in some qualitative way. Pupil example: "Because it's bigger."	1
Concrete II Operational	A rule, usually addition, is used to calculate the increase or decrease. Pupil example: "I added 10 + 6 = 16 so 2 + 6 = 8."	2
Formal I Operational	The subject calculates by multiplying or using simple ratios. Pupil example: "10/16 x 8 = 5. I multiplied."	3
Formal II Operational	The subject uses proportions. Pupil example: "5/8 = 10/16. It's proportional."	4

Sample pupil responses and their scoring are shown in Table 3.2. Student answers were recorded in notes and in audio tape recording. The grading of responses was done from notes and replaying the tapes.

Table 3.2

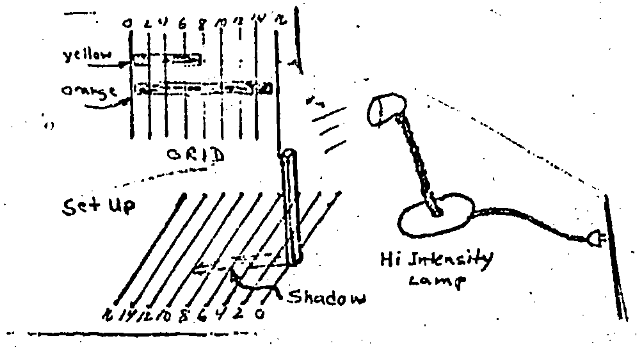
Sample Pupil Responses

Answer	Reason	Score
5	I guessed	0
About 4	It has to go down	1
2	It goes down 6	2
5	I multiplied 10/16 x 8	3
5	Because it goes the same way 10/16 is 5/8	4

Paper-Pencil Tests

The twenty-five tasks were then written as paper-pencil items and all items were given to all 25 pupils. Because the writer questioned what form to use for the items, distractors for the paper-pencil items were written in the four different forms illustrated. The item forms were distributed throughout the test.

Flag Pole



Introduction (stem): The orange rod you can see is about 16 units long. The yellow one is about 8.

When I set up the orange rod and the lamp, the rod has a shadow 10 units long.

Predict (question): The number of units of shadow I would get if I set up the yellow rod in the same way without moving the lamp.

Form I

Pupil solves the problem for his answer which he records, and selects a description indicating his method of solution.

____
Answer you found

Reason

a - I guessed
b - I added
c - I multiplied
d - I used a ratio

Form II

Pupil selects an answer and an appropriate reason.

a - 5 5/8 = 10/16
b - About $4\frac{1}{2}$ short is half as tall
c - 4 I subtracted a little less
d - 2 I subtracted 6

Form III

Pupil selects an answer and a reason from identical answers but different reasons.

a - 5 because 5/8 = 10/16
b - 5 because 10/16 x 8 = 5
c - 2 because 8 - 6 = 2
d - 2 because it should be smaller

Form IV

Pupil selects a method. Select the approach you would use.

a - I guess
b - I use a proportion
c - I would add
d - I would multiply

Pilot Study Results

Pupil results on tasks of this pilot study were analyzed to confirm the probable existence of levels of proportional reasoning and to examine the success of their measurement with designed tasks and paper-pencil items.

Task Interviews

Levels of proportional reasoning were evident in the results. As shown in Table 3.3, pupils did have a range of task scores.

Table 3.3

Pupil Average Scores on Pilot Tasks

Level	0	Trans.[a]	I	Trans.	II	Trans.	III	Trans.	IV
Pupils		1	2	4	3	4	2	8	1

[a] Trans. = Transitional

The pupil results were also used to analyze the discrimination power and the consistency of the tasks.

All pupil task scores were arranged in the pattern shown in Table 3.4. Here it can be seen that task I-1 Thermometer shows discrimination for cnly one pupil scored. This suggested that this task should not be used in further testing.

The underlined scores (<u>3</u>, <u>0</u>) are scores which differ by 2 or more from the average score that pupil received. Such a wide difference suggested that this task may not have been measuring

Table 3.4

Rating of Pilot Task Performance

	Tasks					
Pupils	I-1 Thermometer	I-2 Folds	I-3 BB Cr	I-4 Recipe	I-5 Sq A	Average
A	2	3	0	3	2	2.0
B	2	4	3	3	3	3.0
C	2	4	4	4	4	3.6
D	0	0	0	3	0	.6
E	2	3	1	0	3	1.8

the same thing as other tasks. This recipe task was rewritten before it was used again. Description of all tasks, paper-pencil items and pupil scores may be obtained from the writer.

Paper-Pencil Tests

Levels of proportional reasoning were present as found in the paper-pencil testing. These levels are summarized in Table 3.5.

Table 3.5

Pilot Paper-Pencil Average Scores

Level and (Range of Average Scores)				
0	I	II	III	IV
(0 - 0.4)	(0.5 - 1.4)	(1.5 - 2.4)	(2.5 - 3.4)	(3.5 - 4.0)
Pupils				
	2	9	7	3

There was no perceptible difference in pupil scores with different distractor formats. Pupils who regularly solved problems by guessing would candidly indicate that they guessed when asked or would solve the problem in that way when a solution was required.

The items lacked good consistency, had a wide range of discrimination and showed variation in difficulty. In Table 3.6 it was noted that items 2.2 and 3.3 had average scores of 3.0 while items 2.4, 2.5, 4.2, 5.3 and 5.5 each had an average score of 1.9.

Table 3.6

Average Scores of Paper-Pencil Problems

Problem	Average Score
1.1	2.8
1.2	2.7
1.3	2.8
1.4	2.4
1.5	2.4
2.1	2.4
2.2	3.0
2.3	2.5
2.4	1.9
2.5	1.9
3.1	2.2
3.2	2.4
3.3	3.0
3.4	2.9
3.5	2.2
4.1	2.0
4.2	1.9
4.3	2.0
4.4	2.6
4.5	2.8
5.1	2.6
5.2	2.1
5.3	1.9
5.4	2.7
5.5	1.9

That a relationship between task scores and paper-pencil scores existed was evidenced by the contingency analysis in Table 3.7. The hypothesis that the relationship here was due to chance was rejected after the chi-square statistic was computed. Chi-square here was 19.97. For nine degrees of freedom this hypothesis may be rejected for 98 of 100 cases. This calculation is found in Appendix A.

Table 3.7

Contingency Table of Average Task and Paper-Pencil Scores

Average Paper-Pencil Score	Average Task Score 1	2	3	4	Totals
1	1	1			2
2	2	4	2	1	9
3	2	1	3	1	7
4	—	—	1	2	3
Totals	5	6	6	4	21

Implications for Phase II

Paper-pencil items did appear to measure proportional reasoning and the results were comparable to those of other researchers (Karplus, Karplus and Wollman, 1974). This implied that a thorough research study to develop a paper-pencil test should be attempted.

Variations between task measures were evident. This suggested that exacting descriptions should be made of the task interviews and three task measures based in the literature should be given to all pupils tested with tasks in the next phase. A larger number of pupils should be involved in task testing in the next phase in a way to give more pupils at each reasoning level.

The results suggested that the paper-pencil items would need much refinement. There appeared to be no clear support for pupil solution of the problem or selection of just an answer over just selecting the description of the method of solution. It was reasoned that paper-pencil items should be rigorously designed, written in sets for each of the four levels and empirically improved through large volume and repeated testing.

Certain questions, including the higher ordered proportions, direct as cube, inverse as square, appeared to be at a different level. Proportions involving circular areas gave very different results.

It was decided that proportions should not involve circular areas; the items with higher order proportions should be carefully screened.

CHAPTER 4

PHASE II - TASK INTERVIEW TESTING

This phase of the study was the task testing of a selected group of 40 eighth grade science pupils. This phase accomplished a Piagetian task measure of these pupils' proportional reasoning ability. The pupil responses to task measures and the pupil performance on task measures were the basis for construction and selection of paper-pencil items for the test instrument desired in the study.

Setting

The writer, employed by the Bloomington School District, chose to use Bloomington as the site for the study because of the convenience of working within the district and the relevance of this study to the Bloomington science program.

Demographic and pupil test data from elementary schools of the junior high attendance areas were used to establish socioeconomic and pupil ability rankings. This information was gathered by the school district in gaining Title I Elementary Secondary Education Act (ESEA) designation of target schools. Data of this sort were available from the Information Office of the Bloomington Schools. Table 4.1 shows a composite of the rankings of elementary schools by socioeconomic status and by pupil achievement test grades listed for each junior high attendance area.

Table 4.1

Socioeconomic Comparison of Bloomington Junior High Schools

School	Composite Elementary School Ranking Socioeconomic	 Pupil Tests
Penn	8	7
Portland	18	17
Oak Grove	13	13
Olson	7	8

Oak Grove Junior High seemed to be a school that would provide a median type of pupil population. At Oak Grove, pupils were modularly scheduled with science-mathematics a scheduled instructional block. It was possible at this school to give task interviews within a pupil's scheduled science time or independent study time. An 8 x 8 foot room off the science office was used for the task interviews. In this room were a table, a chair for the subject, a chair for the interviewer, a tape recorder to record task interviews and 19 small boxes, each holding the equipment for one of the tasks. An average of 25 minutes was spent with each pupil in completing all five tasks.

Sample Selection

A random sample of 40 pupils was selected from the Oak Grove grade eight pupil population of 485 pupils. This random sample had the following composition as compared with the total population as shown in Table 4.2.

Table 4.2

Comparison of Characteristics of Initial Sample with Total Population

	Bloomington Grade 8 Pupils	Oak Grove Grade 8 Pupils	Sample of 40 Oak Grove Pupils
Number			40
% male	51	51	70
% female	49	49	30
Average Lorge Thorndike IQ	110	110.5	111.4

Because of the number inequalities in the male-female composition of the sample, it was judged to be atypical. It was decided, therefore, to stratify the population by sex and ability.

The pilot study results were reexamined for correlations between proportional reasoning and the verbal, nonverbal and total IQ scores of the Lorge-Thorndike measure. Piagetian levels obtained from task interviews were found to have the following product moment correlation coefficients with Lorge-Thorndike IQ measures: nonverbal, .67; verbal, .71; total, .71. The calculation of these values is found in Appendix A.

The intent was to select a sample of approximately equal numbers of boys and girls and to have a range of abilities to ensure that all levels of proportional reasoning would be represented. Pupil nonverbal Lorge-Thorndike scores were mapped out (see Table 4.3). Choice was made by numbering consecutively

Table 4.3

Pilot Sample Characteristics

Lorge-Thorndike nonverbal scores	Sample Boys	Girls	Boys & Girls	All Oak Grove
118 and above	5	8	13	149
99 to 117	11	4	15	247
98 and below	5	7	12	86
Totals	21	19	40	482

all persons (boys and girls) within the Lorge-Thorndike level and then selecting with computer generated random numbers. When a randomly identified student was found to have moved from the district, another random number was used in the same manner.

The levels and the sample sizes within the levels were chosen, not to ensure a sample representative of all grade 8 pupils, but to ensure a sample with pupils at each of the four levels of proportional reasoning. Deliberately, larger proportions of pupils were thus chosen from the lower and from the higher Lorge-Thorndike ranges.

Basic Design

The task interview phase was used to measure proportional reasoning levels of 40 pupils through intensive interviews wherein the pupil would manipulate physical objects while completing the proportional reasoning tasks the pupil was assigned. The interviewer followed a general format but asked open and probing questions after the manner of Chittenden and Bybee. The interviewer's format was reviewed by Dr. Edward Chittenden during the October 1974 Educational Testing Service Criterion-Referenced Testing Seminar and by Dr. Roger Bybee in meetings with the writer in December 1973.

Task items involved proportionality with direct, inverse, direct-as-the-square and inverse-as-the-square proportions. The cognitive content of the task was obtained from a variety of areas. Physical tasks were those arising out of some physical law or action. Geometric tasks were those arising out of geometric figures. The nature of these task items is summarized in Table 4.4.

Task 1, the Shadow Task, and Task 19, Incline, were adapted by Hensley (1974) from the work of Inhelder and Piaget (1958). Task 2, Mr. Tall, was a task used by Karplus and Karplus (1970). Task 3, the Sled Task, was an adaptation of a task of Piaget (1970). Task 15, Pulley, and Task 16, Ruler, were those designed by Karplus, Karplus and Wollman (1974). Wollman, Hensley and Karplus extended permission for the writer's use of these tasks. The first three tasks, termed "literature tasks," were given to all 40 subjects. The other tasks, largely designed by the writer and termed "derived" tasks were each given to at least five subjects.

This pattern of task assignment used with pupils meant that the first five pupils had tasks 1, 2, 3, 4 and 5. The second five pupils had tasks 1, 2, 3, 6 and 7. The third five pupils had tasks 1, 2, 3, 8 and 9; the fourth five pupils had tasks 1, 2, 3,

Table 4.4

Task Specifications

		Proportionality				
	Title	Direct	Inverse	Direct as Square	Inverse as Square	Cognitive Content
1.	Shadow		Physical			Light
2.	Mr. Tall	Geometric				Scaling
3.	Sled			Physical		Motion - Acceleration
4.	Angle	Geometric				Similar
5.	Balance	Physical				Lever
6.	Flag Pole	Physical				Light
7.	BB Square	Physical		Geometric		Area
8.	Pattern			Geometric		Scaling
9.	Frosting				Geometric	Inverse Square Law
10.	Paint	Physical				Chemical Proportions
11.	Speed	Physical				Motion - Uniform
12.	Boyle		Physical			P/V - Gas Laws
13.	Population			Physical		Density
14.	Probability	Physical				Statistics
15.	Pulley	Physical				Displacement
16.	Ruler	Physical				Displacement
17.	Weight	Physical				Statistics
18.	Light & Shadow	Physical				Light
19.	Incline	Physical				Simple Machines
	Totals	11	2	2	1	
		Physical	Physical	Physical	Geometric	
		2		2		
		Geometric		Geometric		

10 and 11; the fifth five pupils had tasks 1, 2, 3, 12 and 13; the sixth five pupils had tasks 1, 2, 3, 14 and 15; the seventh five pupils had tasks 1, 2, 3, 16 and 17; and the last or eighth five pupils had tasks 1, 2, 3, 18 and 19.

Interview tasks were designed with written description of the testing protocol, the scoring and the setting. Protocols were to be open ended with the examiner making notes, asking for certain pupil responses and recording the interview on tape.

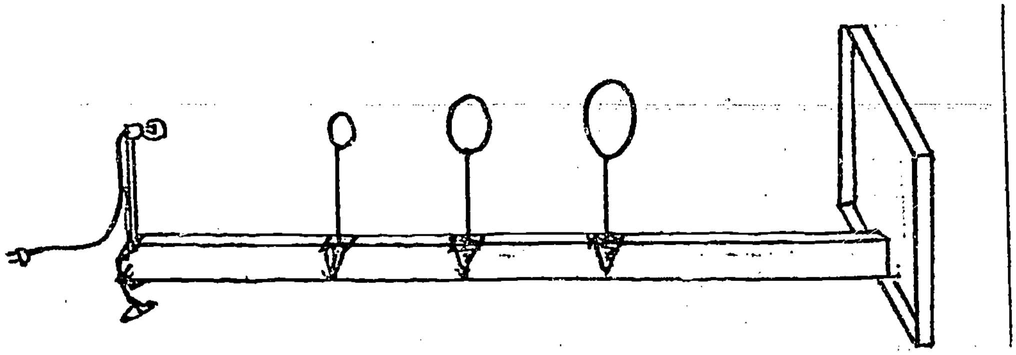
The description for Task 1, Shadows, follows. The complete set of task descriptions may be found in Appendix B.

1. Projection of Shadows (Hensley, 1974)

Thinking Tested:

Schema of Proportions
Inverse proportion - Physical

Material:



A screen, 30 cm x 30 cm, is used to observe the shadows. The shadows are made by three wire rings, 3.0 cm, 6.0 cm and 9.0 cm in diameter. Each ring has a support wire. The length of the support wire is such that the center of each ring is 12.5 cm above the bottom of the support wire. The rings are made from different colors of wire as follows: 3.0 cm (white), 6.0 cm (red), 9.0 cm (black). The rings are held vertically on a meter stick by optic bench screen holders. The meter stick has only marks at each 10 cm length. Each mark is labeled with the following letters: N, R, M, K, G, F, A, B and O. A clear light bulb is supported at one end of the beam. The center of the bulb is 12.5 cm above the top of the beam. The light is turned on and off by connecting or disconnecting the cord to the 6 volt battery. One meter stick marked in centimeters and millimeters is provided for the pupil to use.

Introduction:

"Here is a board, a light and a screen. I can put up one ring (6.0 cm) on the board (at 50 cm) and then when I turn on the light (do it), I get a shadow of the ring on the screen."

Question:

Initially seek out predictions of the effects of ring size and ring position on the shadow with questions such as: "What would you predict will happen if I use this smaller (3.0 cm) ring?" "What else could change the size of the shadow?" "How?" Do what is suggested.

Culminating Question:

"How might I make just one shadow using two rings?" "Explain why this works?"

Scoring Criteria:

Stage	Criteria	Score
I	The subject represents the shadow in the way the object appears to him. He does not perceive how the shadow is formed on the screen.	0
IIA	The subject recognizes that the size of the shadow depends on the size of the object. His knowledge goes no further.	1
IIB	In addition to the ring-size dependence of the shadow demonstrated in IIA, the subject suggests qualitatively that the distance affects the shadow size, the closer the object is to the screen, the smaller the shadow.	2
IIIA	The subject quantitatively compensates between distance and shadow size, between distance and diameter, but is not generalized as a rule. The subject begins to measure distance from the light source.	3
IIIB	From the start the subject measures both the distance from the light source and the diameter of the rings. He looks for a numerical hypothesis based on the divergent structure of the light rays. The subject is able to state in a numerical form the general relation for the two rings to have just one shadow.	4

Phase II Results

Pupil responses to task interviews were collected in pupil notes, observer notes and audio tape records. Pupil responses were scored by the writer according to criteria as described. For each task in Appendix B, overall calculation of correlations between these task scores was not made but postponed for analysis with the final results of Phase III. The scores and the averages were used at that time.

For a qualitative analysis of results, a composite listing was made of all pupil scores, the average scores on literature based and derived tasks, and the overall average. The task scores in this phase were more consistent than task scores in the pilot phase. The average pupil task levels are listed in Table 4.5. These averages cluster at Level II. Some pupils did achieve every level.

Table 4.5

Pupil Task Averages by Level

Task	Level				
		I	II	III	IV
	(0-0.4)	(0.5-1.4)	(1-5-2.4)	(2.5-3.4)	(3.5-4.0)
Literature tasks	0	6	22	9	3
All tasks	0	4	22	11	3

The difficulty of the literature tasks was estimated by averaging the pupil scores obtained for each of these three tasks. They were respectively: task 1, 2.40; task 2, 2.30 and task 3, 2.08.

Implications for Phase III

Recorded pupil responses were retained for building the paper-pencil items of Phase III. Pupils on task 3 had a low overall average. Because it was suspected that task 3 had a higher difficulty, multiple choice answers were designed with clear illustrations of the motion that the item questioned.

It was not conclusive that any tasks should be eliminated. All tasks were written as items at each of the four levels of proportional thinking, insofar as possible. All of these tasks were the content of test items. Some 76 items were used for the first testing in Phase III.

CHAPTER 5

PHASE III - PAPER-PENCIL TESTING

Phase III of the study was the design and selection of items for a paper-pencil instrument to measure proportional reasoning. Paper-pencil testing started with a set of 76 items administered to the 40 pupils who had been tested with interview tasks in Phase II. The content of the items was that of the 19 Phase II tasks. As many as four items were written for each task covering the four proportional reasoning levels.

Pupil performance was used to judge item effectiveness in the selection of a set of 24 items from an initial set of 76 items. This selection and the continued item improvements made through further testing are described in this chapter.

Test Versions and Sample Selection

Ten versions of the test were administered. Each version was an improvement over previous ones as a consequence of the changes in items or the replacement of some items with others. Table 5.1 summarizes the characteristics of each version, the pupil samples that were tested and the relationship between the versions.

Version I consisted of 76 items over the four levels of proportional reasoning. This was administered to 40 eighth grade

Table 5.1

Test Versions and Pupil Samples

	Test		Pupil Sample	
Version	Characteristics	Number	Description	Selection
I	76 items 4 levels	40	Grade 8 "transitional"	Pupils selected randomly within three intelligence levels for task testing
II A	24 items 6 at each of 4 levels	29	Grade 8 "transitional"	Randomly selected from 385
II B	12 items per pupil in a "matrix" sample 6 at Level I; same for all Another 6 from among Levels II, III and IV	27	Grade 5 "non-masters"	One total class
II C	Same test for all 6 at Level I; 6 at Level II; 6 at Level III; 12 at Level IV	77	Chemistry pupils "masters"	Chemistry classes at one high school
III A	29 items; 6 at each Level I, II, III and IV. Five additional items for Level II	393	Grade 8 "transitional"	All Grade 8 pupils in one school
III B	12 items per pupil in a "matrix" strategy. The same 6 Level I for all. Another 6 chosen from Levels II and III	30	Grade 5 "non-masters"	One total class
IV A	30 items, 6 at each Level I, II, III and IV; additional Level III items	77	2 separate Grade 8 groups	77 pupils selected randomly from 385
		195	"transitional"	195 as half of the total Grade 8 population
IV B	30 items, 6 at each level and 6 additional Level IV items	69	Physics classes "masters"	Physics classes in one high school
V A	30 items, 6 at each level and 6 additional Level IV items	427	Grade 8	All Grade 8 pupils in one school
V B	Identical with V A except for the substitution of 2 items and rescoring		"transitional"	

pupils selected randomly within three intelligence levels for task testing.

Version II A, which resulted from review of Version I results, had two related verions, II B and II C. Version II A, the basic set of items, consisted of 24 items, six items at each of the four proportional reasoning levels. Twenty-nine pupils, randomly selected from a group of 385 grade eight pupils, were tested with this version.

Version II B had three forms designed so that responses of a class of 27 fifth grade pupils, supposed non-masters, to Level I items could be analyzed thoroughly and some measurement could be made of the other items. Each of the forms had twelve items. Six of the items in each form were the six Level I items from Version II A. The additional six items were selected from each of the other three levels.

Version II C was a 30 item adaptation of Version II A that was used with 77 high school chemistry pupils, supposed masters, to thoroughly analyze Level IV items. An additional six Level IV items were used along with the Version II A items in order to consider some replacement of Level IV items.

Version III A, which was administered to 393 grade eight pupils, was the result of the improvements in Version II. Twenty-nine items were used in this version, six at Level I, eleven at Level II, six at Level III and six at Level IV. The additional Level II items were intended for consideration for improvement of Level II.

Version III B, administered to 30 fifth grade pupils, was designed as two forms with 12 items each. Six Level I items of Version III A and three items each from Levels II and III of Version III A were used in the two forms. A special purpose of this testing was the improvement of Level I items.

Version IV A was a set of 30 items that was administered to 272 eighth grade pupils. Seventy-seven of these pupils were randomly selected from the 385 grade eight pupils of a school. The additional 195 pupils were the grade eight pupils enrolled in second semester science classes in another school. The test contained six Level I items, six Level II items, twelve Level III items and six Level IV items. Overall item improvement was intended from this testing as was the possible replacement of some Level III items.

Version IV B contained most of the items used in Version IV A with the exception that six items were used at Level III and twelve items at Level IV. The responses of the supposed masters who took the test, 69 high school physics pupils, were used to improve the upper levels of the test.

Versions V A and V B were administered to 427 grade eight pupils, essentially all the grade eight pupils in one junior high. The purpose of this testing was to develop descriptive statistics regarding the final version of the test. Version V A and V B were the single test that was to be the final test version of 24 items. Thirty items were used. The 24 items that were scored as the basic test consisted of six for each of the four levels. Six additional Level IV items were included. With the replacement of two of the original Level IV items by two from the additional six items which were part of Version V A, Version V B came into being upon rescoring the papers.

Basic Design

The paper-pencil testing was carried out to select a final form of 24 items, six items at each of four levels. An initial set of 76 items were written. Each item of the initial 76 item set was constructed according to procedures for good item construction after Mehrens and Lehman (1972). Only procedures 5-9 inclusive were pertinent.

5. Prepare a table of specifications
6. Decide upon the type of format to be used
7. Prepare test items
8. Evaluate
9. Revise

The table of specifications used was that to be found in Table 5.2. It can be seen that the items were to sample all levels and to be written in both a geometric and physical context. Content of the test item came from the nineteen tasks used in task interviews. Pupil responses to these tasks were helpful in forming the items.

The paper-pencil test items, the item key and the distractors were written to specific criteria from Inhelder and Piaget (1958). This was in accord with the specifications of

Table 5.2

Specifications of Paper-Pencil Items Desired

Context	Stage and Level				Approximate Totals
	Concrete Stage		Formal Stage		
	Level I	Level II	Level III	Level IV	
Geometric	a	a	a	a	30
Physical	a	a	a	a	50
Total	20	20	20	20	80

[a] Exact numbers in each context were not established ahead of time.

Glaser and Cox (1968) for criterion-referenced measure. As Glaser and Nitko (1971) prescribed, the classes of behavior for each level were specified as clearly as possible before the test was constructed.

Paper-pencil test item format, criteria and test examples are illustrated by level in Figures I, II, III and IV. The key is located as the first answer in these examples. In practice, however, the locations of the key and distractors were varied by setting out all possible combinations of the first four answers and then randomly assigning them.

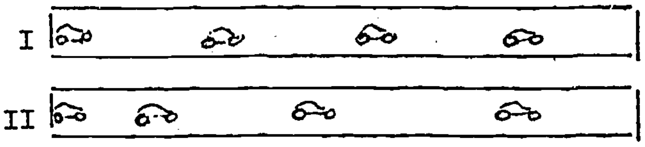
Answer "E," <u>I have no answer</u>, was always placed as the last answer. Thus, a pupil need not enter a guess when no answer seemed plausible.

Item Design Concrete I Stage (Level I)

	Stage	Score	Criteria
Key	Concrete I	4	Subject compensates in a qualitative way. May match two direct ordered relations or use addition or subtraction to contrast or calculate ratios A > B > C > D • • • • J > K > L > M
Distractor	Reasoned Guess	3	Subject makes erroneous connection but one which involves appropriate elements
Distractor	Reasoned Guess	2	Subject makes reverse ordered connection but involves elements
Distractor	Illogical Guess	1	Subject guesses or makes no ordered connection, nonsensical
Distractor	None	0	Subject makes no response

Item Example ($11C_1$)

A car moving at a constant speed of 30 mph will, if pictured at one second intervals, look like:



Answer	Stage
A. I because it moves equal distances each second	Concrete I
D. II because it is increasing its distance	Reasoned Guess
C. II because it changes	Reasoned Guess
B. None of these because it is moving	Illogical Guess
E. I have no answer	None

Figure I. Level I Item Design and Example: Test Item 5

Item Design Concrete II Stage (Level II)

	Stage	Score	Criteria
Key	Concrete II	4	Subject orders corresponding relations (with inverse) $A > B > C > D$ $J < K < L < M$
Distractor	Concrete I	3	Subject compensates in some qualitative, non-ordered way (or direct - not inverse)
Distractor	Reasoned Guess	2	Subject makes erroneous connection but one which involves elements
Distractor	Illogical Guess	1	Subject guesses or makes no connection between how things change
Distractor	None	0	Subject makes no response

Item Example ($14C_2$)

These nature hunt groups are chosen for a nature hike. The teacher with the most pupils to help is:

Mrs. Andrews	- 5 pupils
Mr. Denton & Mrs. Felk	- 8 pupils
Mr. Holt	- 6 pupils

Answer	Stage
A. Mr. Holt because 6/1 is larger than 5/1 is larger than 8/2	Concrete II
C. Mr. Denton and Mrs. Felk because they have the most pupils	Concrete I
B. Mr. Denton and Mrs. Felk because 2/8 is larger than 1/5 is larger than 1/6	Reasoned Guess
D. Mrs. Andrews because she has fewer pupils	Illogical Guess
E. I have no answer	None

Figure II. Level II Item Design and Example: Test Item 21

Item Design Formal I Stage (Level III)

	Stage	Score	Criteria
Key	Formal I	4	Subject multiples, uses simple ratios, contrasts ratios and can order them 5/25 2/25 5/25 x 10 = 2
Distractor	Concrete II	3	A rule, usually addition or subtraction, is used to contrast or calculate ratios
Distractor	Concrete I	2	Subject compensates in some qualitative way
Distractor	Guess	1	Subject guesses or makes no connection between how things change
Distractor	None	0	Subject does not respond

Item Example ($10F_1$)

Jim uses 4 heaping teaspoons of Tang powder with an 8 oz. glass of water. How much Tang is needed for the same mixture with 12 oz. of water?

Answer	Stage
A. About 6 teaspoons because 12/8 x 4 tsp. = 6 tsp.	Formal I
B. About 8 teaspoons because 8 oz. + 4 oz. = 12 oz. and 4 tsp. + 4 tsp. = 8 tsp.	Concrete II
C. More than 4 teaspoons because there is more water	Concrete I
D. 4 teaspoons because it is the same mixture	Guess
E. I have no answer	None

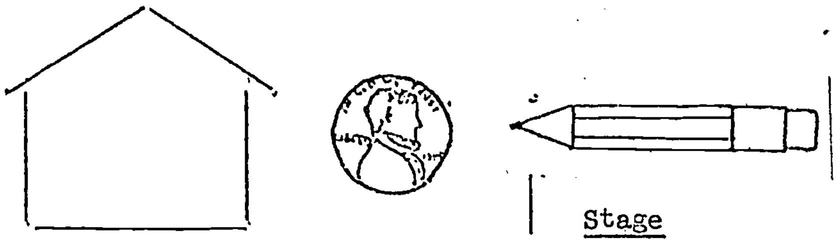
Figure III. Level III Item Design and Example: Test Item 11

Item Design Formal II Stage (Level IV)

	Stage	Score	Criteria
Key	Formal II	4	The subject calculates using proportions and recognizes the appropriate proportions to be used: $\frac{A}{B} = \frac{C}{D}$ or $\frac{A}{B} = \frac{C}{D} = \frac{E}{F}$
Distractor	Formal I	3	The subject multiplies or uses simple ratios
Distractor	Concrete II	2	A rule, usually addition or subtraction, is used to calculate the increase or decrease
Distractor	Concrete I	1	The subject compensates in some qualitative way
Distractor	None	0	The subject guesses or makes no connection between how things change

Item Example ($2F_2$)

Sketch #1 of a house is 5 pencil widths or 2 pennies high. Sketch #2 of this house is not shown. Sketch #2 looks the same but is 8 pencil widths high. How high must sketch #2 be in pennies?



Answer	Stage
B. About 3 because $\frac{2}{5} = \frac{3.2}{8}$	Formal II
C. About 3 because $\frac{2}{5} \times 8 = 3.2$	Formal I
A. About 3 because $8 - 5 = 3$	Concrete II
D. About 3 because it has to be more	Concrete I
E. I have no answer	None

Figure IV. Level IV Item Design and Example: Test Item 22

Phase III Results/Interpretations

Each testing period was followed by an analysis of results and an improvement of the item set. Deficient items were modified or replaced. In the first stage, item analysis consisted of comparing the overall results with expectations. In later stages of analysis the response patterns of masters and non-masters were contrasted. In the last stages a biserial r was calculated to evaluate the correlation of scores of masters with the levels assigned by testing and a report of the mean scores of item masters and non-masters.

Version I

Item writing for Version I produced 76 items. Table 5.3 summarizes the content and levels of these items. Seventeen items were written at the Concrete I stage, 17 at the Concrete II stage, 18 at the Formal I stage and 24 at the Formal II stage. In total, 20 items were written with geometric context and 56 with physical context. Usually four items were written from each task although as many as five and as few as one were written.

It was intended that the final planned array for Version II after item selection would be that of Table 5.4.

Observed pupil performance was used to select items for Version II. The test was taken by 40 pupils who had been selected to give performance at every level of proportional reasoning and who had demonstrated such proportional reasoning in task testing.

Piagetian Stage
F_2 or G_2 Formal II
F_1 Formal I
C_2 Concrete II
C_1 Concrete I

Table 5.3

Content and Stage of Version I Paper-Pencil Items

				Proportionality				
Content	P=Physical G=Geometrical Context	Mult'n of Relations	Inverse Mult'n of Relations	Ordering Proportions	Direct	Inverse	Direct as Square	Inverse as Square
1. Shadow	P	C_1	C_2	F_1	F_2			
2. Mr. Tall	G	C_1	C_2	F_1		F_2		
3. Sled	P	C_1	C_2	F_1			F_2 G_2	
4. Angle	G	C_1	C_2	F_1	F_2			
5. Balance	P	C_1	C_2	F_1	F_2			
6. Flag Pole	P	C_1	C_2	F_1	F_2			
7. BB Square	G	C_1	C_2	F_1			F_2 G_2	
8. Pattern	G	C_1	C_2	F_1			F_2 G_2	
9. Frosting	G	C_1	C_2	F_1				F_2 G_2
10. Paint	P	C_1	C_2	F_1	F_2			
11. Speed	P	C_1	C_2	F_1	F_2			

Table 5.3 (continued)

Content and Stage of Version I Paper-Pencil Items

				Proportionality				
	P=Physical		Inverse				Direct	Inverse
	G=Geometrical	Mult'n of	Mult'n of	Ordering			as	as
Content	Context	Relations	Relations	Proportions	Direct	Inverse	Square	Square
12. Boyle	P	C_1	C_2	F_1		F_2		
13. Population	P	C_1	C_2	F_1			F_2 G_2	
14. Probability	P	C_1	C_2	F_1	F_2			
15. Pulley	P	C_1	C_2	F_1	F_2			
16. Ruler	P	C_1	C_2	F_1	F_2			
17. Weight	P	C_1	C_2	F_1	F_2			
18. Light & Shadow	P			F_1				F_2
19. Incline	P				F_2			
	56 Physical 20 Geometrical	17 C_1	17 C_2	18 F_1	11F_2	2 F_2	8 F_2	3 F_2
					24 F_2			
		76 items						

Table 5.4

Version II Test Item Content and Stage

Content	Stage (Levels)				
	Concrete		Formal		
	Level I	Level II	Level III	Level IV	Total
Geometric & Physical	6		6	6	24

These general decision rules, as shown in Table 5.5, were applied:

1. Choose items which approximate these levels of pupil performance:

Level I	50 - 60 % correct
Level II	40 - 55 % correct
Level III	30 - 45 % correct
Level IV	20 - 35 % correct

   Such percentages were chosen from recognition that correct answers to four of the six levels would be mastery. It was also expected (Hensley, 1974; Karplus and Karplus, 1970) that most pupils would achieve Level I, 70 per cent would achieve Level II, 25 per cent Level III and 10 per cent Level IV.

2. Use items with a variety of content and have both geometric and physical contexts within the selected items.

3. Change items in accord with Piaget theory and item design requirements for answers which have defined characteristics.

Because a combination of these rules was applied, an item was not rejected upon failure to meet any one rule.

Table 5.5

Characteristics of Selected Version I Items for Version II

Level I Items

Test Item	$1C_1$	$2C_1$	$4C_1$	$9C_1$	$11C_1$	$14C_1$	Average
% Correct N = 40	53	56	43	63	58	53	54.3
Decision	Use	Change	Change	Use	Use	Use	

Level II Items

Test Item	$1C_2$	$3C_2$	$5C_2$	$6C_2$	$11C_2$	$14C_2$	Average
% Correct N = 40	38	35	28	25	60	68	42.3
Decision	Change	Change	Change	Change	Use	Use	

Level III Items

Test Item	$2F_1$	$8F_1$	$10F_1$	$11F_1$	$17F_1$	$18F_1$	Average
% Correct N = 40	40	38	55	48	28	25	39.0
Decision	Use	Use	Change	Use	Change	Change	

Level IV Items

Test Item	$1F_2$	$4F_2$	$9G_2$	$11F_2$	$17F_2$	$19F_2$	Average
% Correct N = 40	14	24	24	28	10	31	21.8
Decision	Use	Use	Change	Use	Use	Use	

Version II

Version II, prepared through the selection process previously described, consisted of a basic set of 24 items. Version II was used in a different form with each of three groups:-

Version	Characteristics	Population
II A	24 items; 6 from each level; 2 forms	29 randomly selected Grade 8 pupils
II B	12 items per pupil 3 forms each with 6 Level I items and 6 items from the other levels	27 Grade 5 pupils (one class) Probable non-masters
II C	30 items; 6 for each level; 6 additional items from Level III; 2 test forms	77 Grade 11 pupils (chemistry) Probable masters

All testing was done with at least two forms of the test in which items were randomly ordered. Form 2 had the reverse item order from Form 1.

Decision rules for improvement of Version II were more complex than for Version I. The scoring provided for a classification of a pupil's level of proportional reasoning. The assigned reasoning level was then used to categorize responses. It was possible then to note how the items discriminated between proportional reasoning levels.

A pupil was assigned as a master of a particular level when he achieved correct responses for four of the six items assumed to be written at that level. It was reasoned that with six items per level and four responses per item (Level E response always was

"I have no solution"), the probability of success by pure guessing would be one-fourth per item. For six items, then, it was probable that two items might be answered correctly by pure chance.

Through test scoring, the masters and non-masters for each level were identified. Since all pupils were tested on all items, the scoring may be thought of as a classification scheme where 0 denotes non-mastering and 1 denotes mastering at respective levels (see Figure V). A person mastering all levels would follow the sort of performance on the right. A person failing all levels would follow the performance on the left.

This Version II scoring accomplished an assignment of each pupil to a performance index based upon his meeting or failing the criteria of achieving correct responses to four of the six items at each level. In Table 5.6 there is a listing of all possible performance indices arranged by the level they probably represent. The number of eighth grade pupils, masters in proportional reasoning, are listed by the performance index they achieved. As anticipated, most of the eleventh grade pupils, 78 per cent, achieved above Level II. These results suggested, however, that too many eighth grade pupils were being classified in Level 0 or Level I.

The responses of grade 5 pupils, non-masters, were valuable in evaluating the Level I items. Grade 5 results, Version II B, were obtained by hand scoring. The results, as shown in Table 5.7, suggested that Level I items were working appropriately.

	Performance Index	Failing	Performance Index	Passing
Level I items	0	Fails Level I	1	Passes Level I
Level II items	00	Fails Levels I and II	11	Passes Levels I and II
Level III items	000	Fails Levels I, II and III	111	Passes Levels I, II and III
Level IV items	0000	Fails all levels = Preoperational Stage - Level 0	1111	Passes all levels = Formal II - Level IV

Figure V. Performance Index

Table 5.6

Performance of "Masters" and "Transitional" Pupils on Versions II A and II C

Level	Performance Index[a]	Grade 8 Pupils "Transitional" N = 29	Grade 11 Chemistry Pupils "Masters" N = 75
	0000	11	1
Level 0 (Preoperational)	0001	0	0
	0010	0	1
	0011	0	0
	0100	0	0
	0101	0	0
	0110	0	0
	0111	0	0
Level I	1000	10	1
	1001	0	0
	1011	0	0
	1010	5	8
Level II	1100	0	5
	1101	0	0
Level III	1110	3	36
Level IV	1111	0	23

[a] This notation describes the levels passed and failed, e.g., 1111 means

Passed Level I
Passed Level II
Passed Level III
Passed Level IV

Table 5.7

Version II B Results

Responses	Level I Items						Level II Items						Level III Items						Level IV Items			
	1	2	4	9	11	14	1	3	5	6	11	14	2	8	10	11	17	18	1	9G2	17	19
A	1	$\underline{14}$	3	1	$\underline{10}$	2	1	0	$\underline{5}$	$\underline{3}$	$\underline{1}$	$\underline{5}$	3	$\underline{0}$	$\underline{0}$	4	5	1	1	$\underline{1}$	3	1
B	7	3	2	2	2	8	$\underline{2}$	1	3	3	0	2	1	1	4	3	1	$\underline{1}$	$\underline{3}$	4	2	1
C	4	2	2	$\underline{15}$	4	8	1	4	2	0	4	1	$\underline{0}$	3	1	$\underline{7}$	$\underline{2}$	0	0	2	1	1
D	$\underline{12}$	2	$\underline{15}$	5	0	$\underline{7}$	3	$\underline{4}$	4	1	2	1	2	2	1	3	0	3	2	1	$\underline{3}$	$\underline{0}$
E	2	6	4	3	9	1	0	1	6	3	0	0	1	3	3	0	1	1	1	1	1	6

Correct answers are underlined.

Items $11C_1$ and $14C_1$ could have been too hard since they were answered correctly by fewer pupils. Results from other levels confirm that these items do discriminate.

Table 5.8 lists responses for all grade 8 pupils: grade 8 Level 0 pupils (0000) and grade 8 Level I pupils (1000).

Table 5.8

Level I Item Results for Grade 8 Pupils on Version II A

Item number	Per cent correct by student description All N=29	 0000 N=11	 1000 N=10	Comment
$14C_1$	62	36	70	okay
$11C_1$	62	9	90	okay
$9C_1$	69	55	70	okay
$4C_1$	72	27	100	okay
$2C_1$	48	9	60	change
$1C_1$	69	36	90	okay

The first criterion for item improvement was that items for Level I should be answered correctly by approximately 66 per cent of the eighth grade pupils. Item $2C_1$ did not meet this criterion.

Contrasting the results of Level 0 and Level I pupils gives some estimation of how well each item discriminated between masters and non-masters. Item $11C_1$ was especially good at discrimination, as shown in Table 5.8. Item $2C_1$ discriminated well but should have been correctly answered by more persons. Item 2 it was concluded, needed improvement. Very familiar objects wer substituted for the pictures of the problem. Version II item decisions are summarized in Table 5.9.

Table 5.9

Version II Item Decisions

Level I Items							
Test Item	$14C_1$	$11C_1$	$9C_1$	$4C_1$	$2C_1$	$1C_1$	Aver
% Correct Responses N = 29	62	62	69	72	48	69	63
Decision	Use	Use	Use	Use	Change Example	Use	
Level II Items							
Test Item	$14C_2$	$11C_2$	$6C_2$	$5C_2$	$3C_2$	$1C_2$	Aver
% Correct Responses N = 29	52	59	62	7	38	59	46
Decision	Use	Use	Change Example	Change Ratio	Use Only 2 Charts	Reduce Ambiguity	
Level III Items							
Test Item	$18F_1$	$17F_1$	$11F_1$	$10F_1$	$8F_1$	$2F_1$	Aver
% Correct Responses N=29	21	52	45	45	55	38	43
Decision	Change Ratio	Use	Use	Use	Change Ratio	Use	
Level IV Items							
Test Item	$19F_2$	$17F_2$	$11F_2$	$9G_2$	$4F_2$	$1F_2$	Aver
% Correct Responses N = 29	31	10	28	24	24	14	22
Decision	Use	Replace Item	Replace Item	Use	Replace Item	Use	

Version II needed some improvement. Version II had the beginnings of appropriate discrimination but items at each level needed changes.

Version III A and Version III B

Version III A was constructed from the experience in testing with Version II. These decision rules were used:

1. Items within a level should have homogeneity in their overall difficulty.
2. Items should discriminate between the responses of persons identified with levels of reasoning, that is, Level III pupils should have better performance on Level III items than Level II pupils.

Selected items were randomly ordered through the test. Two versions of the test were used in all testing. One version had the reverse order of items from the other. The key and distractors for the items were randomly ordered. The population tested with Version III included all grade 8 pupils in one junior high school (see Figure VI). Thirty grade 5 pupils, one class at an elementary school, were tested with Version III B. Version III B differed from Version III A, since it included the lower three levels.

Test deficiencies were evidenced by the very large number of pupils failing to meet success by the criteria for Level I and then showing success for higher levels. Of 227 pupils who failed to correctly answer four of the six Level I items, only 99 failed to meet the criteria at the other three higher levels. It was

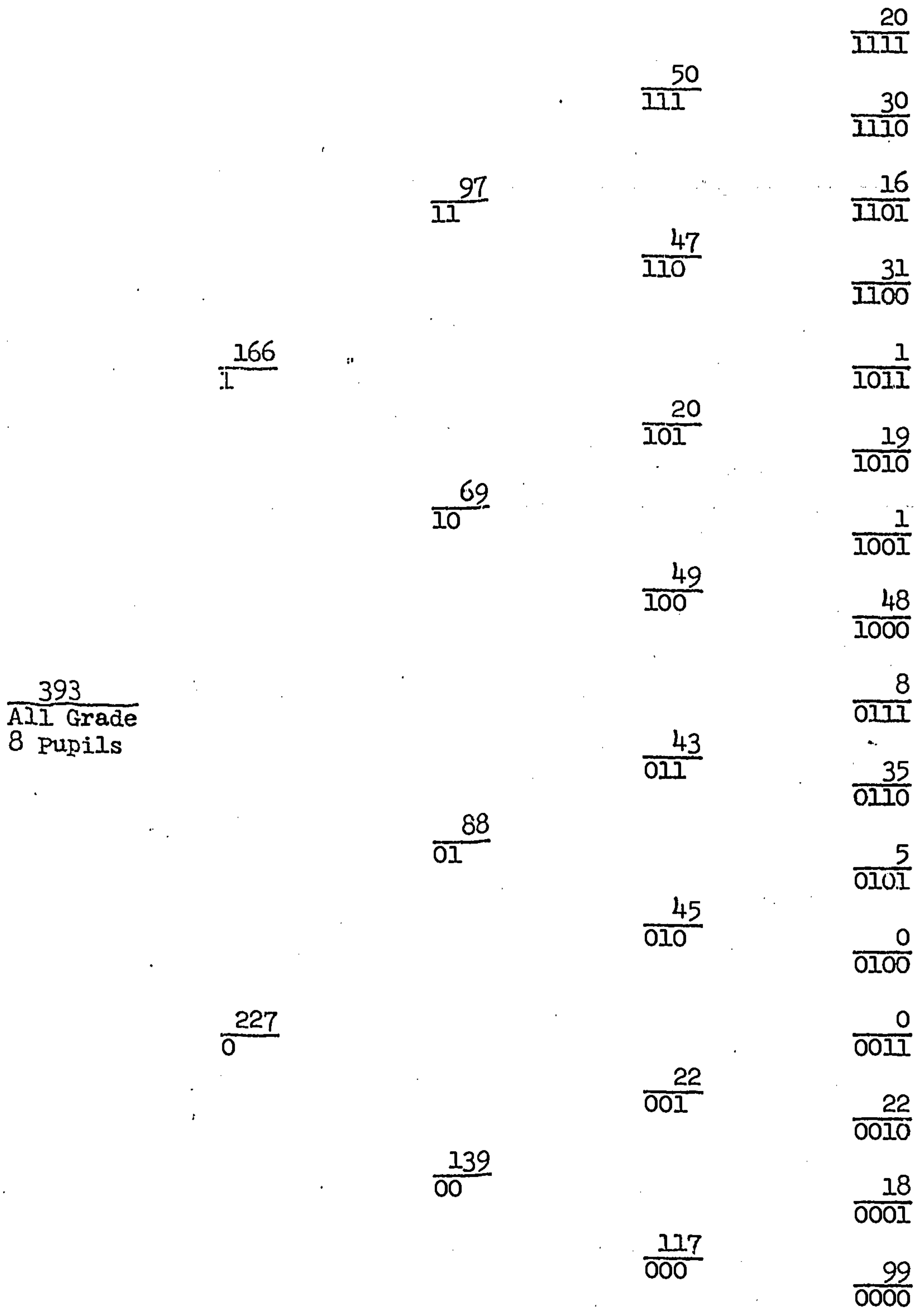


Figure VI. Grade 8 Pupil Performance on Test Version III A

found that two of the six items for Level I had been incorrectly keyed and that some program problem had not carried through the old classification. The items themselves were likely better than performance indicated.

Test analysis followed the same pattern as explained for Version II. A summary of these improvements is provided in Table 5.10.

Table 5.10

Version III A Item Decisions

Level I Items							
Test Item	$14C_1$	$11C_1$	$9C_1$	$4C_1$	$2C_1$	$1C_1$	Average
% Correct Responses N = 393	63	62	69	72	68	64	66
Decision	Use	Change only 2 examples	Use	Change Make more discriminating	Use	Use	
Level II Items							
Test Item	$14C_2$	$11C_2$	$6C_2$	$5C_2$	$3C2$	$1C_2$	Average
% Correct Responses	61	53	38	52	60	69	56
Decision	Use	Change Responses	Replace	Change 1 answer	Use	Use	
Level III Items							
Test Item	$18F_1$	$17F_1$	$11F_1$	$10F_1$	$8F_1$	$2F_1$	Average
% Correct Responses	52	68	42	49	27	43	47
Decision	Use	Add plaus. answer	Change 1 answer	Use	Change Pbm stem	Use	
Level IV Items							
Test Item	$19F_2$	$15F_2$	$10F_2$	$9G_2$	$5F_2$	$1F_2$	Average
% Correct Responses	34	34	62	21	34	21	34
Decision	Use	Change order of answers	Remove words frm ratio	Use	Use	Add more numbers	

Version IV A

Version IV A was prepared from analysis of Version III results as previously described. Version IV A had thirty items. Twenty-four of these were the six items for each of Levels I, II, III and IV. An additional six items at Level III were included to provide improvement of Level III. Test Version IV A was taken by 272 pupils. Of these pupils, 77 were those randomly selected from 385 grade 8 pupils at Olson Junior High, Bloomington; 195 of these pupils were those eighth grade pupils taking science in the second semester at Portland Junior High, Bloomington.

Version IV B had thirty items. The twenty-four items providing the core test of six items for each of the Levels I, II, III and IV were the same as those of Version IV A. The additional six items, however, were from Level IV to support improvement of Level IV items. Test Version IV B was taken by 69 pupils who were physics pupils at Lincoln High School, Bloomington. By maturity and ability these pupils were assumed to be masters of proportional reasoning.

It was intended that this testing be used to improve the items selected for test Version V. In addition to previous item selection techniques, the point biserial measure of item discrimination was calculated. Decision rules for item improvement were:

1. Items within a level should have homogeneity in their overall difficulty as evidenced in:

    a. the total percentage of persons correctly answering the item

b. the percentage of persons attaining the level who correctly answer the item

c. the number getting the item right and the number getting the item wrong

2. Items within a level should discriminate between responses of persons mastering that level and those not mastering the level as evidenced in:

   a. pupils coded as masters of the level should have performance on items of that level that distinctly exceeds that of non-masters

   b. the average scores over the test of those who are masters of the level should be approximately the same

   c. r biserial values for each item should approximate or exceed .5000

Version IV A results are described in Figure VII. Of the 272 pupils tested, 232 or 85 per cent were identified distinctively with a certain level. Table 5.11 summarizes the proportional reasoning levels assigned.

Table 5.11

Proportional Reasoning Levels of Grade 8 Pupils on Version IV A

	Number	Level	Stage	Per cent
	35	0		13
	26		Transitional	9
	62	I	Concrete I	23
	12		Transitional	4
	76	II	Concrete II	28
	2		Transitional	1
	55	III	Formal I	20
	4	IV	Formal II	1
Total	272			

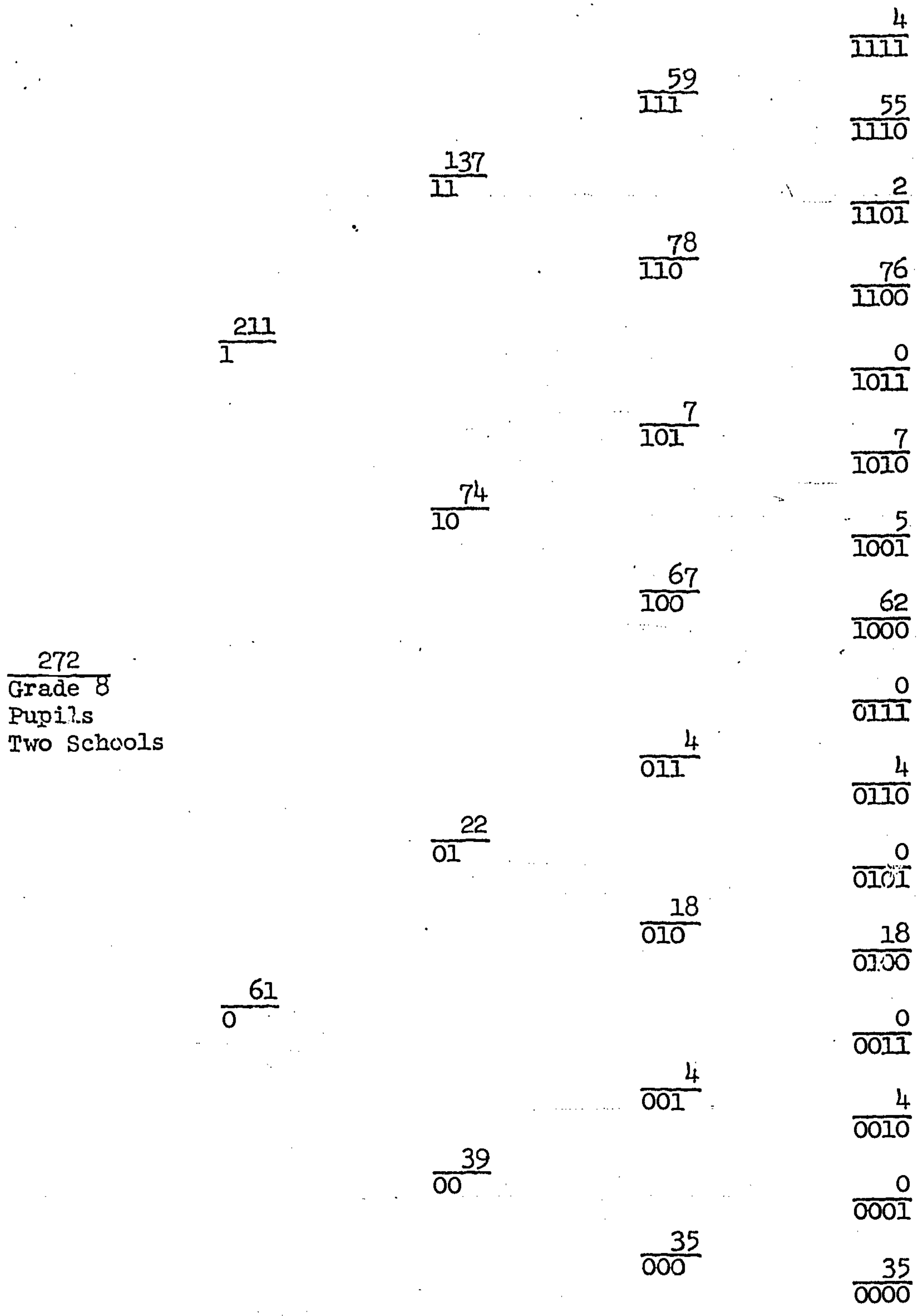


Figure VII. Pupil Performance on Test Version IV A

Grade eight responses by items are described in Table 5.12.

Table 5.12

Version IV A Item Decisions

Level I Items

Test Item	$14C_1$	$11C_1$	$9C_1$	$4C_1$	$2C_1$	$1C_1$	Average
% Correct Responses N = 272	63	71	69	62	66	55	64
Decision	Use	Use	Use	Use	Add table	More diagram detail	

Level II Items

Test Item	$14C_2$	$11C_2$	$10C_2$	$5C_2$	$3C_2$	$1C_2$	Average
% Correct Responses N = 272	77	51	68	68	60	69	65
Decision	Use	Use	Use	Replace	Use	Use	

Level III Items

Test Item	$18F_1$	$17F_1$	$11F_1$	$10F_1$	$8F_1$	$2F_1$	Average
% Correct Responses N = 272	58	34	65	48	37	43	48
Decision	Use	Use	Replace	Simplify ratios	Use	Use	

Level IV Items

Test Item	$19F_2$	$15F_2$	$10F_2$	$9G_2$	$5F_2$	$1F_2$	Average
% Correct Responses N = 272	21	18	38	19	34	29	27
Decision	Use	Use	Use	Use	Use	Use	

It was apparent that Level I items were too difficult and Level II items too easy. Item discrimination information from Table 5.13 was used as indicated.

Version IV B

Test Version IV B consisted of thirty items. The twenty-four items forming the core of the test were identical to those of test Version IV A. The additional six items, however, were from Level IV to allow improvement of Level IV items. Test items were randomly ordered in the test. The test was administered in two forms. One form had the reverse order of the other form.

Test Version IV B was taken by sixty-nine physics pupils at the same time as test Version V A was being administered. Results from Version IV B were not available for improvement of Version V A. Pupil performance on Version IV B is summarized in Figure VIII.

Decision rules for improvement of the items of Version IV B included information from calculation of the point biserial measure of item discrimination. The decision rules were:

1. Items within a level should have homogeneity in their overall difficulty as evidenced in:

    a. the total percentage of persons correctly answering the item

    b. the percentage of persons attaining the level who correctly answer the item

    c. the number getting the item wrong

Table 5.13

Item Discrimination Version IV A

	Question	# Getting Item Correct	# Getting Item Wrong	Average Score on This Level		Point Biserial Correlation	T Value
				Corrects	Wrongs		
	1-1	197	75	82.7	48.9	.618*	12.91
	1-2	247	25	77.7	31.3	.547*	10.72
	1-3	217	55	78.8	52.1	.438*	8.00
	1-4	203	69	81.9	48.3	.598*	12.24
	1-5	164	108	84.9	56.0	.576*	11.58
	1-6	170	102	86.1	52.3	.668*	14.75
Level I Average		199.7	72.3	82.0	48.2		
	2-1	198	74	71.6	40.1	.563*	11.20
	2-2	166	106	74.7	45.4	.590*	11.99
	2-3	193	79	72.3	41.4	.580*	11.70
	2-4	202	70	70.3	43.1	.491*	9.27
	2-5	155	117	71.1	53.0	.370*	6.54
	2-6	119	153	77.6	52.2	.521*	10.02
Level II Average		172.2	99.8	72.9	46.0		

* Significant at the .001 level

Table 5.13 (continued)

n Discrimination Version IV

	# Getting	# Getting	Average Score on		Point	
	Item	Item	This Level		Biserial	T
Question	Correct	Wrong	Corrects	Wrongs	Correlation	Value
3-1	131	141	57.6	29.2	.612*	12.70
3-2	96	176	59.4	33.9	.524*	10.11
3-3	155	117	54.6	37.4	.581*	11.74
3-4	175	97	53.1	24.4	.593*	12.09
3-5	52	220	46.5	42.0	.075**	1.24
3-6	91	181	59.7	34.4	.513*	9.82
Level III Average	129.6	142.4	56.9	33.6		
4-1	80	192	39.0	17.9	.510*	9.73
4-2	75	197	40.7	17.8	.543*	10.62
4-3	83	189	39.0	17.5	.523*	10.08
4-4	48	224	39.6	20.8	.381*	6.77
4-5	70	202	36.9	19.6	.401*	7.18
4-6	37	235	37.8	21.9	.290*	4.97
Level IV Average	65.5	206.5	38.8	19.3		

* Significant at the .001 level
** Significant at the .1 level

Table 5.13 (continued)

Item Discrimination Version IV A

n	# Getting Item Correct	# Getting Item Wrong	Average Score This Level Corrects	Wr[illegible]	Point Biserial Correlation	T Value
5-1	153	119	37.7	[illegible]	.614*	12.78
5-2	45	227	30.4	[illegible]	.084**	1.38
5-3	52	220	50.3	[illegible]	.560*	11.10
5-4	29	243	48.9	[illegible]	.373*	6.61
5-5	97	175	46.2	[illegible]	.710*	16.55
5-6	56	216	50.3	[illegible]	.586*	11.87
Level V Average	129.6	142	56.9	[illegible]		

* Significant at the .001 level

** Significant at the .1 level

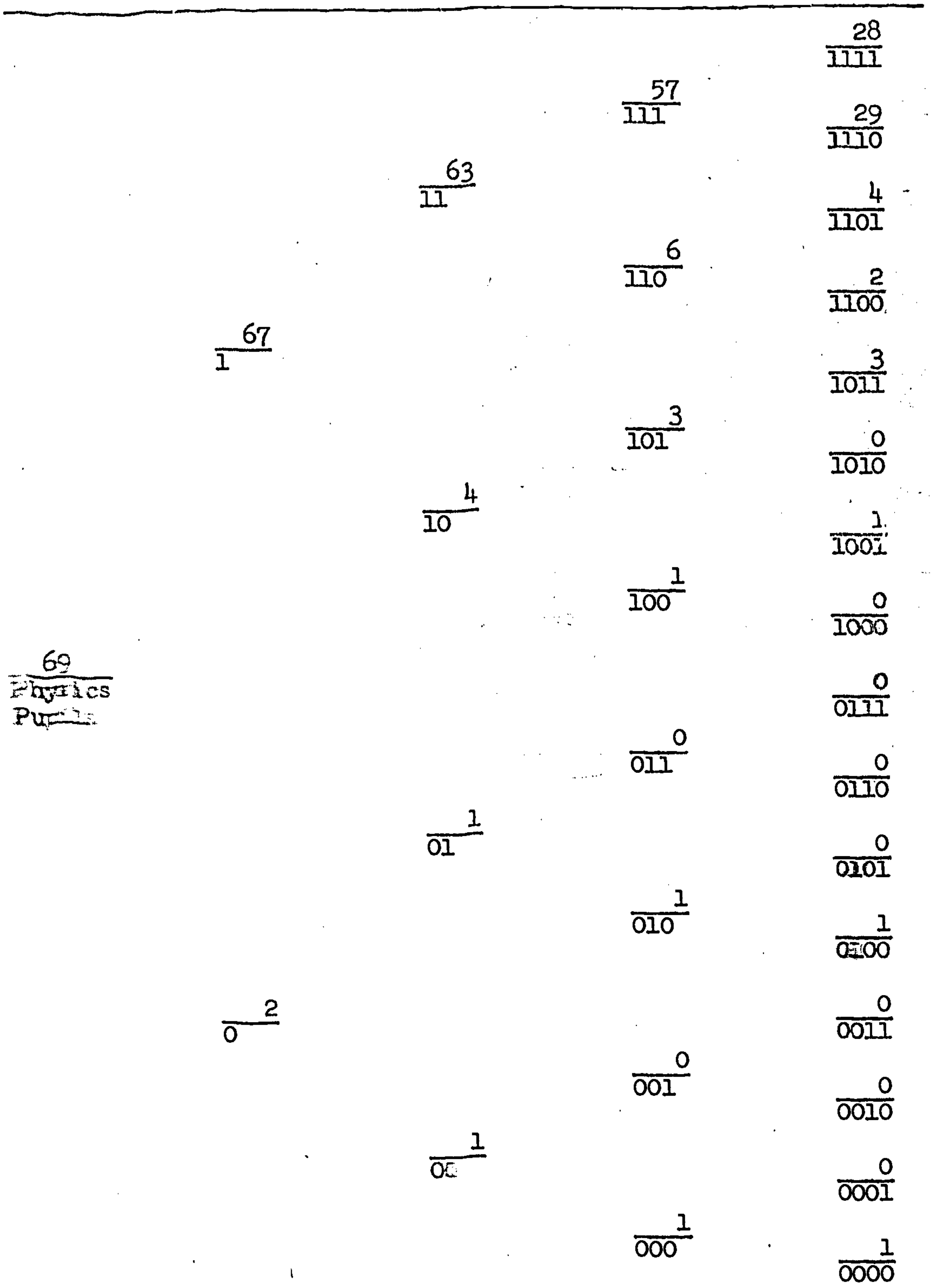


Figure VIII. Pupil Performance on Test Version IV B

2. Items within a level should discriminate between responses of persons mastering that level and those not mastering the level as evidenced in:

   a. pupils coded as masters of a level should have performance on items of that level that clearly exceeds that of non-masters

   b. the average scores over the test of those who are masters of a level should be approximately the same

   c. point biserial values for each item should approximate .500 or better

That physics pupils were indeed masters was confirmed by their performance as summarized in Table 5.14.

Table 5.14

Version IV B Item Responses of Physics Pupils

Level I Items							
Test Item	$14C_1$	$11C_1$	$9C_1$	$4C_1$	$2C_1$	$1C_1$	Average
% Correct Responses N = 69	91	94	96	91	93	91	93
Level II Items							
Test Item	$14C_2$	$11C_2$	$10C_2$	$5C_2$	$3C_2$	$1C_2$	Average
% Correct Responses N = 69	93	86	91	81	86	84	87
Level III Items							
Test Item	$18F_1$	$17F_1$	$11F_1$	$10F_1$	$8F_1$	$2F_1$	Average
% Correct Responses N = 69	84	70	87	84	62	90	80
Level IV Items							
Test Item	$19F_2$	$15F_2$	$10F_2$	$9G_2$	$5F_2$	$1F_2$	Average
% Correct Responses N = 69	52	74	57	74	54	35	58

Item discrimination information summarized in Table 5.13 and the information from Table 5.14 supported the replacement of item $1F_2$ in Version V B.

Version V A

Test Version V A contained thirty items. Twenty-four items were the core of the test. Each of the four proportional reasoning levels had six test items from this set of twenty-four. The additional six items were from Level IV to support improvement of Level IV items from pupil performance on this test and the performance of masters on test Version IV B.

Items were randomly ordered in the test. The test was administered in two forms. One form had the reverse order of the other form.

Test Version V A was administered to 427 grade eight pupils at Oak Grove Junior High School. Included were most of the original forty pupils who participated in task testing. Pupil performance on test Version V A is summarized in Figure IX.

Improvements of this version were possible through the rescoring of Level IV items. Decision rules for such improvements included information from calculation of the point biserial measure of item discrimination. The decision rules were:

1. Items within a level should have homogeneity in their overall difficulty as evidenced in:

   a. the total percentage of persons correctly answering the item

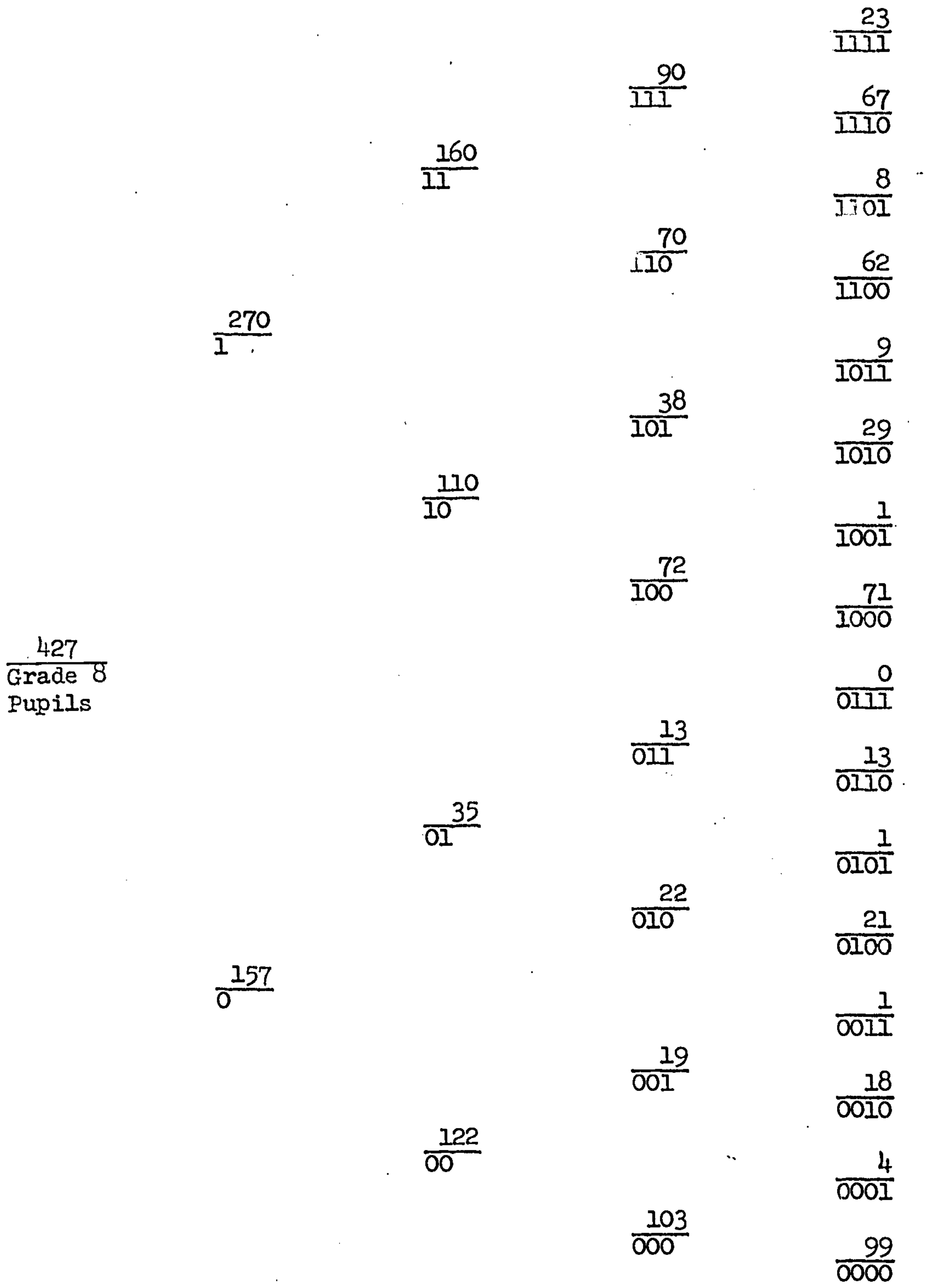


Figure IX. Pupil Performance on Test Version V A

b. the percentage of persons attaining the level who correctly answer the item

c. the number getting the item wrong

2. Items within a level should discriminate between responses of persons mastering that level and those not mastering the level as evidenced in:

a. pupils coded as masters of a level should have performance on items of that level that clearly exceeds that of non-masters

b. the average scores over the test of those who are masters of a level should be approximately the same

c. point biserial values for each item should approximate .500 or better

Seventy-five per cent (322) of the 427 total grade eight pupils were clearly identified with a proportional reasoning level. Summarizing Figure IX results, the proportional reasoning levels assigned were those of Table 5.15.

Table 5.15

Proportional Reasoning Levels of Grade 8 Pupils on Version V A

	Number	Level	Stage	Per cent
	99	0	Preoperational	23
	58		Transitional	14
	71	I	Concrete I	17
	39		Transitional	9
	62	II	Concrete II	15
	8		Transitional	2
	67	III	Formal I	16
	23	IV	Formal II	5
Total	427			

Pupil responses by [illegible] are summarized in Table 5.16.

Table 5.16

Version V A Item Responses of Grade 8 Oak Grove Pupils

Level I Items							
Test Item	$14C_1$	$11C_1$	$9C_1$	$4C_1$	$2C_1$	$1C_1$	Average
% Correct Responses N = 427	68	71	72	59	64	57	65
Level II Items							
Test Item	$14C_2$	$11C_2$	$10C_2$	$5C_2$	$3C_2$	$1C_2$	Average
% Correct Responses N = 427	67	55	69	35	50	53	55
Level III Items							
Test Item	$18F_1$	$17F_1$	$11F_1$	$10F_1$	$8F_1$	$2F_1$	Average
% Correct Responses N = 427	46	34	55	57	39	59	48
Level IV Items							
Test Item	$19F_2$	$15F_2$	$10F_2$	$9G_2$	$5F_2$	$1F_2$	Average
% Correct Responses N = 427	33	37	45	16	25	26	30

It ~~was~~ apparent ~~that~~ changes from Version IV A were improvements with the exception of the replacement of item $5C_2$. These results suggested that items $9G_2$ and $5F_2$ needed improvement. Results from Version IV B, physics masters, supported the change of item $5F_2$. Results on $9G_2$ by masters was commendable suggesting that this item was likely a higher order proportional reasoning level. The item discrimination information of Table 5.15 confirmed the need for replacement of items $9G_2$ and $5F_2$ and suggested that appropriate replacement items would be items $12F_2$ and $2F_2$.

Version V B

Test Version V B was obtained by a reworking of the V A results. Items $9G_2$ and $5F_2$ were replaced with items $12F_2$ and $2F_2$. The results for these items were appropriately assigned and the overall test results recalculated. Pupil performance on this, the final test version, is summarized in Figure X. Seventy-four per cent (317) of the 427 total pupils were clearly identified with a proportional reasoning level. Table 5.17 summarizes the Figure X results in terms of percentages of pupils attaining each proportional reasoning level.

Table 5.17

Proportional Reasoning Levels of Grade 8 Pupils on Version V B

	Number	Level	Stage	Per cent
	98	0	Preoperational	23
	58		Transitional	14
	67	I	Concrete I	16
	42		Transitional	10
	60	II	Concrete II	14
	10		Transitional	2
	60	III	Formal I	14
	32	IV	Formal II	7
Total	427			

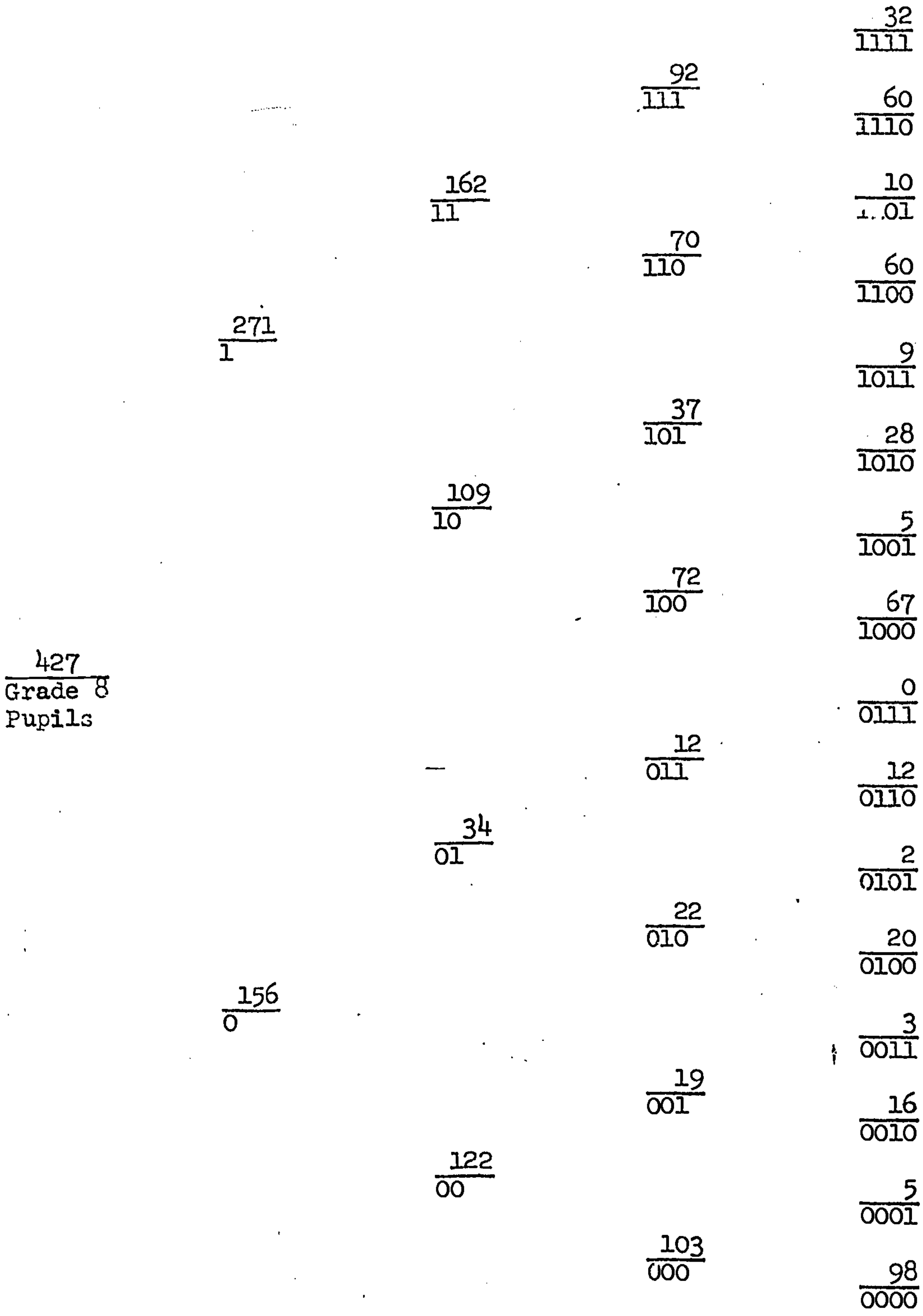


Figure X. Pupil Performance on Test Version V B

Table 5.18 presents pupil responses by item for Version V B. The replacement of the two Level IV items did improve the test.

Table 5.18

Version V B Responses of Grade 8 Oak Grove Pupils

Level I Items							
Test Item	$14C_1$	$11C_1$	$9C_1$	$4C_1$	$2C_1$	$1C_1$	Average
% Correct Responses N = 427	68	71	72	59	64	57	65
Level II Items							
Test Item	$14C_2$	$11C_2$	$10C_2$	$5C_2$	$3C_2$	$1C_2$	Average
% Correct Responses N = 427	66	55	69	35	50	53	55
Level III Items							
Test Item	$18F_1$	$17F_1$	$11F_1$	$10F_1$	$8F_1$	$2F_1$	Average
% Correct Responses N = 427	46	34	55	57	39	59	48
Level IV Items							
Test Item	$19F_2$	$15F_2$	$10F_2$	$12F_2$	$2F_2$	$1F_2$	Average
% Correct Responses N = 427	33	37	45	28	33	26	34

Table 5.19 presents data which confirm the homogeneity of items by level and relates the discrimination these items have. There is consistency between the number getting the items correct and wrong by level. The average scores on the items of those who

Table 5.19

Version V B Item Discrimination

Question	Test Item Number	# Getting Item Correct	# Getting Item Wrong	Average Score on This Level Corrects	Average Score on This Level Wrongs	Point Biserial Correlation*	T Value
1-1	1	289	138	76.1	41.9	.598	15.38
1-2	5	302	125	74.6	41.9	.558	13.85
1-3	20	306	121	74.6	40.9	.568	14.22
1-4	15	253	174	77.9	46.4	.579	14.66
1-5	9	273	154	73.9	49.4	.441	10.12
1-6	23	243	184	80.1	45.1	.649	17.57
Level I Average		278	149	76.2	44.3	.565	14.30
2-1	21	284	143	64.9	34.3	.561	13.96
2-2	12	236	191	67.2	39.0	.545	13.40
2-3	18	293	134	65.1	31.6	.605	15.64
2-4	14	148	279	72.0	45.4	.491	11.62
2-5	8	213	214	67.6	41.7	.504	12.02
2-6	2	225	202	66.7	41.2	.494	11.73
Level II Average		233	194	67.2	38.9	.533	13.06

* All biserial correlations are significant at the .001 level

Table 5.19 (continued)

Version V B Item Discrimination

	Test	# Getting	# Getting	Average Score on		[illegible] nt	
	Item	Item	Item	This Level		[illegible] rial	T
Question	Number	Correct	Wrong	Corrects	Wrongs	[illegible] ation*	Value
3-1	7	198	229	65.4	34.1	.586	14.92
3-2	24	146	281	65.6	39.7	.461	10.71
3-3	17	234	193	61.5	32.9	.535	13.04
3-4	11	245	182	61.9	30.7	.579	14.65
3-5	13	168	259	66.1	37.3	.528	12.82
3-6	3	254	173	61.2	30.1	.574	14.45
Level III Average		208	219	63.6	34.1	.544	13.43
4-1	16	139	288	47.4	21.8	.559	13.90
4-2	19	157	270	44.6	21.7	.515	12.38
4-3	4	192	235	42.1	20.4	.505	12.07
4-4	10	120	307	39.6	17.5	.488	11.52
4-5	22	141	286	39.4	16.0	.540	13.23
4-6	6	109	318	47.6	24.2	.476	11.17
Level IV Average		143	284	43.4	20.3	.513	12.38

* All biserial correletions are significant at the .001 level

got the item correct and those who got it wrong are similar. Item discrimination, as measured by the point biserial correlation coefficient, does consistently approximate .500. T-value suggests that these correlation values are not due to chance.

Summary

Paper-pencil items were improved through the changes empirically based on test results of non-master pupils, transitional pupils and master pupils.

Performance of comparable pupils on the five versions is shown in Table 5.20. The items, which are reported under Version I, are those 24 of the 76 that were used in Version II. Increased item homogeneity is evident in the decreasing range of percentage correct. Higher average values in most levels were also achieved in the later versions.

Table 5.20

Percentage Correct on Test Versions by Grade 8 Pupils

	Level I		Level II		Level III		Level IV	
Version	Range	Avg.	Range	Avg.	Range	Avg.	Range	Avg.
I (24 items only)	43-63	54	25-68	42	25-55	39	10-31	22
II	48-72	63	7-62	46	21-55	43	10-31	22
III	62-72	66	38-69	56	27-68	47	21-62	34
IV	55-71	65	51-77	55	34-58	48	18-38	34
V	57-72	65	50-66	55	34-59	48	26-37	34

CHAPTER 6

CHARACTERISTICS OF THE INSTRUMENT

In this chapter, criteria for validity, reliability and discrimination of the instrument are stated. The statistical analysis of the instrument is described and judgments are made regarding the instrument's performance with respect to the stated criteria.

Validity

Content Validity

Validity of a test is a measure of the degree to which the test measures what it is intended to measure. One component of validity is content validity. In accord with Cronbach (1960), a test has content validity if the items in the test require behaviors for their resolution that are proper to the trait being measured. The purpose of this test was to measure four levels of proportional reasoning. Items were written for each of the four levels. Each item used, as the question stem, a situation that had been used in task testing or had appeared in the literature. Specifications for writing the responses were that the key, correct answer, would be a response at the level tested and the distractors would be plausible for lower levels of reasoning.

This logical relationship of item design to theory is demonstrated in the following examples (see Figures XI, XII, XIII, and XIV) of item design taken from the test's final version. The test had strong content validity because the items in each level met the specifications for proportional reasoning of Piaget and Inhelder (1958).

Concurrent Validity

Concurrent validity, as defined by Cronbach (1960), exists when the test correlates highly positively with direct test measures of the same trait as the initial test. Concurrent validity of the paper-pencil test was assumed to be acceptable when the pupil paper-pencil test scores showed a positive correlation of at least .30 with their corresponding task interview scores. The criterion value of .30 was based on the range of reported inter-task correlations -.15 to .55 (Lawson, Nordland and DeVito, 1975). Table 6.1 summarizes the correlations for thirty-five pupils who were measured with both tasks and the paper-pencil test.

Tasks 1, 2 and 3 are, respectively, the shadow task, Mr. Tall task and the sled task. Rate 4, Rate 8, and Rate 16 are three rating schemes used to evaluate paper-pencil results. Under Rate 4 every pupil was assigned to one of four proportional reasoning levels, namely I, II, III or IV, with no transitional stages. Under Rate 8 transitional stages were identified, namely 0, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, and 4.0. Under Rate 16 the values them-

Item Design Concrete I Stage (Level I)

	Stage	Score	Criteria
Key	Concrete I	4	Subject compensates in a qualitative way. May match two direct ordered relations or use addition or subtraction to contrast or calculate ratios A < B < C < D • • • • J > K > L > M
Distractor	Reasoned Guess	3	Subject makes erroneous connection but one which involves appropriate elements
Distractor	Reasoned Guess	2	Subject makes reverse ordered connection but involves elements
Distractor	Illogical Guess	1	Subject guesses or makes no ordered connection - nonsensical
Distractor	None	0	Subject makes no response

Item Example

Mary buys three tickets to a raffle where 90 tickets are sold. Jane buys one ticket to a raffle where 30 tickets are sold. Sue buys three tickets to a raffle where 300 tickets are sold.

Which girls have about the same chance of winning?

Answer	Stage
D. Jane and Mary because three chances in 90 is the same as one in 30	Concrete I
B. Sue and Mary because each have three tickets	Reasoned Guess
A. Jane and Mary because theirs are the least tickets	Reasoned Guess
C. All girls have the same chance	Illogical Guess
E. I have no answer	None

Figure XI. Level I Item Design and Example: Test Item 1

Item Design Concrete II Stage (Level II)

	Stage	Score	Criteria
Key	Concrete II	4	Subject orders corresponding relations (with inverse) $A < B < C < D$ $J > K > L > M$
Distractor	Concrete I	3	Subject compensates in some qualitative, non-ordered way (or direct - not inverse)
Distractor	Reasoned Guess	2	Subject makes erroneous connection but one which involves elements
Distractor	Illogical Guess	1	Subject guesses or makes no connection between how things change
Distractor	None	0	Subject makes no response

Item Example

Four cars have different speeds: Car A is the fastest, Car B the next fastest, Car C the next fastest and Car D the next fastest. The fastest car takes the least time to go 200 miles, the next fastest car the next least time and so on. Which car is the third fastest and takes the third least time to go 200 miles?

Answer				Stage
A.	Car C because:			Concrete II
	1st fastest	2nd fastest	3rd fastest	
	Car A	Car B	Car C	
	1st least time	2nd least time	3rd least time	
B.	Car C because:			Concrete I
	1st most fast	2nd most fast	3rd most fast	
	Car A	Car B	Car C	
	1st most time	2nd most time	3rd most time	
C.	No car because they don't match up			Reasoned Guess
B.	Car B because:			Illogical Guess
	1 - Car D	2 - Car C	3 - Car B	
E.	I have no answer			None

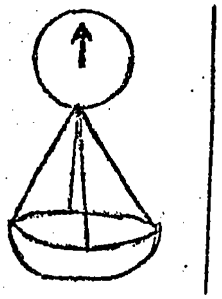
Figure XII. Level II Item Design and Example: Test Item 12

Item Design Formal I Stage (Level III)

	Stage	Score	Criteria
Key	Formal I	4	Subject multiplies, uses simple ratios, contrasts ratios and can order them 5/25 2/25 5/25 x 10 = 2
Distractor	Concrete II	3	A rule, usually addition or subtraction, is used to contrast or calculate ratios
Distractor	Concrete I	2	Subject compensates in some qualitative way
Distractor	Guess	1	Subject guesses or makes no connection between how things change
Distractor	None	0	Subject does not respond

Item Example

Jane is weighing out apples on this supermarket scale. What will fourteen apples weigh if six apples weigh 2 pounds?



Answer	Stage
C. 4 2/3 lbs. because 2/6 x 14 = 4 2/3	Formal I
B. 3 or 4 lbs. because it is more	Concrete II
A. 10 lbs. because 6 + 8 = 14 or 2 + 8 = 10	Concrete I
D. 5 because 2 + 2 + 1 = 5	Guess
E. I have no answer	None

Figure XIII. Level III Item Design and Example: Test Item 24

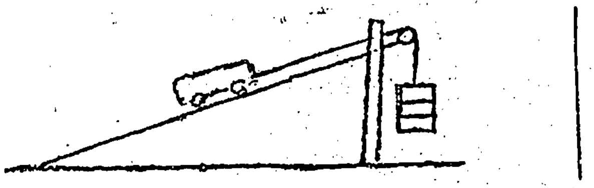
Item Design Formal II Stage (Level IV)

	Stage	Score	Criteria
Key	Formal II	4	Subject calculates using proportions and recognizes the appropriate proportion to be used. $\frac{A}{B} = \frac{C}{D}$ or $\frac{A}{B} = \frac{C}{D} = \frac{E}{F}$
Distractor	Formal I	3	Subject multiplies or uses simple ratios
Distractor	Concrete II	2	A rule, usually addition or subtraction, is used to calculate the increase or decrease
Distractor	Concrete I	1	Subject compensates in some qualitative way
Distractor	None	0	Subject guesses or makes no connection between how things change

Item Example

On the ramp illustrated, the cart and its weight are balanced by weights on the string. What amount of weight is needed to balance 400 g of cart weight at 20°?

	Weight	
Angle	Cart	String
10°	200g	35
10°	300g	52
20°	300g	100
20°	400g	?



Answer	Stage
D. 133 because $\frac{100}{300} = \frac{133}{400}$	Formal II
A. 133 because $\frac{100}{300} \times 400 = 133$	Formal I
C. 177 because it goes up 17 for every 100	Concrete II
B. 150 because it is more	Concrete I
E. I have no answer	None

Figure XIV. Level IV Item Design and Example: Test Item 16

Table 6.1

Pearson Correlation Coefficients for
Tasks and Paper-Pencil Ratings
N=33

	Task 1	Task 2	Task 3	Task Av	Rate 4	Rate 8
Task 1						
Task 2	.59 S=.001*					
Task 3	.37 S=.018	.27 S=.062				
Task Av	.83 S=.001	.77 S=.001	.73 S=.001			
Rate 4	.40 S=.011	.31 S=.04	.25 S=.079	.41 S=.009		
Rate 8	.36 S=.020	.29 S=.052	.24 S=.085	.38 S=.015	.99 S=.001	
Rate 16	.35 S=.023	.28 S=.058	.23 S=.096	.36 S=.019	.98 S=.001	1.00 S=.001

* S is significance level

selves were used and ordered in this manner:

0000; 1000, 0010, 0001, 0011; 1100, 0101, 1001, 0100;
1110, 0110, 0111, 1010; 1111, 1101, 1011

See Chapter 5 for a complete description of these ratings.

Correlations exceeding the .30 level were reported for Task 1 with all ratings, for Task 2 with Rate 4, for Task 3 with no ratings, for the task average with all ratings.

The test was assumed to have acceptable concurrent validity since the paper-pencil results reported as Rate 8 (reasoning levels and transition scores) had a Pearson correlation coefficient of .38 with the average task score which exceeded the minimum .30 level and was significant at the .015 level.

Construct Validity

According to Cronbach (1971), a test has construct validity if it measures the attribute it is said to measure. It follows then that if the test does not measure other things, it is acceptable. Comparison of pupil test performance was made with pupil task scores and with pupil intelligence scores measured with the Lorge-Thorndike verbal, nonverbal and total test.

The test had groups of questions for each of the successively more difficult levels. The observed pupil difficulty levels between groups of questions were compared.

It was assumed that construct validity would be evident in the convergence of scores of other measures of the same test. Correlations between task scores and the paper-pencil scores would be high, positive and higher than task score correlations with intelligence test scores.

The Pearson correlations using the scores of the thirty-five pupils participating in both task and paper-pencil testing were .36 between average task score and paper-pencil test rating, .53 between task scores and Lorge-Thorndike nonverbal IQ and .35 between task scores and Lorge-Thorndike verbal IQ. Although the correlation between task and paper-pencil scores was positive and high, it was exceeded by the value for task and nonverbal IQ correlation. It must be mentioned that the correlation between paper-pencil scores and Lorge-Thorndike nonverbal IQ was .58 and between paper-pencil scores and Lorge-Thorndike verbal IQ was .30. It is suspected that the high correlation with Lorge-Thorndike nonverbal is from some relationship with what is being measured and also from the continuous data provided by Lorge-Thorndike scores.

Additionally, it is a construct of Inhelder and Piaget (1958) that successive levels of proportional reasoning require progressively more sophisticated reasoning. Similarly, construct validation suggests that the difficulty level of items would be expected to show an increasing difficulty with higher levels of the test. This is illustrated in Figure XV.

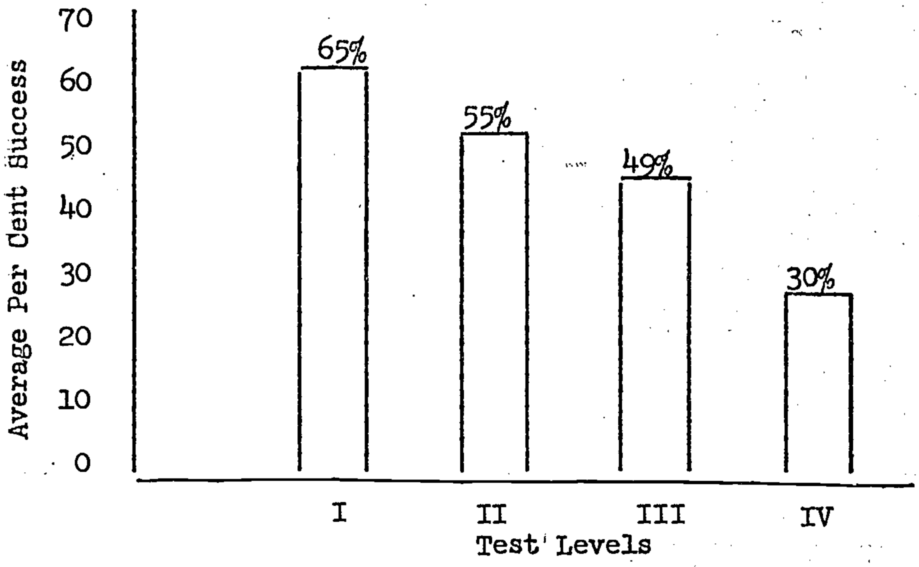


Figure XV. Average Per Cent Success of 427 Eighth Grade Pupils at the Four Test Levels

Further support for this difficulty construct was obtained by comparing the expected difficulty rank of items by group and the observed difficulty rank. It was expected that in each level all items would have identical ranking, that is $\frac{1+2+3+4+5+6}{6}$ for every item in Level I. The following array in Table 6.2 resulted.

Table 6.2

Comparison of Observed and Expected Item Difficulties (# Right)

	Test Item	Expected Rank	Observed Rank
Level I	1	3.5	4
	2	3.5	2
	3	3.5	1
	4	3.5	7
	5	3.5	6
	6	3.5	8
Level II	7	9.5	5*
	8	9.5	11
	9	9.5	3*
	10	9.5	19*
	11	9.5	14
	12	9.5	13
Level III	13	15.5	15
	14	15.5	20
	15	15.5	12
	16	15.5	10*
	17	15.5	17
	18	15.5	9*
Level IV	19	21.5	22
	20	21.5	18
	21	21.5	16
	22	21.5	23
	23	21.5	21
	24	21.5	24

* Items of evident discrepancy in rank order.

A measure of the continuity of this type of order is the Spearman rank correlation coefficient (Glass and Stanley, 1970) which for this array has a value of .87. This value suggests good construct validity in terms of difficulty rankings.

Discriminant Validity

A test has discriminant validity if it discriminates between the trait it measures and other traits. Evidence of discriminant validity was expected in smaller correlations of paper-pencil proportional reasoning scores with notebook averages than correlation of paper-pencil proportional reasoning scores with teacher-test scores. This should be evidenced also in smaller correlations of paper-pencil proportional reasoning scores with verbal IQ scores than with nonverbal IQ scores.

Pearson correlation coefficients with test rating (0, 1, 1.5, 2.0, 2.5, 3, 3.5, 4) were for small group average, .42; class test average, .60; notebook average, .22; verbal intelligence, .58; nonverbal intelligence, .64. These were all statistically significant at the .001 level.

Convergent Validity

A test has convergent validity if its measurement corresponds to other measurements of the same trait. Convergent validity would be evidenced in high positive correlations with other tests measuring the same trait. That is, correlations between task scores and paper-pencil scores should be high, positive and higher than those with intelligence scores.

Convergent validity would be evidenced in results that compare with the results of other researchers. That is, the proportion of persons measured to be formal operational should correspond to the proportions reported in the literature. There should be noted a positive correlation between proportional reasoning level and age (Inhelder and Piaget, 1958; Karplus and Peterson, 1970; Lawson, 1973; Hensley, 1974).

Convergent validity would be evidenced in the identity of components of proportional reasoning. That is, components of proportional reasoning should account for much of pupil achievement and intelligence. Pearson correlation coefficients with task scores for the thirty-five person sample taking both tests and tasks were: paper-pencil tests, .36; Lorge-Thorndike verbal, .35; Lorge-Thorndike nonverbal, .53.

The proportions of eighth grade pupils successful at each level reported in this test were: Level I, 77 per cent; Level II, 56 per cent; Level III, 36 per cent; and Level IV, 13 per cent. Corresponding values reported for a sample of 75 eighth to tenth grade pupils were: Levels I and II, 49% and Levels III and IV, 36 per cent (Karplus and Peterson, 1970). For a sample of 30 eighth grade pupils, the results were: Level I and below, 100 per cent; Level II, 70 per cent; Level III, 20 per cent and Level IV, one per cent (Hensley, 1974).

The correlation between test rating and age was found to be -.0498, which was not statistically significant at the .05 level.

The age correlation of other researchers cited was reported over ranges of ten to thirty years. The age range of the sample was about one year.

A principal components analysis identified two principal components. The first accounting for 44.8 per cent of the variance, the second 4.7 per cent. The first component loads heavily on measures of pupil achievement and intelligence. The test had acceptable convergent validity by these measures.

Summary of Validity

In summary, the test had high content validity, acceptable concurrent validity, good construct validity, high discriminant validity and acceptable convergent validity.

Reliability

Reliability is concerned with the fact that repeated measures should duplicate each other (Stanley, 1971). Measures of reliability center on the variability of response. In a criterion-referenced test, then reliability may have a special meaning. As a criterion for reliability, it was expected that the same person or comparable person taking the paper-pencil instrument or a comparable paper-pencil instrument should exhibit a comparable percentage of mastery. A classical one-form reliability measure (Hoyt, 1941) was calculated. Individual pupil scores and the total number of correct responses were used. The reliability coefficient, equivalent to the Kuder-Richardson Twenty value, was .78. Data and calculations of this are in Appendix C.

In a second approach, the criterion-referenced nature of the testing and the scoring by category were acknowledged and Livingston's (1972) approach was used.

This approach afforded a correction for the criterion level and the variance limitation of criterion-referenced testing. The relationship used was:

$$r_c = \frac{r_x \ \sigma_x^{2}(x) + (\bar{X} - C_x)^2}{{}^{2}(x) + (\bar{X} - C_x)^2}$$

where:

r_c = criterion-referenced reliability

r_x = classical measure of reliability (Hoyt, 1941)

σ^2 = variance of the test scores

$\bar{X}$ = mean of test scores

C = criterion level

The criterion-referenced reliability thus obtained (r_c) was .84, when the criterion level C was taken as 15. This was the level value for assignment of pupils to be either concrete or formal level proportional reasoners. Calculations may be found in Appendix C.

The reliability of the test, .84, compared favorably with other attempts, which ranged from .23 to .76, in the literature. Using Spearman-Brown split half measures, Lawson and Renner (1975) reported $r_H = .76$ for a biology reasoning level test, $r_H = .71$ for a chemistry reasoning level test and $r_H = .59$ for a physics reasoning level test. DeAvilla and Struthers (1967) used Cronbach's alpha measure of reliability and reported these results for a set of cartoon format paper-pencil tests: conservation, .694; causality, .550; relations, .001; logic, .227; and total test, .717.

Reliability was also measured on a test-retest basis and analyzed with the tetrachoric correlation coefficient and the Pearson correlation coefficient (Nie, et al., 1975). The tetrachoric measure (r_t) relates the reliability of the test to discriminate concrete and formal proportional reasoning levels. The Pearson correlation coefficient describes the relation of test-retest scores on the 24 test items.

The relationships were: $r_t = .40$ and $r = .68$ for a population of 94 fifth grade pupils; $r_t = .70$ and $r = .70$ for a population of 419 eighth grade pupils and $r_t = .32$ and $r = .47$ for a population of 149 eleventh grade chemistry pupils. Past testing had suggested that such fifth grade pupils would be largely non-masters of formal level proportional thinking, eighth grade pupils would be at the transitional stage between concrete and formal level proportional thinking and eleventh grade chemistry pupils would be masters of formal proportional thinking. In the manner suggested by Zeiky (1974), a sample of 338 fifth grade, eighth grade and chemistry pupils was randomly selected from those tested to comprise a sample of approximately equal numbers of probable non-masters, transitional and masters. This composite sample test-retest relationships were $r_t = .84$ and $r = .83$. Appendix C contains the calculation data for these values.

Summary of Reliability

In summary, the test has high reliability as a criterion-referenced test. This reliability supports its use as an excellent group measure of proportional reasoning and a good individual measure of proportional reasoning.

Item Difficulty

Piaget has described developmental levels of proportional reasoning (Inhelder and Piaget, 1958). The successive developmental levels require progressively more sophisticated reasoning. It was expected that the paper-pencil items would show increasing difficulty as the higher levels were measured. It was also expected that within a level item difficulties would be similar. Table 6.3 presents these item difficulties in terms of the percentage of grade eight pupils from Oak Grove Junior High School getting the item correct. There was increasing difficulty with higher levels as expected. The average percentage of pupils getting items correct by levels was: Level I, 65 per cent; Level II, 55 per cent; Level III, 49 per cent; and Level IV, 34 per cent.

Item Discrimination

It was expected that items selected for the test should demonstrate discrimination between masters and non-masters such that:

1) differences in percentages correct should be in agreement with the measured reasoning level of the pupils (see Appendix E)

Table 6.3

Item Difficulties in Terms of Performance for 427 Grade 8 Pupils

Level	Item in Final Test Version	Percentage Getting Item Correct	Average for Level
I	1	68	65%
	5	71	
	20	72	
	15	59	
	9	64	
	4	57	
II	21	67	55%
	10	55	
	18	69	
	14	35	
	8	50	
	2	53	
III	7	46	49%
	23	34	
	17	55	
	11	57	
	13	39	
	3	60	
IV	16	33	34%
	19	37	
	4	45	
	10	28	
	22	33	
	6	26	

2) r biserial values of .50 or above should be reported between masters and non-masters of items

3) item distractors selected by a pupil should match the pupil's reasoning level

Table 6.4 presents the percentage of correct item responses of pupils at five-proportional reasoning levels. The 0 level represents a pupil who was unsuccessful at achieving four or more

Table 6.4

Percentage of Correct Pupil Responses in Relation to Pupil Tested Reasoning Level

	Questions for						Questions for						Questions for						Questions for					
	Level I						Level II						Level III						Level IV					
All N=427	68	71	72	59	64	57	67	55	69	35	50	53	46	34	55	57	39	60	33	37	45	16	25	26
0000 Level 0 N=99	29	29	48	25	41	17	36	31	39	18	30	36	27	19	24	28	16	29	24	26	20	10	15	27
1000 Level I N=71	69	82	80	80	65	70	53	31	41	8	30	41	20	21	45	38	24	46	24	27	34	6	24	15
1100 Level II N=62	90	82	84	74	73	81	92	81	94	55	69	65	26	16	53	58	27	59	21	31	58	16	21	15
1110 Level III N=67	96	94	96	79	84	79	94	73	97	63	81	79	81	61	78	93	67	91	36	39	63	13	22	16
1111 Level IV N=23	100	100	96	83	87	96	100	91	100	65	74	83	91	65	83	87	83	100	91	91	83	57	57	70

correct responses at any of the four proportional reasoning levels: 1 - Concrete I, 2 - Concrete II, 3 - Form I, or 4 - Formal II. A Level I pupil achieved four or more correct responses at Level I but failed criterion achievement at other levels, 1000. A Level II pupil achieved four or more correct responses at both Levels I and II, but failed criterion achievement at Levels III and IV, 1100, and so on for Level III, 1110 and Level IV, 1111. The sharp discrimination across the level was evident at the line on the table separating the master and non-master levels. This line for questions in Level II shows that level respectively 53, 31, 41, 8, 30 and 41 per cent of Level I pupils correctly answered these questions while 92, 81, 94, 55, 69 and 65 per cent of Level II pupils respectively correctly answered them. Clearly the item collections were capable of discriminating the masters from the non-masters.

As an item discrimination index the biserial r correlation coefficient, r_{bis}, was calculated for each item. It was expected that these values would be .50 or greater. As reported in Table 6.5, only six of the twenty-four items failed to meet this criterion. Test items had good discrimination according to this measure.

Item design required that the key, or correct answer, and the distractors, or other answers, all be written at different reasoning levels. This was intended to make the correct answer and other answers appeal to persons at each reasoning level. Level IV items had answers appropriate to all four reasoning

Table 6.5

Item Discrimination

Level	Item	r Biserial	T Value	Significance 425 df
I	1	.5992	15.4292	< .001
	5	.5557	13.7778	
	20	.5673	14.2011	
	15	.5809	14.7110	
	9	.4420	10.1571	
	24	.6473	17.5075	
II	21	.5620	14.0085	< .001
	12	.5471	13.4731	
	18	.6057	15.6926	
	14	.4880	11.5266	
	8	.5061	12.0961	
	2	.4959	11.7713	
III	7	.5871	14.9497	< .001
	23	.4592	10.6555	
	17	.5352	13.0616	
	11	.5797	14.6676	
	13	.5291	12.8531	
	3	.5780	14.6031	
IV	16	.5584	13.8763	< .001
	19	.5317	12.9411	
	4	.4773	11.1979	
	10	.4527	10.4673	
	22	.5243	12.6943	
	6	.4595	10.6646	

levels as illustrated in the problem below:

19. A freeway driver keeps track of the distance he travels. He finds that in 4 minutes he travels 3 miles/ in 10 minutes $7\frac{1}{2}$ miles. If he continues at this speed, how long will it take him to travel 10 miles?

Distance	Time
3 miles	4 min.
$7\frac{1}{2}$ miles	10 min.
10 miles	? min.

A. About 13 minutes because $\frac{4 \text{ min.}}{3 \text{ miles}} = \frac{10 \text{ min.}}{7.5 \text{ miles}} = \frac{13\ 1/3 \text{ min.}}{10 \text{ miles}}$ — Level IV — Formal II

B. About 13 minutes because $10 - 7\frac{1}{2} = 2\frac{1}{2}$ miles and $10 + 2\frac{1}{2} = 12\frac{1}{2}$ min. — Level II — Concrete II

C. About 13 minutes because $\frac{4}{3} \times 10 = 13\ 1/3$ — Level III — Formal I

D. About 14 minutes because $7\frac{1}{2} + 3 = 10\frac{1}{2}$ and $10 + 4 = 14$ — Level I — Concrete I

E. I have no answer. — Level 0

A more complete discussion of this item design may be found in Chapter 5.

A cross tabulation was made of item responses with pupil levels for each item in Level IV. For item 19 the cross tabulation was that found in Table 6.6. In the table it may be read that for 58 pupils of Level III, four selected a Level 0 response, eight selected a Level I response, thirteen selected a Level II response, fifteen selected a Level III response and only eight selected a Level IV response.

These cross tabulations suggested that the item design worked. Pupils did select answers appropriate to their reasoning level. Table 6.7 shows that for only items four and six was this selection pattern not significant above the .001 level.

Table 6.6

Cross Tabulation of Pupil Response and Pupil Level for Item 19

Pupil	Response Level					Totals
Level	0	I	II	III	IV	
0	14	8	18	10	17	67
I	5	13	13	9	11	51
II	5	13	9	8	16	51
III	4	8	13	15	8	58
IV	0	1	1	1	23	26
Totals	28	43	54	43	85	253

Chi-square = 56.16 with 16 degrees of freedom
Significant at < .00001

Table 6.7

Cross Tabulation Significance for Level IV Items

Item in Final Test Version	Chi-square	Significance
16	56.465	< .0001
19	56.161	< .0001
10	52.159	< .0001
4	27.456	.0367
22	78.902	< .0001
6	39.668	.0055

Summary

The test instrument appeared to have high content validity and good construct validity. Reliability of the instrument was good. Items were excellent in their discrimination and generally appropriate in difficulty.

CHAPTER 7

CONCLUSIONS

Review of Purpose and Procedure

The purpose of this study was to develop a paper-pencil instrument to evaluate pupil proportional reasoning levels and to demonstrate how the application of principles of criterion-referenced test design could be used to build, validate and use such a test.

Individual task-testing of a representative group of forty pupils was used to establish a reference group for paper-pencil testing and to determine probable topics for test items. Paper-pencil testing of pupils who by reason of age were assumed to be non-masters, at the transitional stage, and masters was conducted. Analysis of item responses after each testing was used in item improvement. 2027 pupils were tested in arriving at the final test and the description of its characteristics. Five major revisions were made of the item sets comprising the test. The final test form consisted of twenty-four items with four subtests each of six items for Piaget levels Concrete Operational I, Concrete Operational II, Formal Operational I and Formal Operational II. The final test was completed by 90 per cent of the pupils in a 30-minute testing period.

The final test version was analyzed to describe the test characteristics. It was found that:

1) The paper-pencil test results correlated with the initial task results of a group of 35 pupils taking both tests. A value of .36 was obtained for the three task average and the final test scores.

2) Content, concurrent construct, divergent and convergent validity were established for the paper-pencil test. The test by all measures must be considered valid.

3) Reliability was assessed by the Kuder-Richardson-20 approach as modified by Hoyt. The reliability coefficient .77 suggested good reliability for the test. Reliability, calculated according to Livingston (1972) for criterion-referenced test, was .84. The .84 value suggested that the test had high reliability.

   Reliability calculated from test-retest results established a Pearson value of .83 for overall reliability and a value of .84 for the discrimination of formal and concrete levels.

4) Good item discrimination between proportional reasoning levels was established. The item design utilizing correct answers but different reasons was successful.

5) Pupil levels of proportional reasoning determined in the testing agree with those of other researchers (Hensley, 1974; Lawson, 1973; Karplus and Peterson, (1970). In contrast with Inhelder, Piaget's (1958) results, lower proportions of thirteen-year-olds were found to be formal operational in proportional reasoning in this study than in that of Piaget.

Educational Implications

The results of this study tended to confirm the study of Gray (1970) who found that paper-pencil measures of Piaget levels of cognitive development may be developed and that criterion-referenced test theory of Hambleton and Novick (1974) is effective in test design.

Efforts for paper-pencil tests of Piaget measures in other areas of cognitive development could be developed following the strategy used in this study. Control of variables, higher order proportions, causal relationships and functions are examples of areas certain to be of interest in science education.

The group test of this study and others like it should be used by teachers in evaluating the level of proportional reasoning in their classes. It has been expressed as a concern (Almy, 1973), that teachers recognize the level of thinking of their pupils.

Present science curricula, resulting from the activities of the sixties, do demand formal reasoning. The Piaget levels required in the science process skills are formidable (Wood, 1974).

This measurement tool and others developed in this manner should aid teachers in locating the level of their pupils' cognitive development. In an era where broad range achievement and intelligence tests are under criticism, such a specific measure would aid in diagnosis. The large scale testing possible with this paper-pencil instrument will support improvement in curricula, teaching stragegies and organization for instruction.

Curriculum design needs attention. Measures of pupil cognitive development are needed. Group testing with this test and others to determine both the range and mode of these levels wo"ld provide a solid base for curriculum design and would help in correcting past errors.

Limitations of the Study and Suggestions for Further Research

This study was limited to the development of a paper-pencil instrument to measure proportional reasoning in eighth grade pupils. Research is needed in the applicability of this instrument over a broad range of pupil ages. The original attention to reading level and empirical improvement of items would have to be repeated with large groups of pupils at the levels to be tested. Longitudinal studies of cognitive development with a group paper-pencil measure would then be possible.

The results of the study indicate that the test is a valid, reliable measure over the populations tested. Testing across other socioeconomic and cultural groups would extend the generality of the test. Some task testing to establish performance traits, additional items for item improvement would be necessary. The item improvement computer programs used in this study would support additional items for alternative selection.

This study was directed toward the development of a single paper-pencil instrument to measure proportional reasoning. Continued large scale use would allow the development of alternate forms through which further reliability measures could be made and curriculum research supported by pre-post testing with these alternate forms.

The proportional reasoning measure developed in this study should be complemented by the development of parallel measures including control of variables and logic. The test development strategy could follow that which proved to be successful in this study.

SELECTED BIBLIOGRAPHY

Adams, J., et al. (1975). Achievement of Minnesota Students in Mathematics. St. Paul: Minnesota Department of Education Office of Statewide Assessment.

Ahlgren, A. Remarks delivered at AAPT Convention, February 3, 1969, p. 2.

Airasian, P. W., and W. M. Bart. (1975). Validating a priori instructional hierarchies. Journal of Educational Measurement, 12:163-175.

Almy, M. (1964). Wishful thinking about children's thinking? In W. A. Fullagur, H. G. Lewis and C. F. Cumber, Readings for Educational Psychology. New York: Crowell. Pp. 389-401.

________. (1970). Longitudinal studies related to the classroom. In M. F. Rosskopf, et al., Piagetian Cognitive Development Research and Mathematical Education. Washington: National Council of Teachers of Mathematics. ED 077714.

Almy, M., E. Chittenden and P. Miller. (1966). Young Children's Thinking. New York: Teachers College Press.

Ausubel, D. (1965). Some psychological and educational limitations of learning by discovery. The Arithmetic Teacher, 12:290-302.

Ball, D. W. and S. A. Sayre. (1972). Relationships between student Piagetian cognitive development and achievement in science. Unpublished Ed. D. dissertation, University of Michigan, Ann Arbor.

Bart, W. M. (1971). The factor structure of formal operations. British Journal of Educational Psychology, 41:70-77.

________. (1972). Construction and validation of formal reasoning instruments. Psychological Reports, 30:663-670.

Beistel, D. W. (1975). A Piagetian approach to chemistry. Journal of Chemical Education, 52:151-152.

Besel, R. (1973). Using group performance to interpret individual responses to criterion-referenced tests. SWRL Professional Paper 25.

Bridgham, R. G. (1969). Classification, seriation and learning of electrostatics. Journal of Research in Science Teaching, 6:118-127.

Carpenter, T. P., et al. (1975a). Notes from national assessment: basic concepts of area and volume. The Arithmetic Teacher, 22:501-507.

________. (1975b). Results and implications of the NAEP mathematics assessment: secondary school. The Mathematics Teacher, 68:6.

Carver, R. P. (1970). Special problems in measuring charge with psychometric devices. In Evaluative Research: Strategies and Methods. Pittsburgh: American Institute for Research.

Chittenden, E. A. (1974). Personal conversation at Educational Testing Service, Princeton, February, 1974.

Clemenson, R. W. (1970). A comparative study of three fifth grade classrooms on five selected Piaget type tasks dealing with science related concepts. Unpublished Ph. D. dissertation, University of Iowa.

Copeland, R. W. (1974). How Children Learn Mathematics. New York: Macmillan.

Cronbach, L. J. (1960). Essentials of Psychological Testing (2nd ed.). New York: Harper. Pp. 23, 25.

________. (1971). Test validation. In R. L. Thorndike, ed., Educational Measurement. Washington: American Council on Education.

Dale, L. G. (1970). The growth of systematic thinking: replication of Piaget's first chemical experiment. Australian Journal of Psychology, 22:277-286.

Darley, J. G. and G. V. Anderson. (1951). The functions of measurement in counseling. In E. F. Lindquist, ed., Educational Measurement. Washington: American Council on Education. Pp. 68-84.

DeAvilla, E. and J. A. Struthers. (1967). Development of a group measure to assess the extent of pre-logical and pre-causal thinking in primary school age children. Paper presented at the 1967 Annual Convention of the National Science Teachers Association. ED 019136.

DeStefano, J. (1973). Linguistics and logical reasoning. Theory into Practice, 12(5):272-277.

DeVries, R. (1973a). Relationships among Piagetian levels, achievement and intelligence. Paper presented at the American Educational Research Association Meeting, New Orleans, March 1, 1973. ED 079101.

________. (1973b). The two intelligences of bright, average and retarded children. Paper presented at the Biennial Meeting of the Society for Research in Child Development, Philadelphia, March 29, 1973. ED 079102.

Easley, J. A. (1974). The structural paradigm in protocol analysis. Journal of Research in Science Teaching, 11:281-290.

Ebel, R. L. (1971). Criterion-referenced measurements: limitations. School Review, 69:282-288.

Elkind, D. (1961). Quantity conceptions in junior and senior high school students. Child Development, 32:551-560.

________. (1962). Quantity conceptions in college students. The Journal of Social Psychology, 57:459-465.

________. (1975). Piaget. Human Behavior, 4:25-31.

Emrick, J. A. (1971). An evaluation model for mastery testing. Journal of Educational Measurement, 8:4.

Fehr, H. F. (1974). The secondary school mathematics curriculum improvement study: a unified mathematics program. The Mathematics Teacher, 67:25-30.

Flavell, J. H. (1963). The Developmental Psychology of Jean Piaget. Princeton: D. Van Nostrand.

Fremer, J. (1972). Criterion-referenced interpretations of survey achievement tests. Test Development Memorandum. Princeton: Educational Testing Service.

Ginsburg, H. and S. Opper. (1969). Piaget's Theory of Intellectual Development. Englewood Cliffs, N.J.: Prentice Hall.

Glaser, R. (1963). Instructional technology and the measurement of learning outcomes. American Psychologist, 18:519-521.

Glaser, R. and R. C. Cox. (1968). Criterion-referenced testing for the measurement of educational outcomes. In R. Weisgerber, ed., Instructional Process and Media Innovation. Chicago: Rand McNally. Pp. 545-550.

Glaser, R. and A. J. Nitko. (1971). Measurement in learning and instruction. In R. L. Thorndike, ed., Educational Measurement. Washington: American Council on Education. Pp. 625-670.

Glass, C. V. and J. C. Stanley. (1970). Statistical Methods in Education and Psychology. Englewood Cliffs, N.J.: Prentice Hall.

Goodyear, J. and J. Renner. (1975). The multiple-choice test in the science classroom. The Science Teacher, 42:32-34.

Grant, N. and J. Renner. (1975). Identifying types of thought in tenth grade biology pupils. The American Biology Teacher, 37:283-286.

Gray, W. M. (1970). Children's performance on logically equivalent Piagetian tasks and written tasks. Doctoral thesis. Ann Arbor: University Microfilms.

Green, B. F. (1956). A method of scalogram analysis using summary statistics. Psychometrica, 21:79-88.

Green, D. R., M. P. Ford and G. B. Flamer, eds. (1971). Measurement and Piaget. New York: McGraw Hill.

Guttman, L. (1944). A basis for scaling qualitative data. American Sociological Review, 10:255-282.

________. (1947). Cornell scale and intensity analysis. Educational and Psychological Measurements, 7:247-279.

Hall, V. and R. Kingsley. (1968). Conservation and equilibration theory. Journal of Genetic Psychology, 111:195-213.

Hambleton, R. K. and M. R. Novick. (1972). Toward an integration of theory and method for criterion-referenced tests. American College Testing Program Research Report 55. Iowa City: American College Testing Program.

Harris, C. W. (1972). An interpretation of Livingston's reliability coefficient for criterion-referenced tests. Journal of Educational Measurements, 9:27-29.

Hensley, J. H. (1974). An investigation of proportional thinking in children from grades six through twelve. Unpublished doctoral thesis, University of Iowa.

Herron, J. D. (1975). Piaget for chemists. Journal of Chemical Education, 52:146-150.

Hieronymus, A. N. (1971). Today's testing: what do we know how to do. Address, American Educational Research Association Meeting, Minneapolis.

Higgins-Trenk, A. and A. J. H. Gaite. (1971). The elusiveness of formal operational thought in adolescents. Paper presented at 79th meeting of the American Psychological Association, Washington, D.C., September 4. ED 063972.

Hively, W., H. L. Patterson and S. A. Page. (1968). Universe defined system of arithmetic achievement tests. Journal of Educational Measurement, 5:275-290.

Holloway, G. E. T. (1967). An Introduction to the Child's Concept of Geometry. New York: Humanities Press.

Howe, A. (1974). Formal operational thought and the high school science curriculum. Paper presented to the NARST, Chicago, April, 1974. ED 092364.

Hoyt, C. J. (1952). Estimation of test reliability for unrestricted item scoring methods. Educational and Psychological Measurements, 12:756-758.

Inhelder, B. and J. Piaget. (1958). The Growth of Logical Thinking from Childhood to Adolescence. New York: Basic Books.

________. (1969). The Early Growth of Logic in the Child: Classification and Seriation. New York: Norton.

Ivens, S. H. (1970). An investigation of item analysis, reliability and validity in relation to criterion-referenced tests. Unpublished doctoral dissertation, Florida State University.

Jackson, S. (1965). The growth of logical thinking in normal and subnormal children. British Journal of Educational Psychology, 35:255-258.

Jensen, J. (1973). A comparative investigation of the casual and careful oral language styles of average and superior fifth grade boys and girls. Research in the Teaching of English, 7:223-250.

Karplus, E. and R. Karplus. (1970). Intellectual development beyond elementary school. I: deductive logic. School Science and Mathematics, 70:398-406.

Karplus, R. and E. Karplus. (1972). Intellectual development beyond elementary school. III: ratio, a longitudinal study. School Science and Mathematics, 72:735-742.

________. (1974). Proportional reasoning and control of variables. Unpublished paper. Cambridge: Massachusetts Institute of Technology.

Karplus, R. and R. Peterson. (1970). Intellectual development beyond elementary school. II: ratio, a survey. School Science and Mathematics, 70:813-820.

Karplus, R., E. Karplus and W. Wollman. (1974). Intellectual development beyond elementary school. IV: ratio, the influence of cognitive style. School Science and Mathematics, 74:476-482.

Kaufman, B. A. and R. Konicek. (1974). The application of Piaget to contemporary curriculum reform. Paper presents to the National Association for Research in Science Teaching, 47th Annual Meeting, Chicago, April, 1974.

Kavanagh, D. C. (1974). An investigation of a model hierarchy for the acquisition of the concept of speed. Paper presented to the National Association for Research in Science Teaching, Annual Meeting, Chicago, April, 1974.

Keasy, C. (1971). The nature of formal operations in preadolescence, adolescence and middle age. Unpublished doctoral dissertation, University of California, Berkeley.

Kohlberg, L. and C. Gilligan. (1971). The adolescent as a philoscpher. Daedalus, 100:1051-1086.

Kriewall, T. E. (1969). Applications of information theory and acceptance sampling principles to the management of mathematics instruction. Unpublished doctoral dissertation, University of Wisconsin.

Kulm, G. (1973). Sources of reading difficulty in elementary algebra textbooks. Mathematics Teacher, 66:649-652.

Laurandeau, M. and A. Pinard. (1962). Causal Thinking in the Child. New York: International University Press.

Lawson, A. E. (1973). Relationships between concrete and formal operational science subject matter and the intellectual level of the learner. Unpublished doctoral dissertation, University of Oklahoma.

________. (1974) Relationships of concrete and formal operational science subject matter and the developmental level of the learner. Paper presented at the National Association of Research in Science Teaching Convention, April, 1974.

Lawson, A. E. and J. W. Renner. (1975). Relationships of science subject matter and developmental levels of learners. Journal of Research in Science Teaching, 12:347-350.

Lawson, A. E., F. H. Nordland and A. DeVito. (1975). Relationship of formal reasoning to achievement, aptitudes and attitudes in preservice teachers. Journal of Research in Science Teaching, 12:423-431.

Linn, M. and H. Thier. (1975). The effect of experiential science on development of logical thinking in children. Journal of Research in Science Teaching, 12:49-62.

Livingston, S. A. (1972). Criterion-referenced applications of classical test theory. Journal of Educational Measurement, 9:13-28.

Lovell, K. (1961). A follow-up study of Inhelder and Piaget's the growth of logical thinking. The British Journal of Psychology, 52:143-153.

________. (1970). Proportion and probability. In M. F. Rosskopf, et al., Piagetian Cognitive Development Research and Mathematics Education. Washington: National Council of Teachers of Mathematics.

Lovell, K. and I. B. Butterworth. (1966). Abilities underlying the understanding of proportionality. Mathematics Teaching, 37:5-9.

Lovell, K. and J. B. Shields. (1967). Some aspects of a study of the gifted child. British Journal of Educational Psychology, 37:201-208.

Lunzer, E. A. (1965). Problems of formal reasoning in test situations. In P. H. Mussen, ed., Monographs of the Society for Research in Child Development, European Research in Cognitive Development, 30:19-46.

Lunzer, E. A. and P. Pumfrey. (1966). Understanding proportionality. *Mathematics Teaching*, 34:7-12.

Lunzer, E. A., C. Harrison and M. Davey. (1972). The four-card problem and the generality of formal reasoning. *Quarterly Journal of Psychology*, 24:326-339.

McCormack, A. J. and R. V. Bybee. (1970). Piaget and the training of elementary science teachers. Address at NSTA Convention, Cincinnati, Ohio, March 12-16.

McKinnon, J. W. and J. W. Renner. (1971). Are colleges concerned with intellectual development? *American Journal of Physics*, 39:1047-1052.

McLeod, R., G. Berkheimer, D. Fyffe and R. Robison. (1975). The development of criterion-validated test items for four integrated science processes. *Journal of Research in Science Teaching*, 12:415-421.

Mallon, E. J. (1976). Cognitive development and processes: review of the philosophy of Jean Piaget. *The American Biology Teacher*, 38:28-33.

Mandell, A. (1974). *The Language of Science*. Washington: National Science Teachers Association.

Margenau, H. (1950). *The Nature of Physical Reality*. New York: McGraw-Hill.

Mehrens, W. A. and I. J. Lehman. (1972). *Measurement and Evaluation in Education and Psychology*. New York: Holt.

Meyers, S. S. (1970). Questions illustrating the kinds of thinking required in current mathematics tests. Princeton: Educational Testing Service.

Mink, O. G. (1964). Experience and cognitive structure. In R. E. Ripple and V. N. Rockcastle, eds., *Piaget Rediscovered*. Ithaca, N.Y.: Cornell University.

Mogar, M. (1960). Children's causal reasoning about natural phenomena. *Child Development*, 31:59-65.

Nisbet, J. D., et al. (1964). Puberty and test performance, *British Journal of Educational Psychology*, 34:202-203.

Nitko, A. J. (1974). Problems in the development of criterion-referenced tests: the IPI Pittsburgh experience. In Harris, Alkin and Popham, eds., CSE Monograph Series in Evaluation. Los Angeles: Center for the Study of Evaluation.

Nitko, A. J. and T. Hsu. (1974). Using domain-referenced tests for student placement, diagnosis and attainment in a system of adaptive, individualized instruction. Educational Technology, 14:48.

Novak, J. D. (1974). Summary of science education research. A paper presented at the 1974 NARST Convention, Chicago.

Osborne, A. R. (1973). Promoting logical ability. Theory into Practice, 12:286-291.

Osiki, K. J. (1974). A comparison of affective and cognitive development in elementary school students. A paper presented at the 1974 NARST Convention, Chicago.

Phillips, D. R. (1974). Formal operational thought and dogmatism. Paper presented to the National Association for Research in Science Teaching, 47th Annual Meeting, April, 1974, Chicago.

Phillips, D. G. (1974). Changing teachers' perception of "learning": an application of Piaget's theory and experiments. Address at the National Association for Research in Science Teaching, 47th annual meeting, April, 1974, Chicago.

Piaget, J. (1926). The Language and Thought of the Child. London: Kegan Paul.

________. (1964). Development and learning. In R. E. Ripple and V. N. Rockcastle, eds., Piaget Rediscovered. Ithaca, N.Y.: Cornell University. Pp. 7-20.

________. (1970). The Child's Concept of Motion and Speed. New York: Ballantine Books.

________. (1972). Intellectual evaluation from adolescence to adulthood. Human Development, 15:1-12.

Piaget, J. and B. Inhelder. (1963). The Child's Conception of Space. London: Routeledge & Kegan Paul.

Piaget, J. and B. Inhelder. (1969). The Psychology of the Child. New York: Basic Books.

________. (1971). Mental Imagery in the Child. London: Routeledge & Kegan Paul.

Popham, W. J. and T. R. Husek. (1969). Implications of criterion-referenced measurement. Journal of Educational Measurement, 6:1-9.

Raven, R. J. (1972). A multivariate analysis of task dimensions related to science concept learning difficulties in primary school children. Journal of Research in Science Teaching, 9:207-221.

________. (1974). Programming Piaget's logical operations for science inquiry and concept attainment. Journal of Research in Science Teaching, 11:251-261.

Reichard, S., M. Scheiden and D. Rapaport. (1944). The development of concept formation in children. American Journal of Orthopsychiatry, 14:152-162.

Renner, J. W. and A. E. Lawson. (1973). Piagetian theory and instruction in physics. Physics Teacher, 11:165-169.

Ripple, R. E. and V. N. Rockcastle, eds. (1964). Piaget. In Piaget Rediscovered. Ithaca, N.Y.: Cornell University.

Robertson, W. W. and E. Richardson. (1975). The development of some physical science concepts in secondary school students. Journal of Research in Science Teaching, 12:319-329.

Rosskopf, M. F., et al. (1970). Piagetian Cognitive Development Research and Mathematics Education. Washington: National Council of Teachers of Mathematics. ED 077714.

Rowell, J. A. and P. J. Hoffman. (1975). Group tests for distinguishing formal from concrete thinkers. Journal of Research in Science Teaching, 12:157-164.

Sayre, S. A. and D. W. Ball. (1975). Piagetian cognitive development and achievement in science. Journal of Research in Science Teaching, 12:147-156.

Shepler, J. L. (1969). A Study of Parts of the Development of a Unit on Probability and Statistics for the Elementary School. Research and Development Center for Cognitive Learning, Report No. 105. Madison: University of Wisconsin.

Smeslund, J. (1964). Internal necessity and contradiction in children's thinking. In R. E. Ripple and V. N. Rockcastle, eds., *Piaget Rediscovered*. Ithaca, N.Y.: Cornell University.

Stanley, J. C. (1971). Reliability. In R. L. Thorndike, ed., *Educational Measurement*. Washington: American Council on Education.

Steffe, L. P. and R. B. Parr. (1968). *The Development of the Concept of Ratio and Fraction in the Fourth, Fifth and Sixth Years of Elementary School*. Research and Development Center for Cognitive Learning, Report No. TR-49. Madison: University of Wisconsin.

Strauss, S. (1972). Learning theories of Gagne and Piaget: implications for curriculum development. *Teachers College Record*, 74:81-102.

Sund, R. B. and L. W. Trowbridge. (1973). *Teaching Science by Inquiry in Secondary Schools*. Columbus, Ohio: Merrill.

Towler, J. and G. Wheatley. (1971). Conservation concepts in college students, a replication and critique. *The Journal of Genetic Psychology*, 118:265-270.

Trowbridge, L. (1974). Trends and innovations in junior high science teaching in the United States. *The Science Teacher*, 41:12-15.

Tuddenham, R. D. (1971). Theoretical regularities and individual idiosyncrasies. In D. R. Green, M. P. Ford and G. B. Flamer, eds., *Measurement and Piaget*. New York: McGraw Hill.

Webb, R. A. (1974). Concrete and formal operations in very bright six to eleven year olds. *Human Development*, 17:292-300.

While, R. (1974). Indexes used in testing the validity of learning hierarchies. *Journal of Research in Science Teaching*, 11:1, 61-66.

Wohlwill, J. F. (1960). A study of the development of the number concept by scalogram analysis. *Journal of Genetic Psychology*, 97:345-377.

________. (1968). Responses to class-inclusion questions for verbally and pictorially presented items. *Child Development*, 39:449-465.

Wollman, W. and R. Karplus. (1974). Intellectual development beyond elementary school. V: using ratio in differing tasks. School Science and Mathematics, 75:593-613.

Wood, D. A. (1974). The Piaget process matrix. School Science and Mathematics, 74:407-411.

Woodson, M. I. (1974). The issue of item and test variance for criterion-referenced tests. Journal of Educational Measurement, 11:63-64.

Zeiky, M. J. (1974). Methods of setting standards for criterion-referenced item sets and applications and adaptations of classical test theory for application to criterion-referenced measures. An address to the Conference on Criterion-Referenced Testing, Princeton.

APPENDIX A

Pilot Study Results and Calculations

Pearson Correlations between Pilot Task Scores and Written Test and Intelligence Test Scores

Pupil	Task	Paper	Lorge-Thorndike Verbal	 Nonverbal	 Total
1	1.8	1.96	89	97	93
2	3.0	1.40	118	121	120
3	3.6	3.53	-	-	-
4	.6	1.60	75	65	70
5	1.8	3.48	128	142	135
6	3.2	2.48	111	130	121
7	2.8	2.41	108	138	123
8	1.6	2.32	86	101	94
9	3.6	2.54	118	136	127
10	.8	.95	70	85	78
11	3.0	1.88	107	106	107
12	3.0	-	103	121	112
13	3.2	2.16	116	119	118
14	1.0	-	88	97	93
15	3.6	-	-	-	-
16	3.2	2.36	101	105	103
17	2.6	2.24	103	111	107
18	1.4	2.56	81	90	86
19	3.6	1.88	104	108	106
20	2.6	-	84	97	91
21	2.8	3.04	114	130	122
22	3.6	3.76	145	127	136
23	2.4	3.33	111	117	114
24	2.2	2.56	109	120	115
25	2.0	3.12	109	112	111

	N	Σx	$\Sigma^2 x$	Σy	$\Sigma^2 y$	Σxy	$\bar{x}$r	$\bar{y}$	r
Task/Paper	21	52.8	149.6	51.6	137.3	134.2	2.51	2.46	.35
Task/Verbal	23	55.8	153.4	2378	252664	6017	2.43	103	.709
Task/Nonverbal	23	55.8	153.4	2575	295933	6494	2.43	112	.665
Task/Total	22	55.8	153.4	2482	274452	6269	2.43	108	.713
Paper/Total	20	48.0	124.8	2186	244978	5404	2.40	109	.646

Relationships between Task, Paper-Pencil
and Intelligence Test Scores

			Lorge-Thorndike	
Pupil	Task Av.	Paper-Pencil	Non-verbal	Verbal
1	2.3	1.00	111	110
2	3.7	2.25	135	124
3	3.3	2.50	126	108
4	2.3	1.00	124	117
5	2.3	4.00	126	97
6	2.0	2.25	133	111
7	1.3	1.00	97	109
8	1.7	0.00	109	112
9	2.7	0.67	121	118
10	3.7	2.25	121	101
11	1.7	3.00	123	115
12	1.0	1.00	97	93
13	2.0	1.25	88	79
14	4.0	0.00	115	122
15	2.7	2.00	125	117
16	2.7	1.00	113	94
17	1.3	0.00	99	86
18	2.3	0.00	90	90
19	2.3	1.25	111	111
20	2.3	1.00	106	97
21	1.7	2.00	98	106
22	2.3	2.00	105	104
23	3.0	3.00	106	122
24	3.0	3.00	110	120
25	2.3	3.00	126	118
26	1.0	0.00	86	92
27	3.0	3.25	137	120
28	3.3	2.00	129	119
29	2.0	3.50	123	126
30	1.7	0.00	115	106
31	1.3	0.00	82	103
32	2.0	2.50	130	121
33	2.3	1.75	132	98
34	1.7	0.00	121	114
35	2.0	0.00	91	102

	N	Σx	$\Sigma^2 x$	Σy	$\Sigma^2 y$	Σxy	$\bar{x}$	$\bar{y}$	r
Task/Paper	35	80.2	203	53.4	131	133	2.29	1.53	.36
Task/Nonverbal	35	80.2	203	3961	456265	9285	2.29	113	.53
Task/Verbal	35	80.2	203	3677	402276	8623	2.29	105	.35
Paper/Nonverbal	35	53.4	131	3961	456265	6413	1.53	113	.58
Paper/Verbal	35	53.4	131	3683	401531	5874	1.53	105	.30

APPENDIX B

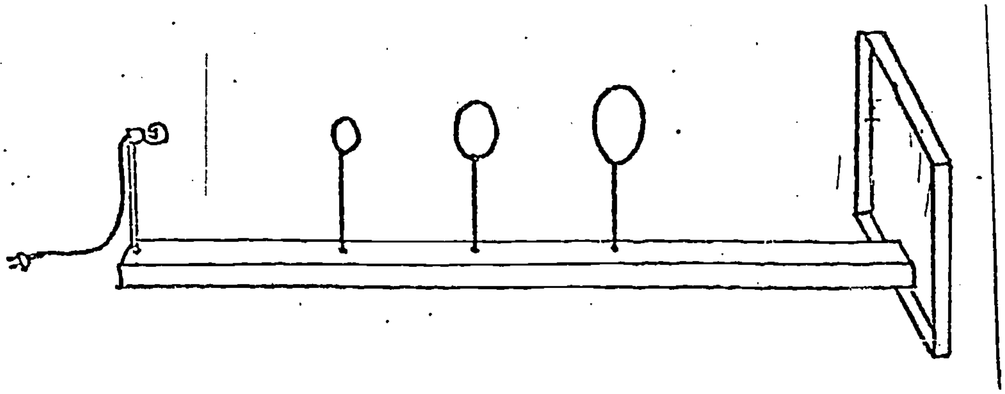
Task Interview Protocols

1. Projection of Shadows (Hensley, 1974)

Thinking tested

Schema of proportions
Inverse proportions - physical

Material



A screen, 30 cm x 30 cm, is used to observe the shadows. The shadows are made by three wire rings, 3.0 cm, 6.0 cm, and 9.0 cm in diameter. Each ring has a support wire. The length of the support wire is such that the center of each ring is 12.5 cm above the bottom of the support wire. The rings are made from different colors of wire as follows: 3.0 cm (white), 6.0 cm (red), 9.0 cm (black). The rings are held vertically on a meter stick by optic bench screen holders. The meter stick has only marks at each 10 cm length. Each mark is labeled with the following letters: N, R, M, K, G, F, A, B and O. A clear light bulb is supported at one end of the beam. The center of the bulb is 12.5 cm above the top of the beam. The light is turned on and off by connecting or disconnecting the cord to the 6 volt battery. One meter stick marked in centimeters and millimeters is provided for the student to use.

Introduction

"Here is a board, a light and a screen. I can put up one ring (6.0 cm) on the board (at 50 cm) and then when I turn on the light (do it), I get a shadow of the ring on the screen."

Question

Initially seek out predictions of the effects of ring size and ring position on the shadow with questions such as: "What would you predict will happen if I use this smaller (3.0 cm) ring?" "What else could change the size of the shadow?" "How?" Do what is suggested.

Culminating Question

"How might I make just one shadow using two rings? Explain why this works?"

Scoring Criteria

Stage	Criteria	Score
I	The subject represents the shadow in the way the object appears to him. He does not perceive how the shadow is formed on the screen.	0
IIA	The subject recognizes that the size of the shadow depends on the size of the object. His knowledge goes no further.	1
IIB	In addition to the ring-size dependence of the shadow demonstrated in IIA, the subject suggests qualitatively that the distance affects the shadow size, the closer the object is to the screen, the smaller the shadow.	2
IIIA	The subject quantitatively compensates between distance and shadow size, between distance and diameter, but is not generalized as a rule. The subject begins to measure distance from the light source.	3

Scoring Criteria (continued)

Stage	Criteria	Score
IIIB	From the start the subject measures both the distance from the light source and the diameter of the rings. He looks for a numerical hypothesis based on the divergent structure of the light rays. The subject is able to state in a numerical form the general relation for the two rings to have just one shadow.	4

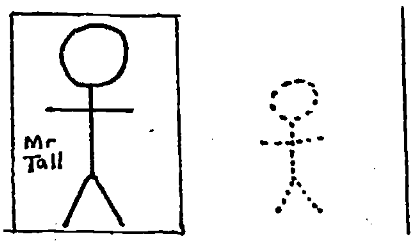
2. Mr. Short and Mr. Tall (Karplus and Karplus, 1970)

Thinking tested

Schema of proportions
Direct proportion - geometric

Material

Paper sketch of Mr. Tall
Large paper clips
Small paper clips
Chart



	Biggies	Smallies
Mr. Tall		
Mr. Short		

	Big	Small
Mr. Tall	3	2
Mr. Short	2	

Introduction

"I have here a picture I call Mr. Tall. He measures about 3 big paper clips, that is, biggies from head to toe." Measure and write on chart. "Mr. Small, whom I don't have here, looks just like Mr. Tall but Mr. Small measures just 2 biggies from head to toe." Write on chart.

Question

"Measure Mr. Tall in small paper clips (smallies) and then predict what height Mr. Small would be if you could measure him in smallies? Explain how you got your answer."

Scoring Criteria

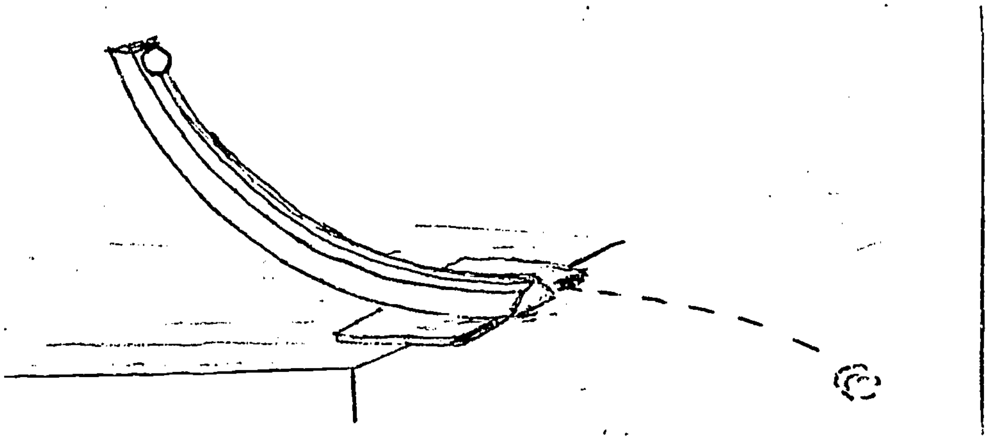
Stage	Criteria	Score
I	Subject guesses, gives answers with no compensations.	0
IIA	Subject qualitatively compensates, "It should be smaller" with no rule.	1
IIB	Subject compensates through inappropriate but consistent addition or subtraction. "It was 2 biggies less so it's 2 smallies less."	2
IIIA	Subject quantitatively compensates. Subject works through some multiple or a multiplication factor.	3
IIIB	Subject states a proportion with numbers in his solution.	4

3. Sled (Piaget, 1970)

Thinking tested

Proportional reasoning
Direct as square
Physical

Material



A 30 cm grooved ruler with a steel backing mounted so that marbles may be rolled down it. Electric stop watch.

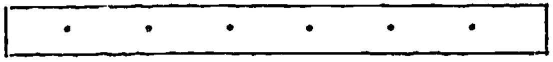
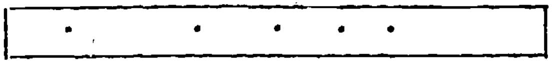
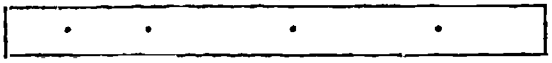
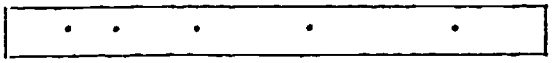
Introduction

"Imagine that this is a hill on which you are sledding and you start at the top and go down like this marble (let the marble roll down chute, have watch running). Imagine you had a watch."

Question

"Suppose, as you called out, each second as you went down the hill someone placed a flag just where you were at that time. Sketch how the flags would be separated. Explain how you got your answer."

Scoring Criteria

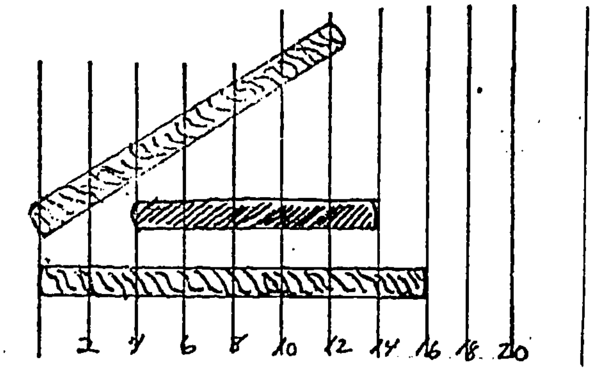
Stage	Criteria	Score
I	Subject's pattern is erratic or he has no pattern	0
IIA	Subject's pattern illustrates some notion of speed 	1
IIB	Subject shows some kind of acceleration but without a constant pattern 	2
IIIA	Subject's pattern relates constant acceleration 	3
IIIB	Subject's pattern relates constant acceleration and subject states an overall rule. "All the time you would go faster and faster." 	4

4. Angle

Thinking tested

Proportional reasoning
Direct proportions
Geometric

Material



Two rods are laid out perpendicular to a numbered measuring grid. The orange rod is 16 units long, the yellow rod is 10 units long. Then the orange rod is turned to another angle.

Introduction

"You can see the orange rod measures 16 units. The yellow rod measures 10. Now, if I turn the orange one, it will cover 12 units."

Question

"Can you predict how many units the yellow rod would cover if I moved it to the same angle? Explain how you got your answer."

Scoring Criteria

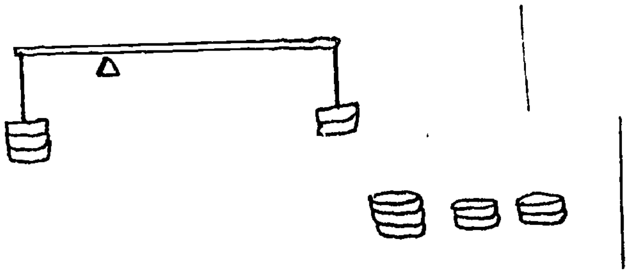
Stage	Criteria	Score
I	Subject guesses. The answer has no support - "looks like it."	0
IIA	Subject qualitatively compensates. "It should be smaller."	1
IIB	Subject compensates quantitatively through addition or subtraction. "Subract." Go back 6.	2
IIIA	Subject quantitatively compensates using some multiplication or fraction. It should be less than 6 difference.	3
IIIB	Subject refers to a general solution. It is proportional. The proportion 10/16 is the same as 5/8.	4

5. Balance

Thinking tested

Proportional reasoning
Direct proportion
Physical

Materials



A light, unequal arm balance has hooks for weights and there are 7-10 identical weights available.

Introduction

"Two weights just balance three on the other side. If I add two more on the right, I will have 4 weights."

Question

"Can you predict how many I will have to add on the left to balance again? How did you get your answer?"

Scoring Criteria

Stage	Criteria	Score
I	Subject guesses or has no answer	0
IIA	Subject compensates qualitatively	1
IIB	Subject compensates using some addition or subtraction 6 - Add up	2

Scoring Criteria (continued)

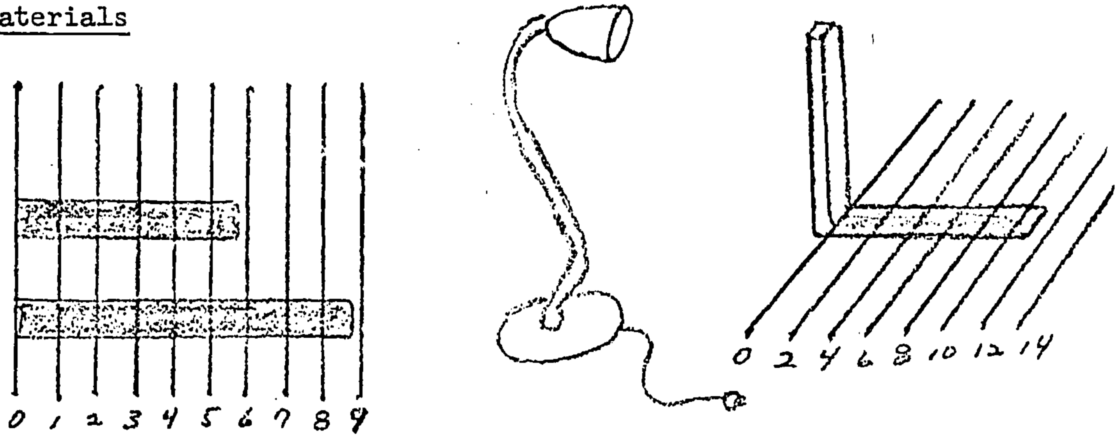
Stage	Criteria	Score
IIIA	Subject uses a ratio or multiplication factor 2=3 so 4=6	3
IIIB	Subject uses an appropriate proportion and states some rule: 1 big thing = 3 small ones 3 big things = 9 small ones	4

6. Flag Pole

Thinking tested

Proportional reasoning
Direct proportion
Physical

Materials



Two rectangular wooden beams are laid out on a measuring grid. A high intensity light source is arranged to produce shadows.

Introduction

"The green rod you can see is about 8 units long. The blue one is about 5. When I set up the blue rod and the lamp, the rod has a shadow 10 units long."

Question

"Predict the number of units of shadow I would get if I set up the green rod in the same way without moving the lamp. How did you get your answer?"

Scoring Criteria

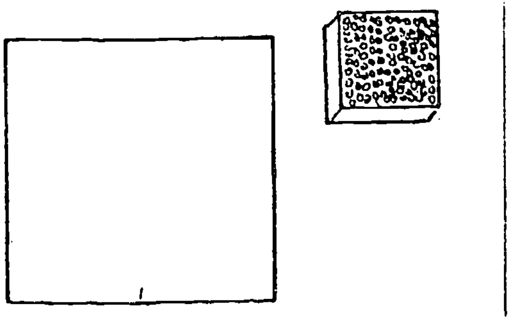
Stage	Criteria	Score
I	No answer or a guess	0
IIA	Subject qualitatively compensates "13 It's smaller"	1
IIB	Subject uses subtraction for a more quantitative compensation "4 I just subtracted"	2
IIIA	A ratio or multiplication factor is used 5/8 = 10/16	3
IIIB	An appropriate proportion is used and a rule stated "The short one is half as tall so the shadow will be half as tall."	4

7. BB Square

Thinking tested

Proportional reasoning
Direct as square
Geometric

Material



A square 2 units on edge, a square 3 units on edge, and a ruler are set out before the subject. The larger square has a small edge so that it may be covered with BBs.

Introduction

"It takes just 140 BBs to cover this small square." Do it.

Question

"Predict how many BBs would be needed to cover the large square. How did you get your answer?"

Scoring Criteria

Stage	Criteria	Score
I	Subject has no answer or guesses	0
IIA	Subject qualitatively compensates "10 because it's less"	1

Scoring Criteria (continued)

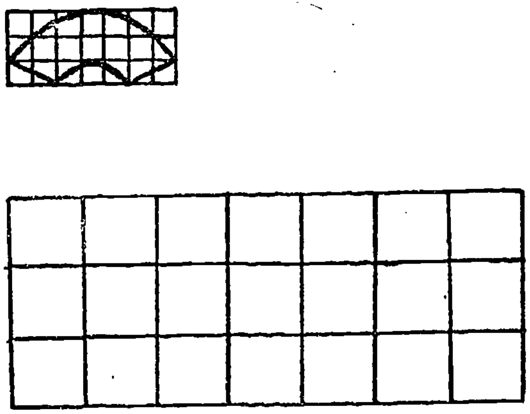
Stage	Criteria	Score
IIB	Subject uses addition to compensate 2 + 1 = 3 140 + 70 = 210	2
IIIA	Subject uses a ratio or a multiplication factor 3/2 = X/140	3
IIIB	Subject uses appropriate proportion employing some rule 9/4 = X/140 About 300. Because it's the area.	4

8. Pattern

Thinking tested

Proportional reasoning
Direct as square proportion
Geometric

Material



A pattern type drawing and a larger grid are presented to the subject.

Introduction

"A small doll sized collar made with the pattern shown uses 12 square centimeters of material."

Question

"How much material is there when I make a collar like this from a pattern drawn on these larger squares?" How did you get your answer?"

Scoring Criteria

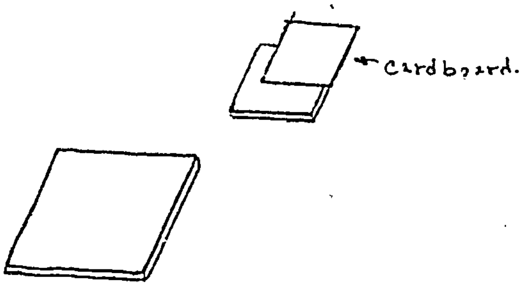
Stage	Criteria	Score
I	Subject guesses or has no answer	0
IIA	Subject qualitatively compensates "20 because it's bigger"	1
IIB	Subject uses addition as a quantitative compensation "36 because 12+12+12=36"	2
IIIA	Subject uses multiplication or a ratio "3x3=9 1/9 = 12/81"	3
IIIB	Subject uses an overall rule "It should be 3 x 3 as much because it goes up as length x width"	4

9. Frosting

Thinking tested

Proportional reasoning
Inverse as square
Geometric

Material



A 4 cm x 4 cm wood square, a 10 cm x 10 cm wood square and a thin cardboard 4 cm x 4 cm square are laid out before the subject.

Introduction

"Imagine that this is frosting which has been spread out just 1/8" thick over this small cake."

Question

"Can you predict what would be the thickness of this same amount of frosting if it were to be spread out over the larger cake? How did you get your answer?"

Scoring Criteria

Stage	Criteria	Score
I	Subject has no answer or reason "I don't know"	0

Scoring Criteria (continued)

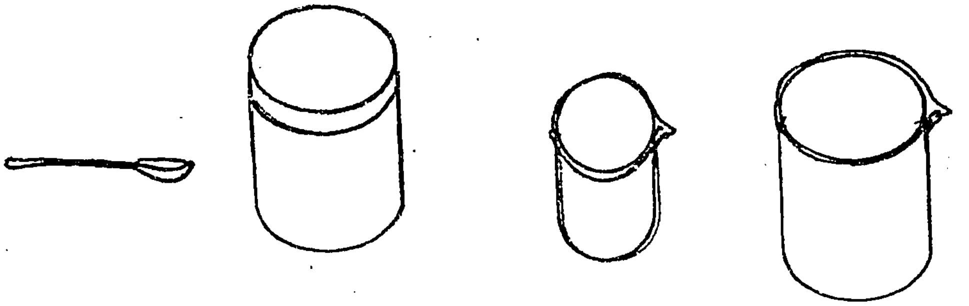
Stage	Criteria	Score
IIA	Subject qualitatively compensates "It would be less"	1
IIB	Subject quantitatively adds or subtracts "It's 6 more so about 1/14 to 1/16"	2
IIIA	Subject calculates using a multiplication factor ratio $16/100 \times 1/8 = 1/50$	3
IIIB	Subject uses an appropriate proportion $\frac{16}{100} = \frac{X}{1/8}$	4

10. Paint

Thinking tested

Proportional reasoning
Direct proportion
Phys_cal

Material



A small (1 ml) measuring spoon, some "Tang" orange drink and a 60 ml and a 250 ml beaker of water are set out on the table.

Introduction

"If I add two measures of Tang to the water in my small 60 ml beaker, I get a certain color and sweetness." Show this.

Question

"How much water should I add to make the same color and sweetness with 5 measures of Tang? How did you get your answer?"

Scoring Criteria

Stage	Criteria	Score
I	Subject guesses or has no prediction	0
IIA	Subject estimates with some qualitative compensation	1

Scoring Criteria (continued)

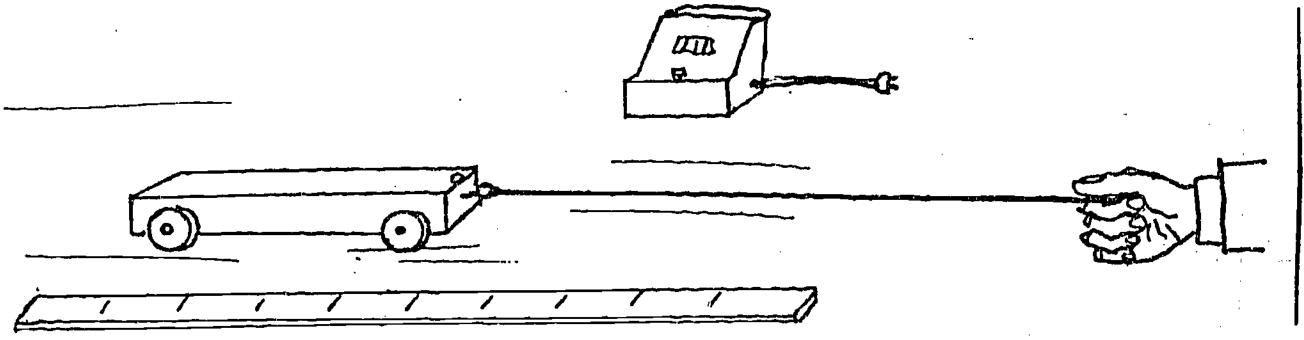
Stage	Criteria	Score
IIB	Subject predicts with some addition or subtraction " 6 because 250/60 = 4 So 2 + 4 = 6"	2
IIIA	Subject utilizes a multiplication factor or ratio "About 8, 60/250 = 4, 4 x 2 = 8"	3
IIIB	Subject utilizes the appropriate proportion and relates some general rule "For the same color it would be proportional" 2/60 = X/250	4

11. Speed

Thinking tested

Proportional reasoning
Direct proportion
Physical

Material



A cart is pulled by the experimenter with a 50 cm length of string. A meter stick graduated into centimeters is used for measuring. An electric timer gives digital readings of time in tenths of a second.

Introduction

"I am going to pull this cart along. I want you to time a 30 cm run. The clock starts when you push it and stops when you push it. Try it. Now do it with the run. Start! Stop! It took ___ seconds to go 30 cm."

Question

"If I were to continue pulling it along in the same way, how long would it take to go 50 cm? Explain how you got your answer."

Scoring Criteria

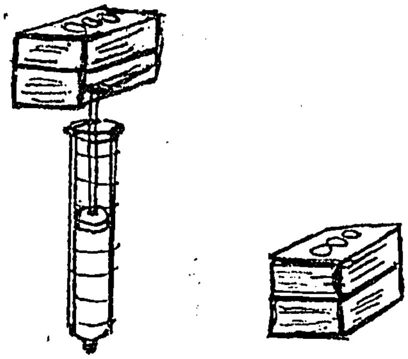
Stage	Criteria	Score
I	Subject guesses or has no prediction	0
IIA	Subject qualitatively compensates "It should be more, about ___ seconds"	1
IIB	Subject quantifies his approach through addition "It's 20 more cm so it should be 20 seconds more"	2
IIIA	Subject consciously applies a ratio or multiplication factor	3
IIIB	Subject recognizes and states a general law. Subject uses proportion. "The car is going the same speed so...."	4

12. Boyle

Thinking tested

Proportional reasoning
Inverse proportion
Physical

Material



Bricks	Syringe
0	30 cc
2	20
4	10

A brick is balanced upon a sealed off graduated syringe to compress the trapped air. Some extra identical bricks are nearby.

Introduction

"This syringe, with its trapped air, feels kind of squashy." Subject tries it. "With no bricks the syringe reads 30 cc; I'm going to add two bricks. Watch what happens." Add reading to chart. "Next see what happens with four bricks." Add reading to chart.

Question

"Can you predict what reading the syringe should have with five bricks on it? How did you get your answer?"

Scoring Criteria

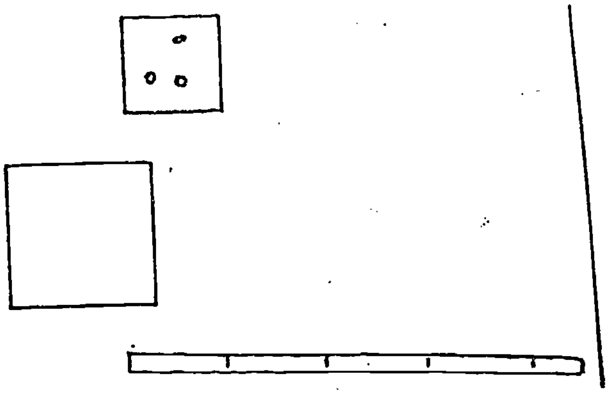
Stage	Criteria	Score
I	Subject has no reason, maybe no answer	0
IIA	Subject estimates qualitatively "It will be less"	1
IIB	Subject uses some subtraction for a somewhat quantitative approach "It should be 3 less"	2
IIIA	Subject calculates quantitatively with some multiplication factor 2 x 20 = 40 4 x 10 = 40 5 x 8 = 40	3
IIIB	Subject calculates from differences using a sort of rule "5 bricks means the volume = 8 Because 4/5 = x/10 so x = 8"	4

13. Population

Thinking tested

Proportional reasoning
Direct as square
Physical

Material



A 50 unit ruler, a square 10 units on edge and a square 18 units on edge were set before the subject. 3 markers were placed on the 2 measure square.

Introduction

"If just 3 cows can live on this much grass, 10 x 10 units, what is the most number of cows that can live on a plot of grass that is 18 x 18 units?"

Question

"How did you get your answer?"

Scoring Criteria

Stage	Criteria	Score
I	Subject guesses or makes no prediction	0

Scoring Criteria (continued)

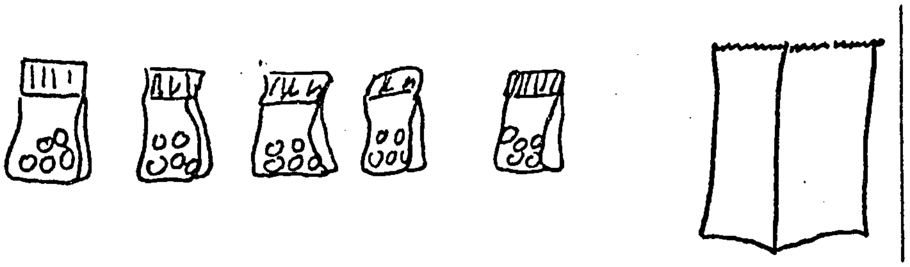
Stage	Criteria	Score
IIA	Subject qualitatively compensates "About 5"	1
IIB	Subject uses addition to quantify his answer "11 cows, 18 is 8 more than 10 8 + 3 = 11"	2
IIIA	Subject uses a ratio or a multiplication factor possibly inappropriately $\frac{10}{18} = \frac{3}{\text{about } 5}$	3
IIIB	Subject projects a general rule into the data and uses appropriate proportions $\frac{52}{92} = \frac{25}{81} = \frac{3}{\text{about } 10}$ "About twice as large a square has 4 times as much grass"	4

14. Probability

Thinking tested

Proportional reasoning
Direct proportion
Physical

Material



5 clear packets each containing 2 red and 3 yellow gum drops and a paper bag are placed in front of the observer.

Introduction

"Notice that this bag has 2 red and 3 yellow gum drops. Suppose you were to close your eyes and reach into the sack. You could then get either a red or a yellow gum drop. Suppose now I empty all of these into the paper bag."

Question

"What chance is there that you would get a red gum drop? How did you get your answer?"

Scoring Criteria

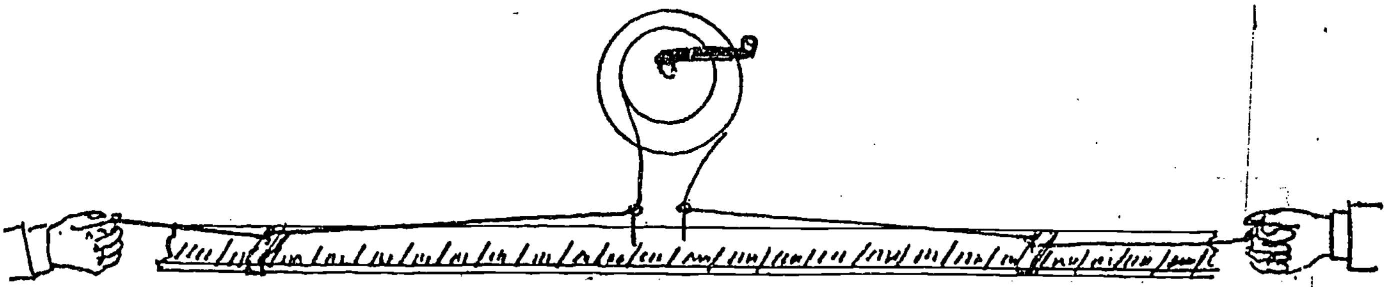
Stage	Criteria	Score
I	Subject has no reason or calculation and possibly no answer "I don't know"	0
IIA	Subject estimates with some qualitative compensation "It's probably yellow because there are more yellow ones"	1
IIB	Subject predicts with some addition or subtraction to compensate "Now there are 5 extra chances for yellow, because there are 5 more yellows"	2
IIIA	Subject quantitatively compensates with a multiplicative or ratio factor "It's 2 to 3 for reds to yellows and now it's 10 to 15 or the same"	3
IIIB	Subject quantitatively compensates relating a general rule "2 to 5 for red and 3 to 5 for yellow. There are 2 reds to 5 candies and 3 yellows to 5 candies. Putting in more keeps the same ratios"	4

15. Pulley (Karplus, Karplus and Wollman, 1974)

Thinking tested

Proportional reasoniлg
Direct proportion
Physical

Material



A system of two pulleys, one 3" in diameter the other 2" in diameter, mounted on the same shaft are arranged so that as one turns the crank one pulley pulls string in while the other lets it out. These strings pull markers along a meter stick.

Introduction

"Hold onto this end (left) while I hold the other (right). Now notice as I wind the crank, your end (subject) has moved 20 cm while mine has moved 15 cm."

Question

"How far will my string move when yours moves 5 cm? How did you get your answer?"

Scoring Criteria

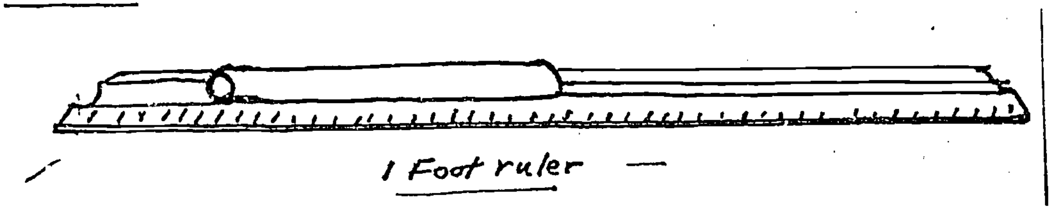
Stage	Criteria	Score
I	Subject guesses. The answer has no reason or calculation. "I can't explain it. I guessed."	0
IIA	Subject estimates with same qualitative compensation outside of any comprehension of the task or any rule. "When I had 10 you had 15, so when I get 6 you should get more, about 8."	1
IIB	Subject quantitatively compensates with addition or subtraction without regard to any physical relationship. "Zero 20 - 5 = 15 so 5 - 5 = 0"	2
IIIA	Subject quantitatively compensates with some multiplication factor. Does not seek out physical rule. "20 matches with 15 so 5 should match with about 4."	3
IIIB	Subject quantitatively compensates seeking out a proportional relationship and a physical rule. "15 is 3/4 of 20 -- so 3.75 is 3/4 of 5. The big pulley goes 4 for the little one's 3."	4

16. Ruler (Karplus, Karplus and Wollman, 1974)

Thinking tested

Proportional reasoning
Direct proportion
Physical

Material



On a centimeter and inch graduated rule, a 4" long pencil is placed.

Introduction

"Notice that this length of pencil extends about 4 units on the inch scale and about 10 units on the centimeter scale."

Question

Suppose I were to put down a pencil that covered 5 inches. How many centimeters might it cover? How did you get your answer?"

Scoring Criteria

Stage	Criteria	Score
I	Subject guesses. Makes no calculation. "I guessed."	0
IIA	Subject estimates with qualitative compensation	1
IIB	Subject quantitatively compensates through addition or subtraction. "10 is 6 more than 4 so for 5 I would get 9."	2

Scoring Criteria (continued)

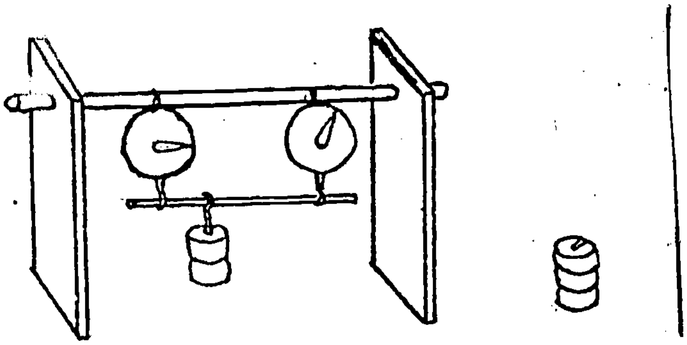
Stage	Criteria	Score
I	Subject guesses. Makes no calculation. "I guessed."	0
	Subject estimates with qualitative compensation	1
IIB	Subject quantitatively compensates through addition or subtraction. "10 is 6 more than 4 so for 5 I would get 9."	2
IIIA	Subject quantitatively compensates without reference to any general relationship. "With 4 it's 10 so with 5 it's about 13."	3
IIIB	Subject quantitatively compensates iterating the relationship of inches and centimeters.	4

17. Weight

Thinking tested

Proportional reasoning
Physical

Materiel



Weights are placed off center on a light rod. Separate spring scales measure the weight on each side of the rod. An additional three weights are nearby.

Introduction

"You can see that these scales show how much weight each set of wheels carry." Examiner lifts slightly one weight.

Question

"Now, can you predict how much each scale will register if I add three more weights for a total of 5 weights? How did you get your answer?"

Scoring Criteria

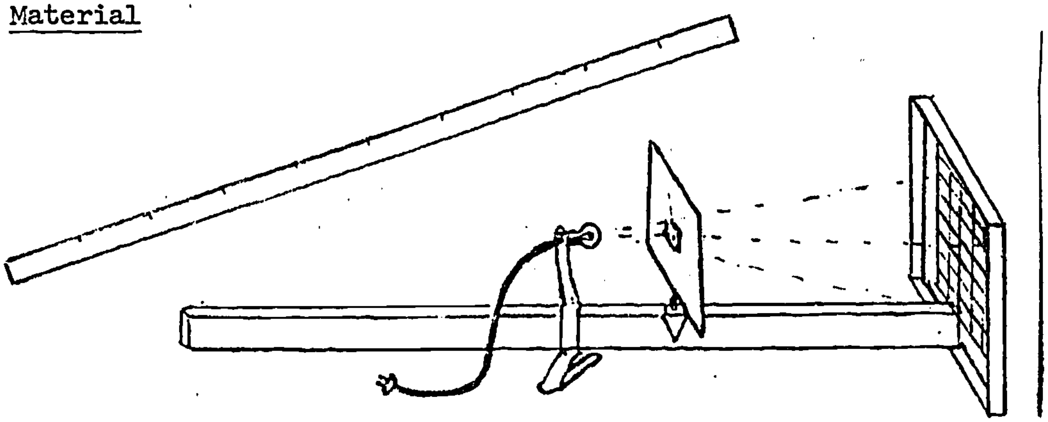
Stage	Criteria	Score
I	Subject has no reason or explanation and possibly no answer. "I guessed."	0
IIA	Subject estimates qualitatively some compensation "About 6 and 2."	1
IIB	Subject compensates with addition "5 and 3 because it's one more" "6 and 4 because it's two more"	2
IIIA	Subject quantitatively compensates with some multiplication "It's 2 to 1 so with 5 it must be about 10/3 to 5/3"	3
IIIB	Subject states a general rule "With 5 it must add up to 10 and be in the ratio 2/1 so it's about 6 and 3"	4

18. Light and Shadow

Thinking tested

Schema of proportions
Direct proportion
Physical

Material



A chart, lamp and "mask" were attached to a meter stick. The lamp and screen can be moved along the meter stick. An observation screen 30 cm x 30 cm has on its surface a grid of 1 cm squares. Light from a bulb goes out through a "mask" with a 1 cm square hole and projected a square of light on the screen. The "light" and "hole are positioned at the same height and at the center of the observing screen. Markings on the meter stick are masked out. Letters note 10 cm marks on the meter stick. A meter stick with centimeter markings is nearby for use in measuring.

Introduction

"Here is a light, a masking screen, and a chart. The way it is now arranged it makes a lighted square with four units on the screen."

Question

Initially seek out correspondence between change of "mask" position and the projection with questions such as: "What would you predict will happen if I were to move the mask toward the light? toward the screen?" Do it. "With the "mask" at this distance from the light, I get a projection just with four units on the screen. What then should I do to get 16 units on the screen? How did you get your answer?"

Scoring Criteria

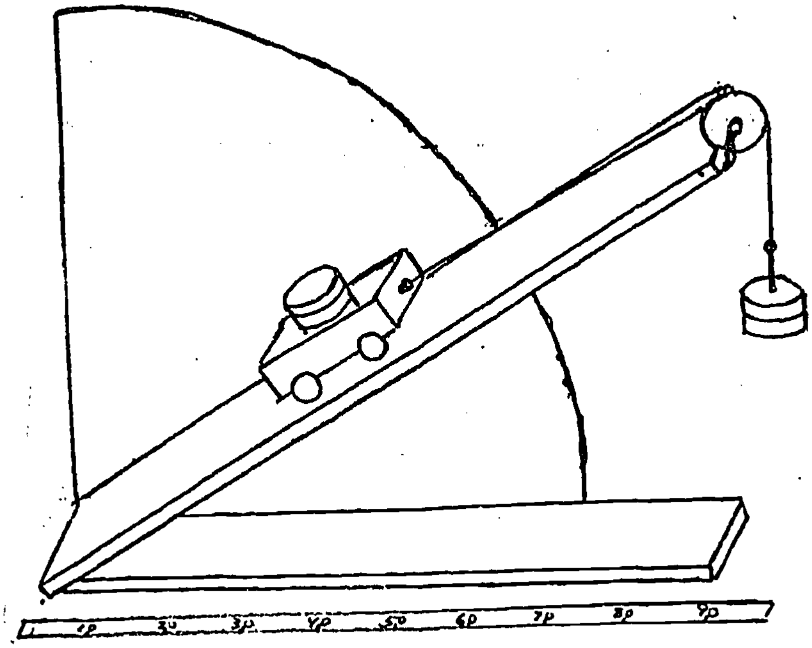
Stage	Criteria	Score
I	The subject views the projection in the way it works. He does not perceive how the projection is formed on the screen.	0
IIA	The subject recognizes how the projection can be changed by moving the "mask."	1
IIB	The subject suggests how changing the "mask" location will change the projection size. The subject may use addition or subtraction to predict same sizes.	2
IIIA	The subject quantitatively calculates same predicted relationship between size and location. The subject measures distances from the light source.	3
IIIB	The subject links "mask" location and projection size with an overall model of what is causing the change. The subject states the relationship in terms of a proportion.	4

19. Incline (Hensley, 1974)

Thinking tested

Overall schema of proportions
Direct proportion
Physical

Material



Welch Scientific Company Inclined Plane, Hall's Carriage, 100 gram slotted weights, weight hanger cord, meter stick.

An inclined plane demonstration device was used. Statements of mechanical advantage, angles and distances were masked out where they were printed on the device.

Introduction

"I have here a cart with some weights on it. It can roll on the incline (demonstrate). It now stays where I put it."

Question

Seek initially all factors the subject can suggest. "What should I do to make the cart move? What else could I do to make it move? Up? Down? What other things could be changed? What general rule can you suggest that will explain what will make the cart move?"

"The cart is now balanced. If I now take off 100 grams, what else should I change to again make it balance? How much should I change it? How did you get your answer?"

Scoring Criteria

Stage	Criteria	Score
I	Subject explains the situation in terms of the totality of the actions which he can perform (he pushes the car up the incline).	0
IIA	The subject perceives the role of the weight on the hook--more weight on the hook, the car moves up the incline. The subject does not perceive the role of the incline.	1
IIB	The subject is able to compensate the effect of weight with a change in the incline.	2
IIIA	Subject coordinates the role of the weight and inclination. The subject can state the overall rule but does not state the proportion with numbers or make a numerical prediction.	3
IIIB	In addition to the attributes at IIIA, the subject gives correct predictions, states the proportion with numbers, and may use the words like its proportions in his explanation.	4

APPENDIX C

Calculations of Final Test Characteristics

Calculation of Criterion-Referenced Reliability for 427 Grade [illegible] Pupils Tested with the Final Version June, 1974

$$r_c = \frac{r_x \quad \sigma_x^2 + (\bar{X} - C)^2}{\sigma_x^2 + (\bar{X} - C)^2}$$

where

r_c = criterion-referenced reliability

r_x = classical reliability estimate (Hoyt, 1941) .779

σ_x^2 = variance of test scores 20.81

$\bar{X}$ = mean of test scores 12.13

C = criterion level 15

$$r_c = \frac{(.779) \quad (20.81) + (15 - 12.13)^2}{20.81 + (15 - 12.13)^2}$$

$$r_c = .842$$

Calculations of Score Reliability for 427 Grade 8 Pupils Tested with the Final Version June, 1975

Score	Frequency
3	4
4	7
5	19
6	22
7	24
8	30
9	22
10	34
11	34
12	41
13	38
14	19
15	31
16	17
17	21
18	24
19	16
20	10
21	33
22	7
23	3
24	1

SV	df	SS	Variance
Total	10247	2561.7048	.2499955
Among items	23	212.2389	9.2277782
Among individuals	426	385.8585	.9057711
Remainder	9798	1963.6074	.200409

$$\text{reliability} = \frac{(\text{Variance among individuals}) - (\text{Remainder})}{\text{Variance among individuals}}$$

(Hoyt, 1941)

$$r_{tt} = \frac{.9057711 - .200409}{.9057711} = .779$$

Mean = 12.13

SD = 4.56

Range = 3-24 (21)

Subjects = 427

Tetrachoric Test-Retest Reliability

Grade 5 Pupils

	Master	Non-master	
Master	5.3% N= 5	8.5% N= 8	13.8% N= 13
Non-master	12.8% N= 12	73.4% N= 69	86.2% N= 81
	18.1% N= 17	81.9% N= 77	100.0% N= 94

$r_t = .40$

Grade 8 Pupils

	Master	Non-master	
Master	41.5% N=174	11.0% N= 46	52.5% N=220
Non-master	14.6% N= 61	32.9% N=138	47.5% N=199
	56.1% N=235	43.9% N=184	100.0% N=419

$r_t = .70$

Chemistry Pupils

	Master	Non-master	
Master	91.3% N=136	4.7% N= 7	96.0% N=143
Non-master	3.4% N= 5	.7% N= 1	4.0% N= 6
	94.6% N=141	5.4% N= 8	100.0% N=149

$r_t = .32$

Composite Sample

	Master	Non-master	
Master	53.0% N=179	7.4% N= 25	60.4% N=204
Non-master	10.4% N= 35	29.3% N= 99	39.6% N=134
	63.3% N=214	36.7% N=124	100.0% N=338

$r_t = .84$

Cross Tabulation of Test-Retest Results by Reasoning Level

Grade 5 Pupils

		Test 2					
		0	1	2	3	4	Tot
Test 1	0	27	10	2	2	0	41
	1	12	13	1	6	0	32
	2	0	1	3	2	2	8
	3	3	5	0	5	0	13
	4	0	0	0	0	0	0
	Tot	42	29	6	15	2	94

Raw chi-square 55.1
with 12 degrees of freedom

Significance < .0001

Grade 8 Pupils

		Test 2					
		0	1	2	3	4	Tot
Test 1	0	25	12	9	5	1	52
	1	22	27	16	13	2	80
	2	4	7	17	26	13	67
	3	3	20	16	79	35	153
	4	0	4	3	19	41	67
	Tot	54	70	61	142	92	419

Raw chi-square 227
with 16 degrees of freedom

Significance < .0001

Chemistry Pupils

		Test 2					
		0	1	2	3	4	Tot
Test 1	0	0	0	0	0	0	0
	1	0	0	0	0	2	2
	2	0	0	1	1	1	3
	3	1	2	2	58	16	79
	4	1	0	1	17	46	65
	Tot	2	2	4	76	65	149

Raw chi-square 51.99
with 12 degrees of freedom

Significance < .0001

Composite

		Test 2					
		0	1	2	3	4	Tot
Test 1	0	33	13	2	5	0	53
	1	16	20	6	10	2	54
	2	0	1	8	12	5	26
	3	6	11	5	82	24	128
	4	1	1	1	21	53	77
	Tot	56	46	22	130	84	338

Raw chi-square 295.0
with 16 degrees of freedom

Significance < .0001

Pearson Correlation of Test-Retest Reliability

Pupils	Test Period	Cases	Mean	SD	Pearson Correl. Coeff.	Level of Significance
5th Grade	1	94	8.6	3.6	.68	.001
	2	94	9.4	4.1		
8th Grade	1	419	14.1	4.3	.70	.001
	2	419	14.5	4.5		
Chemistry (11th Grade)	1	149	19.4	2.6	.47	.001
	2	149	19.0	3.2		
Composite	1	338	14.8	5.7	.83	.001
	2	338	15.1	5.5		

APPENDIX D

Final Paper-Pencil Test

SCIENCE PROBLEM SOLVING TEST

Use of the Test

This test is intended for use with grade 8 pupils, that is persons who are approximately 13 years old. It will be completed within 30 minutes by 90 per cent of such pupils. The test may be used as low as grade 5, that is with about 9-year-olds or as high as grade 12, that is with about 18-year-olds. Use at these extremes will reduce the reliability of measurement. Pupils at the high ages will have scores clustered in the high ranges. Pupils at the low ages will have scores clustered in the low ranges.

Directions for Administering

Pupils should have a good writing surface, a pen or pencil, and answer sheets with A B C D E answers for 24 questions.

Test Scoring

The correct order and answers to test questions are listed. Mastery at each level is four or more of the six correct.

Level I	Level II	Level III	Level IV
1 - D	2 - B	3 - C	4 - C
5 - A	8 - D	7 - B	6 - B
9 - A	12 - A	11 - A	10 - B
15 - D	14 - B	13 - A	16 - D
20 - C	18 - D	17 - C	19 - A
23 - B	21 - A	24 - C	22 - B

Grading master (1) and non-master (0) responses follows this form:

Preoperational	Level I	Level II	Level III	Level IV
0000	1000	1100	1110	1111
	0010	0101	0110	1101
	0001	1001	0111	1011
	0011	0100	1010	

SCIENCE PROBLEM SOLVING TEST TEST FORM # 1 (Version 9) Date 4/23

Directions: Select the answer that most closely is the way you would solve each problem. Mark the letter of your answer on the answer sheet in this manner A X C D E

1 (14C$_1$)
Mary buys 3 tickets to a raffle where 90 tickets are sold --- Jane buys 1 ticket to a raffle where 30 tickets are sold --- Sue buys 3 tickets to a raffle where 300 tickets are sold.

Which girls have about the same chance of winning?

A. Jane and Mary because their's are the least tickets

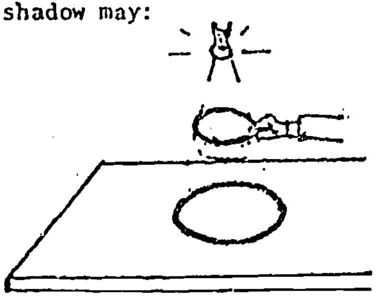
B. Sue and Mary because each have 3 tickets

C. All girls have the same chance

D. Jane and Mary because 3 chances in 90 is the same as 1 in 30

E. I have no answer

2 (1C$_2$)
A ring is held between a table and a light bulb. The light casts a shadow of the ring onto the table. If the ring is moved closer to the table, the shadow may:



A. Become larger because the shadow spreads out

B. Become smaller because the light rays don't spread as much

C. Stay the same because it's the same ring

D. Become larger because the bulb is father away

E. I have no answer

3 (2F$_1$)
A lunchroom is 60 ceiling tile or 25 chairs wide. If a classroom is 12 chairs wide, how wide is this classroom measured in ceiling tiles?

A. Seems to be 50.

B. About 40 because it has to be less.

C. About 29 because $\frac{60}{25}$ is about $\frac{29}{12}$.

D. About 47 because 60 is 35 more than 25 and 47 is 25 more than 12.

E. I have no answer.

4 (10F$_2$)

Here is a recipe for 4 cups of cocoa: Heat to near boiling 4 c. milk
Add with stirring 6 T. sugar
5 T. Cocoa

How many tablespoons of sugar would be needed to make 12 cups of this cocoa?

A. 18 tablespoons because $\frac{6}{4} \times 12 = 18$

B. More than 6 tablespoons because there is more cocoa

C. 18 tablespoons because $\frac{6}{4}$ equals $\frac{18}{12}$

D. 14 tablespoons because 4 c. + 8 c. = 12 c.
so 6 T. + 8 T. = 14 T.

E. I have no answer

5 (11C$_1$)

A car moving at a constant speed of 30 mph will, if pictured at one second intervals, looks like

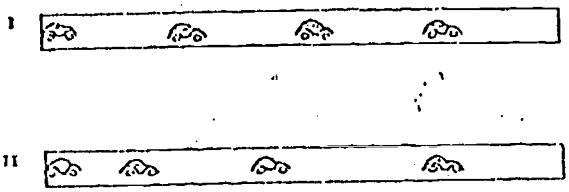
A. I because it moves equal distances each second

B. None of these because it is moving

C. II because it changes

D. II because it is increasing its distance

E. I have no answer



6 (1F$_2$)

A ring 3 inches across is 2 feet from the light and 4 feet from the table. The 3" ring has a 9" shadow. Where should a 4" ring be placed to make the same size shadow?

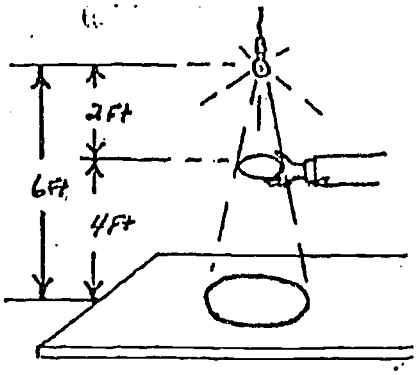
A. The shadow will be larger than 9" wherever the ring is placed.

B. About 3 ft. from the lamp because $\frac{x}{4} = \frac{2}{3}$
and 3x = 8

C. About 3 ft. from the lamp because $\frac{2}{3} \times 4 = 2.7$

D. About 3 ft. from the lamp because the ring is 1" larger 3 + 1 = 4 and 2ft. + 1 ft. = 3 ft.

E. I have no answer



7 (18F1)
A movie projector lens spreads its light out over a 3' x 3' screen 9 feet away. To make the image spread over a 5' x 5' screen, how far back must the screen be moved?

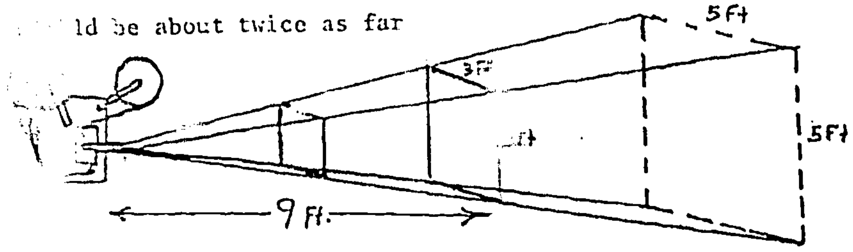
A. About 11 feet. The 5 foot image is 2 more than the 3 foot one and 11 feet is 2 more than 9 feet

B. About 15 feet because 3/9 = 5/15

C. About 12 feet because 9 + 3 = 12

D. About 18 feet because it would be about twice as far

E. I have no answer



8 (3C2)
This person sliding down a hill looks at her watch Each second she puts a stick in the snow. What most likely would be the pattern of these sticks?

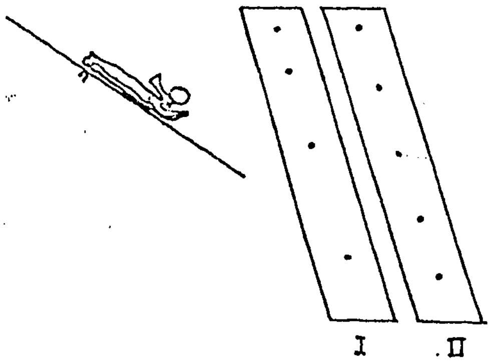
A. I because she moves each second

B. II because she speeds up"

C. I or II because she is moving

D. I because her speed is changing

E. I have no answer



9 (2C1)
A student's desk measures about three textbook lengths or 5 pencil lengths wide. If a teacher's desk is 4 textbook lengths wide, how wide is a teacher's desk measured in pencil lengths?

A. More than 5 pencils because it is bigger than a student desk

B. Less than 5 pencils because it seems that way

C. About 4 pencils because it was 4 textbooks

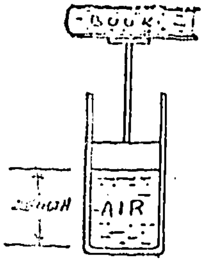
D. 5 pencils because that is what the student desk measured

E. I have no answer

	Text books	Pencils
Student Desk	3	5
Teacher Desk	4	?

10 (12F_2)

Books on top of this air spring compress the spring. For 2 books the spring is 8 cm. long. For 9 books it is 1.8 cm. What should be the spring length for 5 books?



A. About 3 cm. to 4 cm. because it has to be about half between 1.8 cm. and 8 cm.

B. About 3 cm. because $\frac{2 \text{ books}}{9 \text{ books}} = \frac{1.8 \text{ cm.}}{8.0 \text{ cm.}}$

then $\frac{2 \text{ books}}{5 \text{ books}} = \frac{3.2 \text{ cm.}}{8.0 \text{ cm.}}$

C. About 3 cm. because $\frac{2}{5} \times 8 = 3.2$

D. About 5 cm. because 5 books - 2 books = 3 books
and 8 cm. - 3 cm. = 5 cm.

E. I have no answer

11 (10F_1)

Jim uses 4 heaping teaspoons of Tang powder with an 8 oz. glass of water. How much Tang is needed for the same mixture with 12 oz. of water?

A. About 6 teaspoons because $\frac{12}{8} \times 4$ tsp. = 6 tsp.

B. About 8 teaspoons because 8 oz. + 4 oz. = 12 oz.
and 4 tsp. + 4 tsp. = 8 tsp.

C. More than 4 teaspoons because there is more water

D. 4 teaspoons because it is the same mixture

E. I have no answer

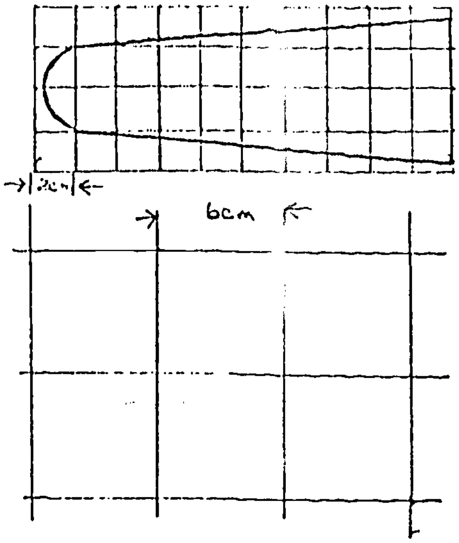
12 (11C_2)

Four cars have different speeds: Car A is the fastest, Car B the next fastest, Car C the next fastest, and, Car D the next fastest. The fastest car takes the least time to go 200 miles, the next fastest car the next least time and so on. Which car is the third fastest and takes the third least time to go 200 miles?

A.	Car C because:	1st fastest CAR A 1st least time	2nd fastest CAR B 2nd least time	3rd fastest CAR C 3rd least time
B.	Car B because	1-CAR D	2-CAR C	3-CAR B
C.	No car because they don't match up			
D.	Car C because:	1st most fast CAR A 1st most time	2nd most fast CAR B 2nd most time	3rd most fast CAR C 3rd most time
E.	I have no answer			

13 ($8F_1$)

A model airplane wing made from the 2 cm. pattern shown measures 19 cm. long. What would be the length of such a wing made from a pattern with squares that are 6 cm.?



A. 57 cm. because 6/2 x 19 = 57

B. 18 cm. because it looks that way

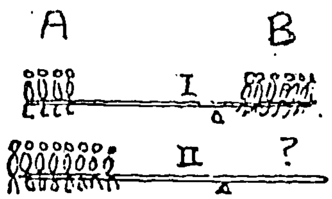
C. 22 cm. because 19 + 3 = 22

D. 19 cm. but the squares would be larger

E. I have no answer

14 ($5C_2$)

Trial I 4 people on side "A" balance 6 of the same size people on side "B"
Trial II 8 people on side "A" should balance how many on side "B"?



A. About 10 becauso 4 more on "A" should balance 4 more on "B"

B. About 12 because it goes up 6 and 6 + 6 = 12

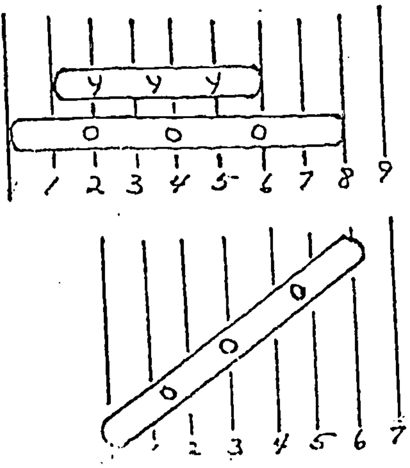
C. About 10 because it takes 4 more and 6 + 4 = 10

D. About 11 because it should be more

E. I have no answer

15 ($4C_1$)

The "O" rod here crosses 8 lines. The "Y" rod crosses 5 lines. The "O" rod, when turned, crosses 6 lines. How many lines would the "Y" rod cross if it were at this angle?



A. About 8 because it should get longer

B. About 5 because the "Y" rod is that long

C. About 6 because the "O" rod was 6

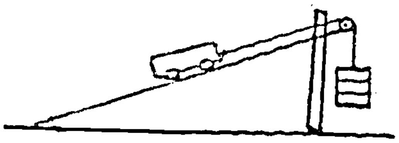
D. About 4 because the "Y" rod is shorter

E. I have no answer

16 (19F2)

On the [illegible] illustrated the cart and its weight is balanced by weights [illegible] the string. What amount of weight is needed to balance [illegible] g of cart weight at 20° ?

	Weight	
Angle	Cart	String
10°	200g	35
10°	300g	52
20°	300g	100
20°	400g	?



A. 133 because $\frac{100}{300}$ x 400 = 133

B. 130 because it is more

C. 177 because it goes up 17 for every [illegible]

D. 133 because $\frac{100}{300} = \frac{133}{400}$

E. I have no answer

17 (11F1)

A car moving at a constant 30 mph travels 88 ft. in 2 seconds. How far will it have traveled by the end of 5 seconds?

A. About 264 feet because 3 x 88 = 264

B. About 100 feet because it is only 3 seconds more

C. 220 feet because $\frac{88}{2}$ x 5 = 220

D. 91 feet because 3 sec. + 2 sec. = 5 sec.
and 88 ft. + 3 ft. = 91 ft.

E. I have no answer

18 (10C2)

Here are some recipes for Kool Aide

	2 quarts	4 quarts	5 quarts
Kool Aide Powder	1 pkg	2 pkg	?
Sugar	½ c	1 c	
Water	2 qt	4 qts	

How much powder is needed for 5 quarts of Kool Aide

A. 2 pkg because it is the same mixture

B. 3 pkg because 4 qts + 1 qt = 5 qts
and 2 pkg + 1 pkg = 3pkg

C. About 3 because it would have to be more

D. 2½ pkg because 4 qts + 1 qt = 5 qts
and 2 pkg + ½ pkg = 2½ pkg

E. I have no answer

19 ([illegible])

A freeway driver keeps track of the distance he travels. He finds that in 4 minutes he travels 3 miles/ in 10 minutes 7½ miles. If he continues at this speed, how long will it take him to travel 10 miles?

Distance	Time
3 miles	4 min
7½ miles	10 min
10 miles	? min

A. About 13 minutes because

$$\frac{4\text{ min.}}{3\text{ miles}} = \frac{10\text{ min.}}{7.5\text{ miles}} = \frac{13\ 1/3\text{ min.}}{10\text{ miles}}$$

B. About 13 minutes because 10 - 7½ = 2½ miles
and 10 + 2½ = 12½ min.

C. About 13 minutes because $\frac{4}{3}$ x 10 = 13 1/3

D. About 14 because 7½ + 3 = 10½
and 10 + 4 = 14

E. I have no answer

20 (9C$_1$)

Imagine that frosting had been spread out ¼ inch thick on top of a small 6" x 6" cake. Predict what the thickness would be if the same amount of frosting were spread out over a 12" x 12" cake?

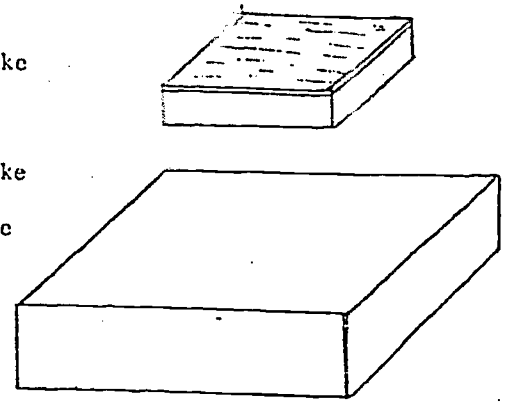
A. More than ¼ inch because it covers less cake

B. Less than ¼ inch because it looks that way

C. Less than ¼ inch because it covers more cake

D. More than ¼ inch because there is more cake

E. I have no answer



21 ([illegible]$_2$)

These nature hunt groups are chosen for a nature hike.

Mrs. Andrews	- 5 students
Mr. Denton & Mrs. Felk	- 8 students
Mr. Holt	- 6 students

The teacher with the most students to help is:

A. Mr. Holt because $\frac{6}{1}$ is larger than $\frac{5}{1}$ is larger than $\frac{8}{2}$

B. Mr. Denton & Mrs. Felk because $\frac{2}{8}$ is larger than $\frac{1}{5}$ is larger than $\frac{1}{6}$

C. Mr. Denton & Mrs. Felk because they have the most students

D. Mrs. Andrews because she has fewer students

E. I have no answer

22 (2[illegible]

Sketch #1 of a house is 5 pencil widths or 2 pennies high. Sketch #2 of this house is not shown. Sketch #2 looks the same but is 8 pencil widths high. How high must sketch #2 be in pennies?

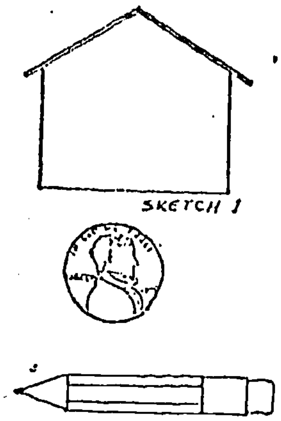
A. About 3 because 8 - 5 = 3

B. About 3 because $\frac{2}{5} = \frac{3.2}{8}$

C. About 3 because $\frac{2}{5} \times 8 = 3.2$

D. About 3 because it has to be more

E. I have no answer



23 (10[illegible])

A ring is held between a table and a light bulb. The light bulb casts a shadow of the ring. If a smaller ring was held in the same place the shadow of the smaller ring would

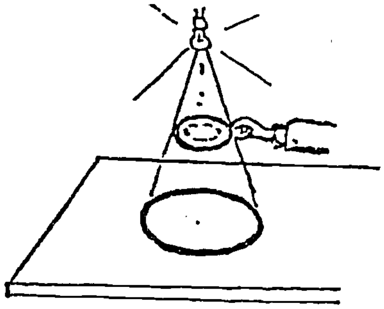
A. Be smaller because the light would change

B. Be smaller because the ring is smaller

C. Be the same size because the ring is in the same place

D. Be larger because it is different.

E. I have no answer



24 (17F1)

Jane is weighing out apples on this supermarket scale. What will 14 apples weigh if 6 apples weigh 2 lbs?

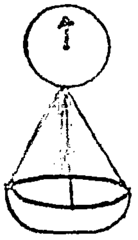
A. 10 lbs because 6 + 8 = 14
so
2 + 8 =10

B. 3 or 4 lbs because it is more

C. 4 2/3 lbs because $\frac{2}{6}$ x 14 = 4 2/3

D. 5 because 2 + 2 + 1 = 5

E. I have no answer



APPENDIX E

Pupil Results and Test Improvements in Versions II-VI

LEVEL I

VERSION III (Test Forms 7 & 8) was changed from VERSION II

14C,

% correct

Mary buys 3 tickets to a raffle where 90 tickets are sold --- Jane buys 1 ticket to a raffle where 30 tickets are sold --- Sue buys 3 tickets to a raffle where 300 tickets are sold.

Which girls have about the same chance of winning?

VERSION II			VERSION III					
All	0000	1000	All	0000	1000	1100		
17	36	10	7	15	0	39	A. Jane and Mary because there are the least tickets	3
10	18	10	10	24	6	61	B. Sue and Mary because each have 3 tickets	2
10	9	10	15	24	0	0	C. All girls have the same chance	1
62	36	70	63	24	94	0	[✓ D.] Jane and Mary because $\frac{3}{90} = \frac{1}{30}$	4
0	0	0	4	8	0	0	E. I have no answer	0

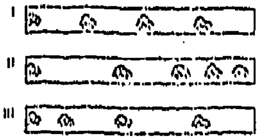
CHANGES (The responses and the question appear appropriate.)

None

11C,

A car moving at a constant speed of 30 mph will, if pictured at one second intervals, look lik

VERSION II			VERSION III					
All	0000	1000	All	0000	1000	1100		
62	9	90	70	40	98	100	[✓ A.] I because it moves equal distances each second	4
3	0	10	6	14	0	0	B. None of these because it is moving	1
21	55	0	14	27	2	0	C. II because it changes	3
10	27	0	6	10	0	0	D. III because it is increasing its distance	2
3	9	0	2	4	0	0	E. I have no answer	0



CHANGES

None

9C,

Imagine that frosting had been spread out ½ inch thick on top of a small 6" x 6" cake. Predict what the thickness would be if the same amount of frosting were spread out over a 12" x 12" cake?

VERSION II			VERSION III					
All	0000	1000	All	0000	1000	1100		
10	18	10	2	8	0	0	A. More than ½ inch because it covers less cake	2
0	0	0	2	7	0	0	B. Less than ½ inch because it looks that way	3
69	55	70	83	58	98	100	[✓ C.] Less than ½ inch because it covers more cake	4
14	28	10	9	20	2	0	D. More than ½ inch because there is more cake	1
7	9	10	3	6	0	0	E. I have no answer	0

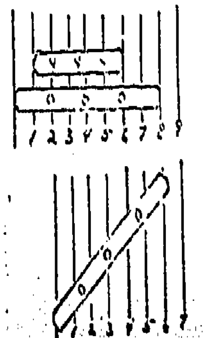
CHANGES

None

4C,

The "O" rod here crosses 8 lines. The "Y" rod crosses 5 lines. The "O" rod, when turned, crosses 6 lines. How many lines would the "Y" rod cross if it were at this angle?

VERSION II			VERSION III					
All	0000	1000	All	0000	1000	1100		
0	0	0	5	16	0	0	A. About 8 because $\frac{8}{5} \times 5 = 8$	3
10	27	0	10	21	2	26	B. About 5 because the "Y" rod is that long	2
7	18	0	4	10	21	10	[✓ C.] About 6 because the "O" rod was 6	1
72	27	100	72	43	71	65	[✓ D.] About 4 because the "Y" rod is shorter	4
10	27	0	9	3	6	0	E. I have no answer	0



CHANGES

None (Wrong key)

version III test forms 7&8 as changed from version II

LEVEL I

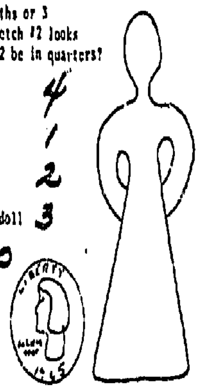
VERSION 2

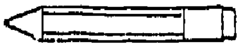
VERSION 3

2C1

Here is sketch #1 of a paper doll. Sketch #1 is 10 pencil widths or 3 quarters high. Sketch #2 of this paper doll is not shown. Sketch #2 looks the same but is 14 pencil widths high. How high must sketch #2 be in quarters?

All	0000	1000		
48	9	60	[✓ A.] More than 3 quarters because paper doll #2 is larger	4
17	45	0	B. Fewer quarters because it seems that way	1
0	0	0	C. 14 quarters because it is 14 pencils	2
21	36	10	D. The same number of quarters since its the same paper doll	3
14	9	30	E. I have no answer	0





A student's desk measures about three textbook lengths or 5 pencil lengths wide. If a teacher's desk is 4 textbook lengths wide, how wide is a teacher's desk measured in pencil lengths?

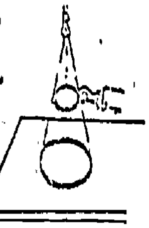
All	0000	1000	1100		
68	59	94	94	[✓ A.] More than 5 because it is bigger than a student desk	4
4	10	0	0	B. Less that 5 because it seems that way	1
11	9	0	6	C. About 4 because it was 4 textbooks	2
7	5	0	0	D. 5 because that is what the student desk measured	3
8	10	6	0	E. I have no answer	0

CHANGES	REASON
Student desk and teacher desk compared in place of paper doll.	More familiar. Wish more succes with this item. Students asked where was the other paper doll.
Simpler integer ratios 10/4 becomes 5/4.	More appropriate to the problem. Intend a simpler problem.

1C1

A ring is held between a table and a light bulb. The light bulb casts a shadow of the ring. If a smaller ring is held in the same place the shadow of the smaller ring would

VERSION II			VERSION III					
All	0000	1000	All	0000	1000	1100		
3	0	0	11	16	0	6	A. Be smaller because the light would change	1
21	55	0	16	21	25	0	B. Be larger because it is different	2
3	9	0	8	10	8	6	C. Be the same size because the ring is in the same place	3
69	36	90	62	42	65	87	[✓ D.] Be smaller because the ring is smaller	4
3	0	10	2	3	2	0	E. I have no answer	0



Responses appear appropriate

CHANGES

None (Wrong key - Version III)

LEVEL II

VERSION 2

VERSION 3

14C₂

These nature hunt groups are chosen for a nature hike. Mrs. Andrews = 5 students
Mr. Denton & Mrs. Folk = 8 students
Mr. Holt = 6 students

The teacher with the most students to help is:

VERSION II All	1110	1100	1000	VERSION III All	1110	1100	1000		
52	100	NO RESPONSE	40	61	87	55	67	(A.) Mr. Holt because $\frac{6}{1}$ is larger than $\frac{5}{1}$ is larger than $\frac{8}{2}$	4
10	0		20	10	3	0	10	B. Mr. Denton & Mrs. Folk because $\frac{2}{8}$ is larger than $\frac{1}{5}$ is larger than $\frac{1}{6}$	2
31	0		30	25	10	45	20	C. Mr. Denton & Mrs. Folk because they have the most students	3
3	0		0	1	0	-	0	D. Mrs. Andrews because she has fewer students	1
3	0		10	2	0	-	2	E. I have no answer	0

CHANGES

None

REASON

Appeared satisfactory - wanted and got about 50% success.

11C₂

Four cars have different speeds: Car A is the fastest, Car B the next fastest, Car C the next fastest, and, Car D the next fastest. The fastest car takes the least time to go 200 miles, the next fastest car the next least time and so on. Which car is the third fastest and takes the third least time to go 200 miles?

VERSION II All	1110	1100	1000	VERSION III All	1110	1100	1000		
59	100	NO RESPONSE	60	53	57	90	25	(A.) Car C because: A 1st fastest ↕ 1st least time; B 2nd fastest ↕ 2nd least time; C 3rd fastest ↕ 3rd least time	4
10	0		20	7	0	0	4	B. Car B	1
14	0		10	12	7	0	6	C. No car because they don't match up	2
17	0		10	16	37	10	6	D. Car C because: A 1st most fast ↕ 1st most time; B 2nd most fast ↕ 2nd most time; C 3rd most fast ↕ 3rd most time	3
0	0		0	10	0	0	58	E. I have no answer	0

CHANGES

None

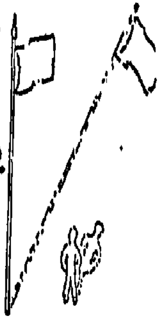
REASON

Appeared satisfactory - wanted and got about 50% success.

6C₂

The large 25 foot flag pole has a shadow 35 feet long. How long a shadow will a 6 foot person have?

All	1110	1100	1000		
62	33	NO RESPONDENTS	70	(A.) About 16 feet because: Flag pole 25 + 10 = 35 so Person 6 + 10 = 16	4
17	67		0	(B.) About 8 feet because the person is less than ¼ as big	3
14	0		20	C. About 7 feet because it should increase like the flag pole	2
0	0		0	D. About 6½ feet because it seems that way	1
3	0		10	E. I have no answer	0



CHANGES (Wrong key)

1. Used the numbers 20, 10 and 5 to give recognizable multiples.
2. "B" here required some contrast of ratio so it was replaced.

A 20 ft. flag pole has a shadow 33 ft. long. A 10 ft. tree has a shadow 25 ft long. How long a shadow will a 5 ft. person have?

All	1110	1100	1000		
38	30	20	68	(A.) About 12 ft. because 38 - 13 = 25 and 25 - 13 = 12	4
28	27	6	17	B. About 12 ft. because it is bigger than the man	2
12	33	16	4	C. About 20 ft. because the man is 5 ft. less.	3
11	10	45	2	D. About 10 ft. because it seems that way	1
11	0	6	8	E. I have no answer	0

REASON

1. Wished more appropriate level - Version II was too hard.
2. Wanted a correct answer obtainable without formal thought.

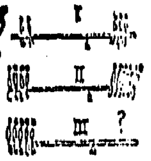
LEVEL II

5C₂

VERSION 2

Trial I 2 people on side "A" balance 3 of the same size people on side "B"
Trial II 4 people on side "A" balance 6 of the same size people on side "B"
Trial III 5 people on side "A" should balance how many on side "B"?

All	1110	1100	1000		
31	33	NO RESPONDENTS	30	A. About 7 because one more on "A" should balance one more on "B"	3
7	33		0	B. About 7 because $\frac{6}{4}$ is less than $\frac{8}{5}$	4
28	33		20	C. About 7 because 4 + 1 = 5 and 6 + 1 = 7	2
17	0		20	D. About 7 because it should be more	1
17	0		30	E. I have no answer	0



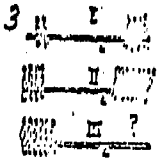
CHANGES

1. Original conditions viz: 2-3 were changed to 2-4
   4-6 were changed to 4-8
   5-? were changed to 6-?

VERSION 3

Trial I 2 people on side "A" balance 4 of the same sized people on side "B"
Trial II 4 people on side "A" balance 8 of the same size people on side "B"
Trial III 6 people on side "A" should balance how many on side "B"?

All	1110	1100	1000		
8	7	0	10	A. About 10 because 2 more on "A" should balance two more on "B"	3
52	73	100	20	B. About 12 because it goes up 4 and 8 + 4 = 12	4
17	0	-	62	C. About 10 because it takes 2 more and 8 + 2 = 10	2
16	10	-	4	D. About 12 because it should be more	1
6	10	-	2	E. I have no answer	0



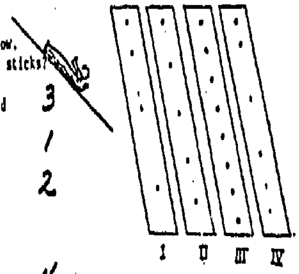
REASON

1. This allowed a correct additive solution since the problem's difficulty was hypothesized to be a result of its use of ratios. Form II was too difficult.

3C₂

This person sliding down a hill looks at her watch. Each second she puts a stick in the snow. What most likely would be the pattern of these sticks?

All	1110	1100	1000		
21	33	NO RESPONDENTS	20	A. II because she travels each second	3
17	0		20	B. III because it is a steep hill	1
21	0		10	C. I II III or IV because she is moving	2
38	67		50	D. I or IV because her speed is changing	4
3	0		0	E. I have no answer	0



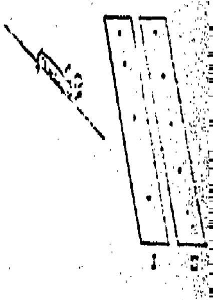
CHANGES

1. Only two examples used in Version III in an attempt to concentrato on reasons.

2. Vocabulary change from travels to moves.

This person sliding down a hill looks at her watch Each second she puts a stick in the snow. What most likely would be the pattern of these sticks?

All	1110	1100	1000	000		
10	0	6	15	13	A. I because she moves each second	3
12	10	0	17	26	B. II because it is a steep hill	1
12	3	3	56	4	C. I or II because she is moving	2
60	87	84	8	43	D. I because her speed is changing	4
6	-	6	4	11	E. I have no answer	0



REASON

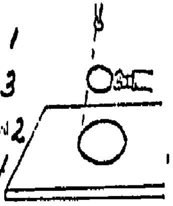
1. Wished to make this question more easily comprehended and answered on the basis of reasons.

2. Student asked about the traveling.

1C₂

A ring is held between a table and a light bulb. The light casts a shadow of the ring onto the table. If the ring is moved, the shadow may:

All	1110	1100	1000		
7	0	NO RESPONDENTS	0	A. Become larger if the ring is closer to the table	1
10	0		0	B. Become smaller if the ring is closer to the light	3
7	0		0	C. Remain the same size regardless of where the ring is placed	2
59	100		70	D. Become larger if the ring is moved closer to the light	4
17	0		30	E. I have no answer	0



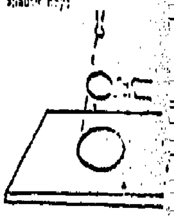
CHANGES

1. Rewording of question stem from "A ring is held between a table and a light bulb" to "If the light is moved closer to the table".

ng of answer and distracters to afford an answer in terms of a physical model.

A ring is held between a table and a light bulb. The light casts a shadow of the ring onto the table. If the ring is moved closer to the table, the shadow may:

All	1110	1100	1000		
21	0	26	15	A. Become larger because the shadow spreads out	1
69	97	74	67	B. Become smaller because the light rays don't spread as much	4
2	0	0	0	C. Stay the same because it's the same ring	2
7	3	0	12	D. Become larger because the bulb is father away	3
2	0	0	6	E. I have no answer	0



REASON

1. Wish to reduce ambiguity of what is desired.

2. Identification of a model is appropriate to this level. The previous answer depended primarily on the experience of the student.

LEVEL III

18F

VERSION 2

8th 1110	1100	11th All	1111	1110	1100
33	NO RESPONDENTS	12	17	0	0
33		57	52	75	0
0		0	0	0	80
33		13	17	11	20
0		16	9	14	0

A movie projector lens spreads its light out over a 3' x 3' screen 8 feet away. To make the image spread over a 5' x 5' screen, how far back must the screen be moved?

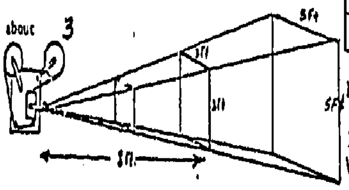
A. About 10 ft. The 5 ft image is 2 more than the 3 ft one, so the 5 ft image should be 2 ft more back. 2

(B.) About 13 ft because 5/3 x 8 = 13 1/3 4

C. About 11 ft because 8 + 3 = 11 1

D. About 15 ft because it should be about twice as far 3

E. I have no answer 0



NGES

Distracter A - From approximate numbers to more explicit
---- 11 feet is 2 more than 9 feet.

Removal of Abbreviations: from ft. to feet.

From 5/3 x 8 = 13 1/3 to 3/9 = 5/15.

Distracter D - removed.

VERSION 3

All	1111	1110	1100
17	5	1	35
52	90	97	49
13	5		10
11	0		3
6	0		3

A movie projector lens spreads its light out over a 3' x 3' screen 9 feet away. To make the image spread over a 5' x 5' screen, how far back must the screen be moved?

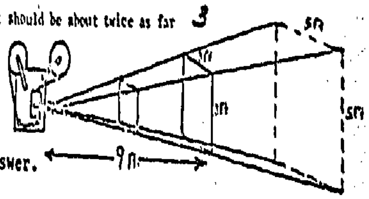
A. About 11 feet. The 5 foot image is 2 more than the 3 foot one and 11 feet is 2 more than 9 feet 2

(B.) About 15 feet because 3/9 = 5/15 4

C. About 12 feet because 9 + 3 = 12 1

D. About 18 feet because it should be about twice as far 3

E. I have no answer 0



REASON

1. I wished to increase the plausibility of the answer. Version 1 was confusing students.

2. Abbreviations could cause confusion.

3. The comparison of the ratio was intended to make this easier and closer to this level.

4. This distracter involves a formal proportion and may inappropriately be attracting formal reasonors.

17F

8th All	1110	1100	11th All	1111	1110	1100	All	1111	1110	1100
21	0	NO RESPONDENTS	0	0	0	20	8	5	3	3
10	0		7	0	6	0	13	5	0	26
52	100		92	100	94	80	68	90	93	71
10	0		1	0	0	0	3		0	0
7	0		0	0	0	0	8		3	0

Jane is weighing out apples on this supermarket scale. What will 14 apples weigh if 6 apples weigh 1½ lbs?

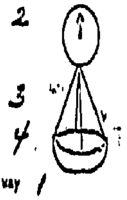
A. 9½ lbs because 6 + 8 = 14
so
1½ + 8 = 9½ 2

B. 3 or 4 lbs because it is more 3

(C.) 3½ lbs because $\frac{1\frac{1}{2}}{6}$ x 14 = 3½ 4

D. 3 or 4 lbs because it looks that way 1

E. I have no answer 0



CHANGES

None - Performance was appropriate

11F

8th All	1110	1100	11th All	1111	1110	1100	All	1111	1110	1100
24	0	NO RESPONDENTS	7	0	8	20	18	5	0	42
10	0		0	0	0	0	4		0	0
45	100		89	100	86	80	65	95	90	52
10	0		0	0	0	0	7		3	3
10	0		4	0	6	0	5		7	3

A car moving at a constant 45 mph travels 198 ft. in 3 seconds. How far will it have traveled by the end of 5 seconds?

A. More than 198 feet because it is still moving 2

B. Less than 198 feet because it is only 2 seconds more 1

(C.) 330 feet because $\frac{198}{3}$ x 5 = 330' 4

D. 200 feet because 3 sec. + 2 sec. = 5 sec.
198 ft. + 2 ft. = 200 ft. 3

E. I have no answer 0

CHANGES

None - Performance was considered appropriate. The item is a good discriminator.

LEVEL III

VERSION 2

VERSION 3

10F,

Grade 8			Grade 11			
All	1110	1100	All	1111	1110	1100
45	100	NO RESPONDENTS	87	96	92	20
28	0		1	0	0	0
21	0		9	4	6	80
3	0		0	0	0	0
3	0		0	0	0	0

CHANGES

None

Jim uses 2 heaping teaspoons of Tang powder with an 8 oz. glass of water. How much Tang is needed for the same mixture with 27 oz. of water?

All	1111	1110	1100		
49	85	57	52	[✓ A.] About 7 teaspoons because $\frac{27}{8}$ x 2 tsp. = 6 3/4 tsp	4
17			45	B. About 21 teaspoons because $\begin{array}{r} 27 \text{ oz} \\ -8 \text{ oz} \\ \hline 19 \text{ oz} \end{array}$ and 2 tsp. + 19 tsp. = 21 tsp.	3
20	15	43	3	C. More than 2 teaspoons because there is more water	2
3				D. 2 teaspoons because it is the same mixture	1
10				E. I have no answer	0

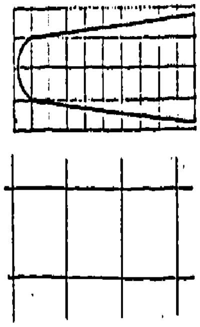
REASON

Tho item seems to be working appropriately.

8F,

Grade 8			Grade 11					
All	1110	1100	All	1111	1110	1100		
7	NO RESPONDENTS	0	5	4	0	20	A. About 3 cm. because it looks that way	3
14		0	0	0	78	0	B. 12½ because 9½ + 3 = 12½	2
55		100	73	74	17	20	[✓ C.] 28½ cm long because 3 x 9½ = 28½	4
17		0	19	22	0	40	D. 9½ cm. but the squares would be larger	1
7		0	3	0	0	20	E. I have no answer	0

A model airplane wing made from the pattern shown measures 9½ cm. long. What would be the length of such a wing made from a pattern with squares that are 3 times as long and 3 times as wide?



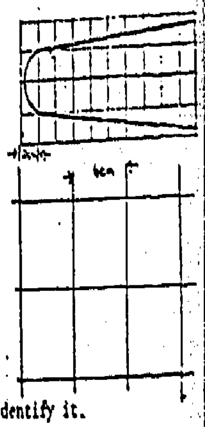
CHANGES

1. Stem was written with measurements rather than the multiple.
Version 11 - ...."squares that are three times as long and ...." Version III - ...."the 2 cm. pattern..."

2. Answers and distractors essentially the same but more integral values.

A model airplane wing made from the 2 cm. pattern shown measures 7 cm. long. What would be the length of such a wing made from a pattern with squares that are 6 cm.?

All	1111	1110	1100		
27	25	20	26	[✓ A.] 57 cm. because 6/2 x 19 = 57	4
18	5	40	29	B. 18 cm. because it looks that way	3
22	20	10	32	C. 22 cm. because 19 + 3 = 22	2
16	10	17	10	D. 19 cm. but the squares would be larger	1
17	40	13	3	E. I have no answer	0



REASON

1. The formal reasoner should infer the multiple rather than just identify it.

2. Students asked questions about answer. It was intended to make this question more discriminating.

2F,

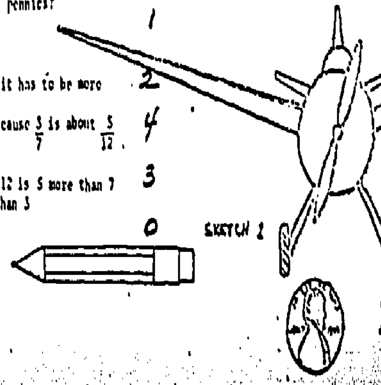
Grade 8			Grade 11			
All	1110	1100	All	1111	1110	1100
10	0	NO RESPONDENTS	1	0	0	20
10	0		1	0	0	0
38	67		88	100	92	40
31	33		1	0	0	0
10	0		8	0	8	40

CHANGES

None

Here is sketch #1 of an airplane. Sketch #1 is 7 pencil widths or 3 pennies high. Sketch #2 of this airplane is not shown. Sketch #2 looks the same but is 12 pencil widths high. How high must sketch #2 be in pennies?

All	1111	1110	1100		
13	0	6	6	A. Seems to be 6	1
17	0	3	23	B. About 7 because it has to be more	2
43	100	87	10	[✓ C.] About 5 because $\frac{3}{7}$ is about $\frac{5}{12}$.	4
22	0	3	55	D. About 8 because 12 is 5 more than 7 and 8 is 5 more than 3	3
5			6	E. I have no answer	0



REASON

Appears to discriminate well.

LEVEL IV

VERSION 3

VERSION 2

19F2

Grade 8			Grade 11		
All	1111	1110	All	1111	1110
10	NO RESPONDENTS	33	24	4	36
10		0	4	4	3
38		0	3	0	6
31		67	57	87	44
10			9	4	6

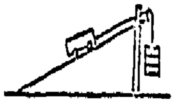
CHANGES

None

On the ramp illustrated the cart and its weight is balanced by weights on the string; What amount of weight is needed to balance 400 g of cart weight at 20^0 ?

All	1111	1110	
20	25	20	A. 133 because $\frac{100}{300} \times 400 = 133$ 3
11	0	3	B. 150 because it is more 1
13	0	17	C. 177 because it goes up 17 for every 100 2
34	75	50	[✓ D.] 133 because $\frac{100}{300} = \frac{133}{400}$ 4
11	0	10	E. I have no answer 0

Angle	Weight: Cart	Weight: String
10^0	200g	33
10^0	300g	52
20^0	500g	100
20^0	400g	?



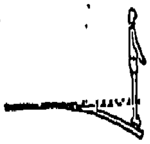
REASON

The item appears to be working appropriately

17F2/15F2 17F2

A 150 pound man standing out on the end of a diving board bends the end of the board down 9 inches. He and his 200 pound companion (total of 350 pounds) bend it 21 incles. How far will the board bend with only the 200 pound person?

Grade 8			Grade 11			
All	1111	1110	All	1111	1110	
69	NO RESPONDENTS	100	48	35	50	A. 12 inches because 21 - 9 = 12 2
14		0	12	9	19	B. 12 inches because $\frac{9}{150} \times 200 = 12$ 3
3		0	1	0	3	C. 12 inches because it is in between 21 and 9 1
10		0	35	57	22	[✓ D.] 12 inches because $\frac{9}{150} = \frac{21}{350} = \frac{12}{200}$ 4
3		0		0	6	E. I have no answer. 0



CHANGES

Replace the item.

15F2

A freeway driver keeps track of the distance he travels. He finds that in 4 minutes he travels 3 miles; in 10 minutes 7½ miles. If he continues at this speed, how long will it take him to travel 10 miles?

Distance	Time
3 miles	4 min.
7½ miles	10 min.
10 miles	? min.

All	1111	1110	
18	30	17	A. About 13 minutes because $\frac{4}{3} \times 10 = 13\frac{1}{3}$ 3
13	5	3	B. About 13 minutes because 10 - 7½ = 2½ miles & 10 + 2½ = 12½ min. 1
25	0	20	C. About 14 because 7½ + 3 = 10½ & 10 + 4 = 14 2
38	65	50	[✓ D.] About 13 minutes because $\frac{4\text{ min.}}{3\text{ miles}} = \frac{10\text{ min.}}{7.5\text{ miles}} = \frac{13\ 1/3\text{ min.}}{10\text{ miles}}$ 4
5		10	E. I have no answer 0

REASON

The item does not discriminate appropriately - too easy but appears to attract an undesired response.

11F2/10F2 11F2

A car is moving along the street at a steady 30 mph. An observer measures these travel distances:

Seconds	Feet
0	0
2	88
5	220

How long will it take the car to travel 400 feet?

Grade 8			Grade 11			
All	1111	1110	All	1111	1110	
38	NO RESPONDENTS	33	31	17	39	A. About 9 seconds because 220 ft. 5 secs. 88 ft. 2 secs. 88 ft. 2 secs. 396 ft. = 9 secs. 2
14		0	5	0	6	B. About 9 seconds because it should be more 1
17		67	31	26	33	C. About 9 seconds because $\frac{88}{2} \times 9 = 396$ 3
28		0	31	57	19	[✓ D.] About 9 seconds because $\frac{2}{88}$ or $\frac{5}{220}$ is about $\frac{9}{400}$ 4
3		0	3	0	3	E. I have no answer 0

CHA[NGES]

Replace the item

10F2

Here is a recipe for 4 cups of cocoa : Heat to near boiling 4 c. milk
Add with stirring 6 T. sugar
5 T. Cocoa

How many tablespoons of sugar would be needed to make 12 cups of this cocoa?

All	1111	1110	
20	0	3	A. 18 tablespoons because $\frac{6}{4} \times 12 = 18$ 3
6	0	3	B. More than 6 tablespoons because there is more cocoa 1
62	100	90	[✓ C.] 18 tablespoons because $\frac{6\text{ T. sugar}}{4\text{ c. cocoa}} = \frac{18\text{ T. sugar}}{12\text{ c. cocoa}}$ 4
7		3	D. 14 tablespoons because 4 c. + 8 c. = 12 c. so 6 T. + 8 T. = 14 T. 2
5			E. I have no answer 0

REASON

The item does not appropriately discriminate.

LEVEL IV

VERSION 2

VERSION 3

9G2

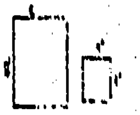
Grade 8			Grade 11		
All	1111	1110	All	1111	1110
24	NO RESPONDENTS	33	47	83	39
34		67	40	17	50
28		0	4	0	6
7		0	4	0	3
7		0	4	0	3

CHANGES

None

Imagine that concrete has been mixed to make a patio 4 ft. x 4 ft. and ½ a foot thick. How thick will this concrete be if it is instead spread out over an 8 ft. x 8 ft. area?

All	1111	1110		
21	75	7	(A.) $\frac{1}{8}$ ft. thick because $\frac{16}{64} = \frac{\frac{1}{8}}{\frac{1}{2}}$	4
49	5	63	B. $\frac{1}{2}$ ft. thick because $\frac{4}{8} = \frac{\frac{1}{4}}{\frac{1}{2}}$	3
21	20	17	C. $\frac{1}{4}$ ft. thick because $\frac{1}{4}$ is less than $\frac{1}{2}$	2
9	0	13	D. $\frac{1}{4}$ ft. thick because it should be less	1
0	0	0	E. I have no answer	0



REASON

The item discrimates appropriately.

4F2/5F2

4F2

Grade 8			Grade 11		
All	1111	1110	All	1111	1110
7	NO RESPONDENTS	33	4	0	6
24		67	64	96	50
38		0	4	0	0
21		0	27	4	42
10		0	1	0	3

The "O" rod shown here in picture "A" crosses 16 lines. The "Y" rod crosses 10 lines. The "O" rod, when turned, crosses 12 lines in picture B. How many lines would the "Y" rod cross if it were turned at this angle?

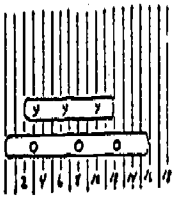
A. About 7 because $\frac{12}{16}$ is greater than $\frac{7}{10}$ 2

(B.) About 7 or 8 because $\frac{12}{16} = \frac{7.5}{10}$ 4

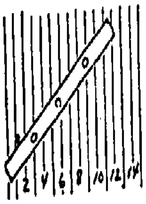
C. About 6 because 16 - 10 = 6
12 - 6 = 6 1

D. About 7 or 8 because $\frac{12}{16} \times 10 = 7.5$ 3

E. I have no answer 0



PICTURE A



PICTURE B

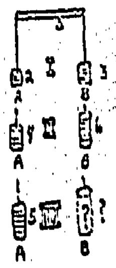
CHANGES

Drop the item and replace it.

5F2

Trial I - Two weights on side "A" balance three of the same weights on side "B"
Trial II - Four weights on side "A". Six on side "B"
Trial III- Five weights on side "A" then should balance how many weights on side "B"?

All	1111	1110		
17	0	13	A. About 8 because $\frac{6}{4} \times 5 = 7.5$	3
34	95	23	(B.) About 8 because $\frac{6}{4} = \frac{7.5}{5}$	4
28	0	13	C. About 7 because 4 + 1 = 5 & 6 + 1 = 7	1
15	5	47	D. About 7 because $\frac{6}{4} < \frac{8}{5}$	2
6	0	3	E. I have no answer	0



REASON

The content appears too complex - it may be adding confusion.

1F2

Grade 8			Grade 11					
All	1111	1110	All	1111	1110	All	1111	1110
0	NO RESPONDENTS	0	5	0	8	8	0	3
14		33	41	74	36	21	50	13
34		67	32	17	39	25	5	40
38		0	12	9	8	31	45	30
14		0	8	0	6	15	0	13

A ring 3 inches across is 2 feet from the light and 4 feet from the table. The 3" ring has a 9" shadow. Where should a 4" ring be placed to make the same size shadow?

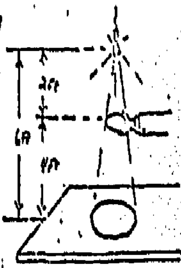
A. The shadow will be larger than 9" wherever the ring is placed. 1

(B.) About 3 ft. from the lamp because $\frac{2}{3} = \frac{2.7}{4}$ 4

C. About 3 ft. from the lamp because $\frac{2}{3} \times 4 = 2.7$ 3

D. About 3 ft. from the lamp because the ring is 1" larger 3 + 1 = 4 and 2ft. + 1 ft. = 3 ft. 2

E. I have no answer 0



CHANGES

None

REASON

The item appears to discriminate appropriately although it is difficult.

LEVEL V

(ADDITIONAL C_2 ITEMS)

VERSION 3

17C_2

The heaviest person is the slowest walker. Sally is heavier than Sue who is heavier than Fran who is a slower walker than Alice. Which person is the 3rd heaviest and the 3rd slowest walker?

All	1000	1100		
20	19	23	A. Fran because 1-Sally 2-Sue 3-Fran	3
8	17	6	B. None because weight / walking: 1 Sally / Alice, 2 Sue / Fran, 3 Fran / Sue, 4 Alice / Sally	1
44	56	35	[✓ C.] Fran because most weight / slowest walking: 1 Sally / Sally, 2 Sue / Sue, 3 Fran / Fran, 4 Alice / Alice	4
17	6	35	D. Sue because least weight / fastest walking: 1 Alice / Alice, 2 Fran / Fran, 3 Sue / Sue, 4 Sally / Sally	2
11	2		E. I have no answer	0

16C_2

Here is a listing of some metric and English measure:

4 inches = 10.2 cm
12 inches = 30.6 cm
_____ = 50 cm

All	1000	1100		
8	2	3	A. About 18 because it has to be more	3
54	81	61	[✓ B.] About 20 because 30 cm + 10 cm + 10 cm = 50 cm and 12 in + 4 in + 4 in = 20 in	4
11	0	6	C. About 19 because it seems that way	1
15	6	26	D. About 32 because 30 cm + 20 cm = 50 cm and 12 in + 20 in = 32 in	2
12	10	3	E. I have no answer	0

15C_2

Sue always drives home on the freeway. Her speed is different each day. Monday is her slowest, Tuesday her next slowest, Wednesday her next slowest, Thursday her next slowest and Friday next slowest. Friday it takes the least time to get home, Thursday the next least, Wednesday the next least and so on. On which day does it take the second least time and is it the second most slow?

All	1000	1100		
22	10	52	A. Thursday or Tuesday because they are second from each end of the week	3
28	23	16	B. Thursday because most speed / most time: 1 Fri. / Fri., 2 Thurs. / Thurs., 3 Wed. / Wed., 4 Tues. / Tues., 5 Mon. / Mon.	2
2	6	0	C. Wednesday because it is the middle	1
34	58	10	[✓ D.] No one day because most time / most speed: 1 Fri. / Mon., 2 Thurs. / Tues., 3 Wed. / Wed., 4 Tues. / Tues., 5 Mon. / Fri.	4
		23	E. I have no answer	0

10C_2

Here are some recipes for Kool Aide

	1 quart	4 quarts	5 quarts
Kool Aide Powder	½ pkg	2 pkg	?
Sugar	¼ c	1 c	
Water	1 qt	4 qts	

How much powder is needed for 5 quarts of Kool Aide

All	1000	1100	1111		
75	90	97	100	[✓ A.] 2½ pkg because 4 qts + 1 qt = 5 qts and 2 pkg + ½ pkg = 2½ pkg	4 2
10	6	3	0	B. 3 pkg because 4 qts + 1 qt = 5 qts and 2 pkg + 1 pkg = 3 pkg	3
8	4	0	0	C. About 3 because it would have to be more	1
2	0	0	0	D. ½ pkg because it is the same mixture	0
5	0	0	0	E. I have no answer	

8C_2

A 12 inch television screen has 80 sq. inches of screen. A 21 inch set should have ____?

All	1000	1100	1111		
28	25	32	30	[✓ A.] About 240 sq. inches because 12 x 12 = 144 and 21 x 21 = 441 (3 times as much)	4
3	4	0	0	B. The same but with larger squares	2
3	4	0	0	C. Less than 80 sq. inches because the squares are larger	1
60	65	65	70	D. More than 80 sq. inches because it is larger	3
6	2	3	0	E. I have no answer	0

LEVEL I

VERSION 3

VERSION 4

14C$_1$

VERSION III			VERSION IV					
All	0000	1000	All	0000	1000	1100	Mary buys 3 tickets to a raffle where 90 tickets are sold --- Jane buys 1 ticket to a raffle where 30 tickets are sold --- Sue buys 3 tickets to a raffle where 300 tickets are sold. Which girls have about the same chance of winning?	
7	15	39	13	18	8	6	A. Jane and Mary because their's are the least tickets	2
10	24	6	9	12	25	6	B. Sue and Mary because each have 3 tickets	3
15	24	0	5	26	0	0	C. All girls have the same chance	1
63	24	94	68	35	69	89	[✓ D.] Jane and Mary because 3 chances in 90 is the same as 1 in 30	4
			4	0	0	0	E. I have no answer	0

CHANGES

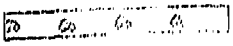
None

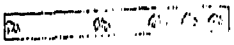
REASON

Responses appeared appropriate

11C$_1$

All	0000	1000	A car moving at a constant speed of 30 mph will, if pictured at one second intervals, look like	
62	9	90	[✓ A.] I because it moves equal distances each second	4
3	0	10	B. None of these because it is moving	1
21	55	0	C. II because it changes	2
10	27	0	D. III because it is increasing its distance	3
3	9	0	E. I have no answer	0

I


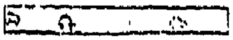
II


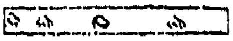
III

CHANGE

Reduce to only two illustrations.

All	0000	1000	1100	A car moving at a constant speed of 30 mph will, if pictured at one second intervals, look like	
71	29	69	100	[✓ A.] I because it moves equal distances each second	4
5	12	15	0	B. None of these because it is moving	1
10	29	0	0	C. II because it changes	2
5	12	0	0	D. "I because it is increasing its distance	3
6	18	7	0	E. I have no answer	0
0	0	8	0	No response	

I


III


REASON

Wish to concentrate on results
Wish to increase correct responses.

9C$_1$

All	0000	1000	All	0000	1000	1100	Imagine that frosting had been spread out ¼ inch thick on top of a small 6" x 6" cake. Predict what the thickness would be if the same amount of frosting were spread out over a 12" x 12" cake?	
10	18	10	6	12	0	11	A. More than ¼ inch because it covers less cake	1
0	0	0	8	18	8	6	B. Less than ¼ inch because it looks that way	3
69	55	70	69	35	85	72	[✓ C.] Less than ¼ inch because it covers more cake	4
14	18	10	8	18	8	6	D. More than ¼ inch because there is more cake	2
7	9	10	6	6	0	6	E. I have no answer	0

CHANGE

None

REASON

The item responses

LEVEL I

VERSION 3

VERSION 4

4C₁

All	0000	1000
0	0	0
10	27	0
7	18	0
72	27	100
10	27	0

The "O" rod here crosses 8 lines. The "Y" rod crosses 5 lines. The "O" rod, when turned, crosses 6 lines. How many lines would the "Y" rod cross if it were at this angle?

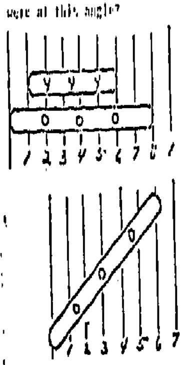
A. About 8 because $\frac{8}{5} \times 5 = 8$ 3

B. About 5 because the "Y" rod is that long 1

C. About 6 because the "O" rod was 6 2

[✓ D.] About 4 because the "Y" rod is shorter 4

E. I have no answer 0



CHANGE

Answer "A" rewritten without including the proportion.

All	0000	1000	1100
4	12	0	0
14	35	0	0
6	18	0	6
62	0	92	72
15	0	8	17

The "O" rod here crosses 8 lines. The "Y" rod crosses 5 lines. The "O" rod, when turned, crosses 6 lines. How many lines would the "Y" rod cross if it were at this angle?

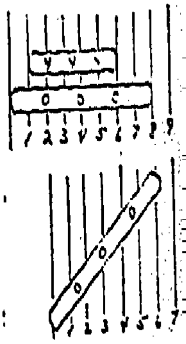
A. About 8 because it should get longer 3

B. About 5 because the "Y" rod is that long 1

C. About 6 because the "O" rod was 6 2

[✓ D.] About 4 because the "Y" rod is shorter 4

E. I have no answer 0



REASON

The problem in the original version suggests thinking inappropriate for this level.
Wish to make this more discriminating.

2C₁

VERSION III				VERSION IV			
All	0000	1000	1100	All	0000	1000	1100
68	59	94	94	66	41	69	89
4	10	0	0	10	35	0	6
11	9	0	6	5	18	0	0
7	5	0	0	4	6	0	0
8	10	6	0	14	0	31	6

A student's desk measures about three textbook lengths or 5 pencil lengths wide. If a teacher's desk is 4 textbook lengths wide, how wide is a teacher's desk measured in pencil lengths?

[✓ A.] More than 5 pencils because it is bigger than a student desk 4

B. Less than 5 pencils because it seems that way 1

C. About 4 pencils because it was 4 textbooks 2

D. 5 pencils because that is what the student desk measured 3

E. I have no answer 0

CHANGE

None

REASON

Appeared to discriminate appropriately.

1C₁

VERSION III				VERSION IV			
All	0000	1000	1100	All	0000	1000	1100
11	16	0	6	6	6	8	6
16	21	25	0	14	35	8	11
8	10	8	6	19	35	15	17
64	42	65	87	55	6	69	61
2	3	2	0	5	18	0	6

A ring is held between a table and a light bulb. The light bulb casts a shadow of the ring. If a smaller ring is held in the same place the shadow of the smaller ring would

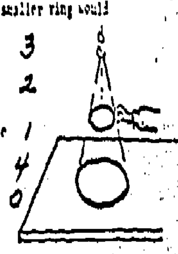
A. Be smaller because the light would change 3

B. Be larger because it is different 2

C. Be the same size because the ring is in the same place 1

[✓ D.] Be smaller because the ring is smaller 4

E. I have no answer 0



CHANGE

None

REASON

Appears to discriminate appropriately.

VERSION 3

VERSION 4

14C2

VERSION III All	1000	1100	1110	VERSION IV All	1000	1100	1110	
61	67	55	87	77	62	94	93	
10	10	0	3	8	8	0	0	
25	20	45	10	13	23	66	7	
1	0	0	0	3	8	0	0	
2	2	0	0	0	0	0	0	

These nature hunt groups are chosen for a nature hike.

Mrs. Andrews	- 5 studen
Mr. Denton & Mrs. Folk	- 8 studen
Mr. Holt	- 6 studen

The teacher with the most students to help is:

[✓ A.] Mr. Holt because $\frac{6}{1}$ is larger than $\frac{5}{1}$ is larger than $\frac{8}{2}$ 4

B. Mr. Denton & Mrs. Folk because $\frac{2}{8}$ is larger than $\frac{1}{5}$ is larger than $\frac{1}{6}$ 2

C. Mr. Denton & Mrs. Folk because they have the most students 3

D. Mrs. Andrews because she has fewer students 1

E. I have no answer 0

CHANGE

None

REASON

Seemingly appropriate discrimination.

11C2

Four cars have different speeds: Car A is the fastest, Car B the next fastest, Car C the next fastest, and, Car D the next fastest. The fastest car takes the least time to go 200 miles, the next fastest car the next least time and so on. Which car is the third fastest and takes the third least time to go 200 miles?

All	1000	1100	1110	
53	23	90	57	[✓ A.] Car C because: A 1st fastest ↕ 1st least time; B 2nd fastest ↕ 2nd least time; C 3rd fastest ↕ 3rd least time 4
7	4	0	0	B. Car B 1
12	6	0	7	C. No car because they don't match up 2
16	6	10	37	D. Car C because: A 1st most fast ↕ 1st most time; B 2nd most fast ↕ 2nd most time; C 3rd most fast ↕ 3rd most time 3
10	58	0	0	E. I have no answer 0

CHANGE

Remove arrows and write out Car A etc....

Four cars have different speeds: Car A is the fastest, Car B the next fastest, Car C the next fastest, and, Car D the next fastest. The fastest car takes the least time to go 200 miles, the next fastest car the next least time and so on. Which car is the third fastest and takes the third least time to go 200 miles?

All	1000	1100	1110	
51	15	67	93	[✓ A.] Car C because: 1st fastest CAR A 1st least time; 2nd fastest CAR B 2nd least time; 3rd fastest CAR C 3rd least time 4
13	31	0	0	B. Car B 1
10	23	0	0	C. No car because they don't match up 2
14	8	22	0	D. Car C because: 1st most fast CAR A 1st most time; 2nd most fast CAR B 2nd most time; 3rd most fast CAR C 3rd most time 3
12	23	11	7	E. I have no answer 0

REASON

Reduce ambiquity.

6C2

A 20 ft. flag pole has a shadow 38 ft. long. A 10 ft. tree has a shadow 25 ft long. How long a shadow will a 5 ft. person have?

All	1000	1100	1110	
35	65	20	30	[✓ A.] About 12 ft. because 38 - 13 = 25 and 25 - 13 = 12 4
28	17	6	27	B. About 12 ft. because it is bigger than the man 2
12	4	16	33	C. About 20 ft. because the man is 5 ft. less. 3
11	2	45	10	D. About 10 ft. because it seems that way 1
11	8	6	0	E. I have no answer 0

CHANGE

is question

10C2

Here are some recipes for Kool Aide

	1 quart	4 quarts	5 quarts
Kool Aide Powder	½ pkg	2 pkg	?
Sugar	½ c	1 c	
Water	1 qt	4 qts	

How much powder is needed for 5 quarts of Kool Aide

All	1000	1100	1110	
5	8	0	7	A. ½ pkg because it is the same mixture 1
16	38	22	0	B. 3 pkg because 4 qts + 1 qt = 5 qts and 2 pkg + 1 pkg = 3pkg 2
9	15	0	0	C. About 3 because it would have to be more 3
68	31	78	93	[✓ D.] 2½ pkg because 4 qts + 1 qt = 5 qts and 2 pkg + ½ pkg = 2½ pkg 4
3	8	0	0	E. I have no answer 0

REASON

Previous change was destructive. The question ($6C_2$) negatively discrimates.

LEVEL II

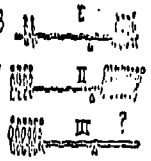
VERSION 3

VERSION 4

5C$_2$

Trial I 2 people on side "A" balance 4 of the same size people on side "B"
Trial II 4 people on side "A" balance 8 of the same size people on side "B"
Trial III 6 people on side "A" should balance how many on side "B"?

All	1000	1100	1110		
8	10	0	7	A. About 10 because 2 more on "A" should balance two more on "B"	3
52	20	100	73	[✓ B.] About 12 because it goes up 4 and 8 + 4 = 12	4
17	62	0	0	C. About 10 because it takes 2 more and 8 + 2 = 10	2
6	4	0	10	D. About 12 because it should be more	1
6	2	0	10	E. I have no answer	0

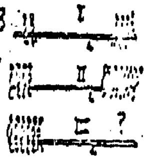


CHANGE

Distractor "D" changed from 12 to 11.

Trial I 2 people on side "A" balance 4 of the same size people on side "B"
Trial II 4 people on side "A" balance 8 of the same site people on side "B"
Trial III 6 people on side "A" should balance how many on side "B"?

All	1000	1100	1110		
9	8	11	0	A. About 10 because 2 more on "A" should balance two more on "B"	3
68	69	78	93	[✓ B.] About 12 because it goes up 4 and 8 + 4 = 12	4
8	8	6	0	C. About 10 because it takes 2 more and 8 + 2 = 10	2
5	8	0	0	D. About 11 because it should be more	1
9	0	6	7	E. I have no answer	0



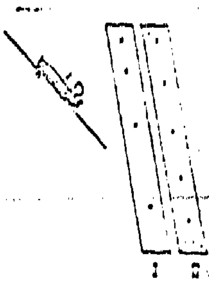
REASON

Wished to have "D" be a more plausible guess.

3C$_2$

VERSION III				VERSION IV					
All	1000	1100	1110	All	1000	1100	1110		
10	15	6	0	6	15	6	0	A. I because she moves each second	2
12	17	0	10	17	23	6	21	B. II because it is a steep hill	1
12	56	3	3	13	15	0	7	C. I or II because she is moving	3
60	8	84	87	60	38	89	64	[✓ D.] I because her speed is changing	4
6	4	6	0	4	8	0	7	E. I have no answer	0

This person sliding down a hill looks at her watch Each second she puts a stick in the snow. What most likely would be the pattern of these sticks?



CHANGE

None

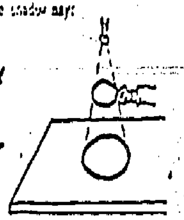
REASON

The problem appears easy yet it does discriminate. When the the results for Grade 5 students (non masters) is examined it appears to be an appropriate question.

1C$_2$

VERSION III				VERSION IV					
All	1000	1100	1110	All	1000	1100	1110		
21	15	26	0	22	30	17	21	A. Become larger because the shadow spreads out	1
69	67	74	97	47	23	56	71	[✓ B.] Become smaller because the light rays don't spread as much	4
2	0	0	0	8	0	11	0	C. Stay the same because it's the same ring	2
7	13	0	3	19	46	17	7	D. Become larger because the bulb is father away	3
2	6	0	0	4	0	0	0	E. I have no answer	0

A ring is held between a table and a light bulb. The light casts a shadow of the ring onto the table. If the ring is moved closer to the table, the shadow may:



CHANGE

None

REASON

Appears nearly too easy.yet does discriminate. Scores of Grade 5 (non-masters) are lower.

LEVEL III

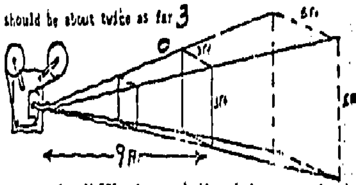
VERSION 3

VERSION 4

18F$_1$

VERSION III				VERSION IV					
All	1100	1110	1111	All	1100	1110	1111		
17	35	1	5	16	44	7	0	A. About 11 feet. The 5 foot image is 2 more than the 3 foot one and 11 feet is 2 more than 9 feet	2
52	49	97	90	58	50	93	100	[✓ B.] About 15 feet because 3/9 = 5/15	4
13	10	0	5	12	6	)	0	C. About 12 feet because 9 + 3 = 12	1
11	3	0	0	0	0	0	0	D. About 18 feet because it should be about twice as far	3
6	3	0	0	6	0	0	0	E. I have no answer	0

A movie projector lens spreads its light out over a 3' x 3' screen 9 feet away. To make the image spread over a 5' x 5' screen, how far back must the screen be moved?



CHANGE

None

REASON

The item has reasonable overall difficulty and discriminates well.

17F$_1$

Jane is weighing out apples on this supermarket scale. What will 14 apples weigh if 6 apples weigh 1½ lbs?

All	1100	1110	1111		
8	3	3	5	A. 9½ lbs because 6 + 8 = 14 so 1½ + 8 = 9½	2
13	26	0	5	B. 3 or 4 lbs because it is more	3
68	71	93	90	[✓ C.] 3½ lbs because $\frac{1½}{6}$ x 14 = 3½	4
3	0	0	0	D. 3 or 4 lbs because it looks that way	1
8	0	3	0	E. I have no answer	0

CHANGE

"D" from a guess question to an addition type answer.

Jane is weighing out apples on this supermarket scale. What will 14 apples weigh if 6 apples weigh 1½ lbs?

All	1100	1110	1111		
9	5	0	0	A. 9½ lbs because 6 + 8 = 14 so 1½ + 8 = 9½	2
23	44	14	0	B. 3 or 4 lbs because it is more	3
34	28	71	67	[✓ C.] 3½ lbs because $\frac{1½}{6}$ x 14 = 3½	4
27	22	14	33	D. 3½ because 1½ + 1½ + ½ = 3½	1
4	0	0	0	E. I have no answer	0

REASON

This gives a clear distractor for a Level 2 reasoner.
The question previously came across too easy probably because it lacked this type of distractor.

11F$_1$

A car moving at a constant 45 mph travels 198 ft. in 3 seconds. How far will it have traveled by the end of 5 seconds?

All	1100	1110	1111		
17	22	0	0	A. More than 198 feet because it is still moving	2
18	17	13	33	B. Less than 400 feet because it is only 2 seconds more	1
42	53	79	67	[✓ C.] 330 feet because $\frac{198}{3}$ x 5 = 330	4
18	28	-	0	D. 200 feet because 3 sec. + 2 sec. = 5 sec. 198 ft. + 2 ft. = 200 ft.	3
5	0	0	0	E. I have no answer	0

REASON

s did not select this distractor. The problem appeared to be too easy.

A car moving at a constant 45 mph travels 198 ft. in 3 seconds. How far will it have traveled by the end of 5 seconds?

All	1100	1110	1111		
18	42	0	5	A. More than 198 feet because it is still moving	2
4	0	0	0	B. Less than 198 feet because it is only 2 seconds more	1
65	52	90	95	[✓ C.] 330 feet because $\frac{198}{3}$ x 5 = 330	4
7	3	3	0	D. 200 feet because 3 sec. + 2 sec. = 5 sec. 198 ft. + 2 ft. = 200 ft.	3
5	3	7	0	E. I have no answer	0

CHANGE

"B" from 198 feet to 400 feet to make it a plausible answer.

LEVEL III

VERSION 3 | VERSION 4

10F1

Jim uses 2 heaping teaspoons of Tang powder with an 8 oz. glass of water. How much Tang is needed for the same mixture with 27 oz. of water?

All	1100	1110	1111	All	1100	1110	1111		
49	52	57	85	48	22	93	67	(A.) About 7 teaspoons because $\frac{27}{8} \times 2$ tsp. = 6 3/4 tsp	4
17	45		0	23	33	0	33	B. About 21 teaspoons because $\begin{array}{r} 27 \text{ oz} \\ -8 \text{ oz} \\ \hline 19 \text{ oz} \end{array}$ and 2 tsp. + 19 tsp. = 21 tsp.	3
20	3	43	15	22	39	7	0	C. More than 2 teaspoons because there is more water	2
3	0	0	0	4	0	0	0	D. 2 teaspoons because it is the same mixture	1
10	0	0	0	1	6	0	0	E. I have no answer	0

CHANGE

None

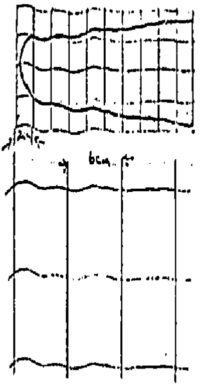
REASON

The item has discrimination. It appears too hard but more use was desired.

8F1

A model airplane wing made from the 2 cm. pattern shown measures 7 cm. long. What would be the length of such a wing made from a pattern with squares that are 6 cm.?

All	1100	1110	1111		
27	26	20	25	(A.) 57 cm. because 6/2 x 19 = 57	4
18	29	40	5	B. 18 cm. because it looks that way	3
22	32	10	20	C. 22 cm. because 19 + 3 = 22	2
16	10	17	10	D. 19 cm. but the squares would be larger	1
17	3	13	40	E. I have no answer	0

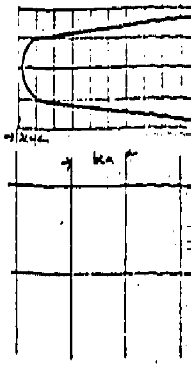


CHANGE

Wing length 7 cm. changed to 19 cm.

A model airplane wing made from the 2 cm. pattern shown measures 19 cm. long. What would be the length of such a wing made from a pattern with squares that are 6 cm.?

All	1100	1110	1111		
37	33	57	33	(A.) 57 cm. because 6/2 x 19 = 57	4
5	0	7	0	B. 18 cm. because it looks that way	3
14	11	7	0	C. 22 cm. because 19 + 3 = 22	2
26	39	29	67	D. 19 cm. but the squares would be larger	1
16	17	0	0	E. I have no answer	0



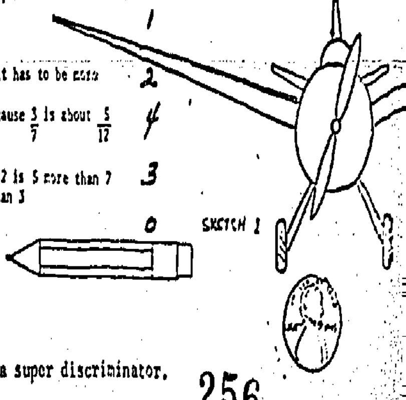
REASON

This was an error in the stem. The problem comes off as too hard.

2F1

Here is sketch #1 of an airplane. Sketch #1 is 7 pencil widths or 3 pennies high. Sketch #2 of this airplane is not shown. Sketch #2 looks the same but is 12 pencil widths high. How high must sketch #2 be in pennies?

All	1100	1110	1111	All	1100	1110	1111		
13	6	6	0	8	6	0	0	A. Seems to be 6	1
17	23	3	0	9	17	0	0	B. About 7 because it has to be more	2
43	10	87	100	35	28	79	100	(C.) About 5 because $\frac{3}{7}$ is about $\frac{5}{12}$	4
22	55	3	0	23	33	14	0	D. About 8 because 12 is 5 more than 7 and 8 is 5 more than 3	3
5	6	0	0	23	17	7	0	E. I have no answer	0



CHANGE

Nono

REASON

Appears to be a super discriminator.

LEVEL IV

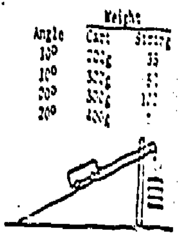
VERSION 3

VERSION 4

19F₂

On the ramp illustrated the cart and its weight is balanced by weights on the string. What amount of weight is needed to balance 400 g of cart weight at 20° ?

All	1110	1111	All	1110	1111		
20	20	25	22	29	0	A. 133 because $\frac{100}{300} \times 400 = 133$	3
11	3	0	9	0	0	B. 153 because it is more	1
23	17	0	25	29	0	C. 177 because it goes up 17 for every 100	2
34	50	75	21	36	100	(D.) 133 because $\frac{100}{300} = \frac{133}{400}$	4
11	10	0	22	7	0	E. I have no answer	0



CHANGE

None

REASON

The item appears to be discriminating appropriately.

15F₂

A freeway driver keeps track of the distance he travels. He finds that in 4 minutes he travels 3 miles/ in 10 minutes 7½ miles. If he continues at this speed, how long will it take him to travel 10 miles?

Distance	Time
3 miles	4 min
7½ miles	10 min
10 miles	? min

All	1110	1111		
34	29	100	(A.) About 13 minutes because $\frac{4 \text{ min.}}{3 \text{ miles}} = \frac{10 \text{ min.}}{7.5 \text{ miles}} = \frac{13\ 1/3 \text{ min.}}{10 \text{ miles}}$	4
21	21	0	B. About 13 minutes because 10 - 7½ = 2½ miles and 10 + 2½ = 12½ min.	1
18	29	0	C. About 13 minutes because $\frac{4}{3} \times 10 = 13\ 1/3$	3
21	14	0	D. About 14 because 7½ + 3 = 10½ and 10-+ 4 = 14	2
6	7	0	E. I have no answer	0

PEASON

Wanted the student tó view the correct answer sooner.

A freeway driver keeps track of the distance he travels. He finds that in 4 minutes he travels 3 miles; in 10 minutes 7½ miles. If he continues at this speed, how long will it take him to travel 10 miles?

Distance	Time
3 miles	4 min.
7½ miles	10 min.
10 miles	? min.

All	1110	1111		
18	17	30	A. About 13 minutes because $\frac{4}{3} \times 10 = 13\frac{1}{3}$	3
13	3	5	B. About 13 minutes because 10 - 7½ = 2½ miles & 10 + 2½ = 12½ min.	1
25	20	0	C. About 14 because 7½ + 3 = 10½ & 10 + 4 = 14	2
38	50	65	(D.) About 13 minutes because $\frac{4 \text{ min.}}{3 \text{ miles}} = \frac{10 \text{ min.}}{7.5 \text{ miles}} = \frac{13\ 1/3 \text{ min.}}{10 \text{ miles}}$	4
5	10	0	E. I have no answer	0

CHANGE

Switched order: A to C, D to A, and C to D.

10F₂

Here is a recipe for 4 cups of cocoa : Heat to near boiling 4 c. milk
Add with stirring 6 T. sugar
5 T. Cocoa

How many tablespoons of sugar would be needed to make 12 cups of this cocoa?

All	1110	1111		
20	3	0	A. 18 tablespoons because $\frac{6}{4} \times 12 = 18$	3
6	3	0	B. More than 6 tablespoons because there is more cocoa	1
62	90	100	(C.) 18 tablespoons because $\frac{6 \text{ T. sugar}}{4 \text{ c. cocoa}} = \frac{18 \text{ T. sugar}}{12 \text{ c. cocoa}}$	4
7	3	0	D. 14 tablespoons because 4 c. + 8 c. = 12 c. so 6 T. + 8 T. = 14 T.	2
5	0	0	E. I have no answer	0

S

ved language from distractor "C".

Here is a recipe for 4 cups of cocoa: Heat to near boiling 4 c. milk
Add with stirring 6 T. sugar
5 T. Cocoa

How many tablespoons of sugar would be needed to make 12 cups of this cocoa?

All	1110	1111		
26	50	0	A. 18 tablespoons because $\frac{6}{4} \times 12 = 18$	3
12	0	0	B. More than 6 tablespoons because there is more cocoa	1
38	36	100	(C.) 18 tablespoons because $\frac{6}{4}$ equals $\frac{18}{12}$	4
16	14	0	D. 14 tablespoons because 4 c. + 8 c. = 12 c. so 6 T. + 8 T. = 14 T.	2
8	0	0	E. I have no answer	0

REASON

The problem came across as too easy. It was suspected that the words with answer "C" might have been a cause.

LEVEL IV

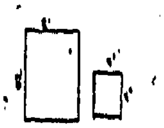
VERSION 3

VERSION 4

9G₂

Imagine that concrete has been mixed to make a patio 4 ft. x 4 ft. and ½ a foot thick. How thick will this concrete be if it is instead spread out over an 8 ft. x 8 ft. area?

All	1110	1111	All	1110	1111		
21	7	75	19	7	67	(A.) $\frac{1}{8}$ ft. thick because $\frac{16}{64} = \frac{\frac{1}{8}}{\frac{1}{2}}$	4
49	63	5	38	71	33	B. $\frac{1}{4}$ ft. thick because $\frac{4}{8} = \frac{\frac{1}{4}}{\frac{1}{2}}$	3
21	17	20	19	14	0	C. $\frac{1}{4}$ ft. thick because $\frac{1}{4}$ is less than $\frac{1}{2}$	2
9	13	0	8	0	0	D. $\frac{1}{4}$ ft. thick because it should be less	1
0	0	0	12	0	0	E. I have no answer	0



CHANGE

None

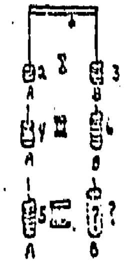
REASON

This item exhibits good discrimination.

5F₂

Trial I - Two weights on side "A" balance three of the same weights on side "B"
Trial II - Four weights on side "A". Six on side "B"
Trial III- Five weights on side "A" then should balance how many weights on side "B"?

All	1110	1111	All	1110	1111		
17	13	0	18	43	33	A. About 8 because $\frac{6}{4} \times 5 = 7.5$	3
34	23	95	24	7	67	(B.) About 8 because $\frac{6}{4} = \frac{7.5}{5}$	4
28	13	0	18	7	0	C. About 7 because 4 + 1 = 5 ↓ 6 + 1 = 7	1
15	47	5	14	36	0	D. About 7 because $\frac{6}{4}$ is less than $\frac{8}{5}$	2
6	3	0	22	7	0	E. I have no answer	0



CHANGE

None

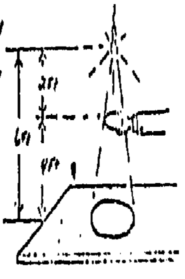
REASON

The item appears to be appropriate.

1F₂

A ring 3 inches across is 2 feet from the light and 4 feet from the table. The 3" ring has a 9" shadow. Where should a 4" ring be placed to make the same size shadow?

All	1110	1111		
8	3	0	A. The shadow will be larger than 9" wherever the ring is placed.	1
21	13	50	(B.) About 3 ft. from the lamp because $\frac{2}{3} = \frac{2.7}{4}$	4
25	40	5	C. About 3 ft. from the lamp because $\frac{2}{3} \times 4 = 2.7$	3
31	30	45	D. About 3 ft. from the lamp because the ring is 1" larger 3 + 1 = 4 and 2ft. + 1 ft. = 3 ft.	2
15	13	0	E. I have no answer	0

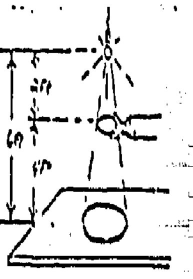


CHANGE

"B" changed with all proportions shown.

A ring 3 inches across is 2 feet from the light and 4 feet from the table. The 3" ring has a 9" shadow. Where should a 4" ring be placed to make the same size shadow?

All	1110	1111		
12	0	0	A. The shadow will be larger than 9" wherever the ring is placed.	1
14	14	33	(B.) About 3 ft. from the lamp because $\frac{6}{9} = \frac{2}{3} = \frac{2.7}{4}$	4
29	43	33	C. About 3 ft. from the lamp because $\frac{2}{3} \times 4 = 2.7$	3
25	21	0	D. About 3 ft. from the lamp because the ring is 1" larger 3 + 1 = 4 and 2ft. + 1 ft. = 3 ft.	2
19	21	33	E. I have no answer	0



REASON

This item is more difficult than desired possibly because a student sees the 6/9 proportion and no place to apply it.

LEVEL V VERSION 4 (LEVEL 3 ITEMS FOR CONSIDERATION)

16F₁

Here is a listing of some metric and some English measures: 4 inches = 10.2 cm. 12 inches = 30.6 cm. ? inches = 100 cm.

A11	1100	1110	1111		
12	17	7	0	A. About 40 inches because it seems that much	1
43	39	64	100	[✓ B.] About 39 inches because $\frac{4\text{ inches}}{10.2\text{ cm.}} \times 100 = 39.2$	4
18	17	14	0	C. About 50 inches because it has to be more	2
17	11	14	0	D. About 80 inches because 30 + 70 = 100 cm. and 12 inches + 70 inches = 32 inches	3
10	17	0	0	E. I have no answer	0

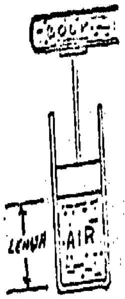
COMMENTS

12F₁

Books balanced on top of this air spring compress the spring. For 2 books the spring is 10 cm long. For 5 books it is 4 cm. long. Predict what length it will be for 8 books?

2 books	10 cm.
5 books	4 cm.
8 books	? cm.

A11	1100	1110	1111		
25	33	21	33	A. About zero (0) because it went down 6 cm. (10 cm. - 4 cm.) for 3 extra books. 3 more books then (5 + 3 = 8) should try to make it go down 6 more.	3
25	28	50	0	[✓ B.] About 3 cm. because since $\frac{2}{5} \times 10$ cm. = 4 cm. then $\frac{2}{8} \times 10$ cm. = 2.5 cm.	4
27	11	14	0	C. About 1 cm. because 5 books - 2 books = 3 books 4 cm. - 3 cm. = 1 cm.	2
13	11	14	0	D. About 2 because it seems that way	1
10	17	0	67	E. I have no answer	0



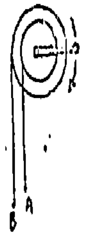
COMMENTS

15F₁

A kind of pulley system here is designed so that turning the crank winds up ends A and B This chart shows how each string moves. How far will B move when A moves 25 cm.

Distance A	Moved B
10 cm.	15 cm.
18 cm.	27 cm.
25 cm.	

A11	1100	1110	1111		
14	17	7	0	A. 36 cm. because it goes up	2
16	17	0	0	B. Less than 42 cm. because 15+27=42	1
[illegible]	22	29	0	C. 35 cm. because 18+7=25 and 27+7=34 and its a little more	3
40	39	64	100	[✓ D.] About 37 because $\frac{15}{10} \times 25 = 37\frac{1}{2}$	4
10	6	0	0	E. I have no answer	0

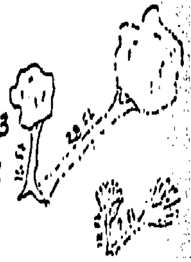


COMMENTS

6F₁

The large 16 foot tree pictured has a shadow 28 feet long. How long a shadow might be cast be the smaller, 12 foot tree?

A11	1100	1110	1111		
4	0	0	0	A. About 20 feet because it seems that way	1
39	50	57	100	[✓ B.] About 21 feet because $\frac{28}{16} \times 12 = 21$	4
29	33	21	0	C. About 24 feet because 16 + 12 = 28 and 12 + 12 = 24	3
16	6	14	0	D. About 24 feet because 16 - 12 = 4 and 28 - 4 = 24	2
13	11	7	0	E. I have no answer	0



COMMENTS

14F₁

John, Mary and Tom each buy a bag of candy - John's bag has 5 mints & 3 gumdrops
Mary's bag has 8 mints & 6 gumdrops
Tom's bag has 4 mints & 3 gumdrops

Which of the persons has the best chance of getting a mint when taking a piece of candy from the bag?

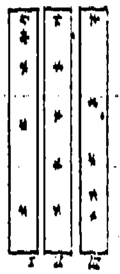
A11	1100	1110	1111		
18	28	14	0	A. Mary because she has the most mints	2
17	22	14	0	B. Mary or Tom because they have 3 more mints than gumdrops	3
13	17	0	0	C. Tom because he has the fewest gumdrops	1
33	22	36	33	[✓ D.] Tom because $\frac{5}{8}$ is more than $\frac{8}{14}$ or $\frac{4}{7}$	4
17	11	29	67	E. I have no answer	0

COMMENTS

3F₁

A "flashing light" rolls down a hill. The flashes at one second apart will most likely sh which of these patterns?

A11	1100	1110	1111			
23	28	43	67	[✓ A.] I	Because each second it goes faster	4
38	28	29	0	B. II	Because it travels each second	2
26	44	0	0	C. I or III	Because it's speed is changing	3
9	0	21	0	D. I, II, or III	Because it is moving	1
4	0	7	33	E. I have no answer		0



COMMENTS

LEVEL I

VERSION 4 VERSION 5 VERSION 6

14C₁

VERSION IV				VERSION V				VERSION VI			
8th				12th				8th			
All	0000	1000	1100	All	0000	1000	1100	All	0000	1000	1100
				0	NOT APPLICABLE				0	0	0
7	15	39		4				11	25	10	3
10	24	6		0				8	16	6	2
15	24	0		3				10	20	13	3
63	24	94		91				68	29	69	90
0	0	0		1				3	9	3	2

Mary buys 3 tickets to a raffle where 90 tickets are sold --- Jane buys 1 ticket to a raffle where 30 tickets are sold --- Sue buys 3 tickets to a raffle where 300 tickets are sold.

Which girls have about the same chance of winning?

No response

A. Jane and Mary because their's are the least tickets 2

B. Sue and Mary because each have 3 tickets 3

C. All girls have the same chance 1

[✓ D.] Jane and Mary because 3 chances in 90 is the same as 1 in 30 4

E. I have no answer 0

12th r_{bis} = .6177 8th r_{bis} = 0.5980

Masters T = 12,9065 T = 15.3799

CHANGE

None

REASON

The item appears to have appropriate discrimination

11C₁

VERSION IV				VERSION V				VERSION VI			
8th				12th				8th			
All	0000	1000	1100	All	0000	1000	1100	All	0000	1000	1100
0	0	8	0	0	NOT APPLICABLE			0	0	0	0
71	29	69	100	94				71	29	82	82
5	12	15	0	4				9	26	4	6
10	29	0	0	1				8	22	1	2
5	12	0	0	0				9	13	10	6
6	18	7	0	0				4	9	3	3

A car moving at a constant speed of 30 mph will, if pictured at one second intervals, look like

No response

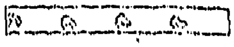
[✓ A.] I because it moves equal distances each second 4

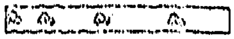
B. None of these because it is moving 1

C. II because it changes 2

D. III because it is increasing its distance 3

E. I have no answer 0

I 

III 

12th r_{bis} = 2.5465 8th r_{bis} = 0.5577

Masters T = 10.7241 T = 13.8522

CHANGE

None

REASON

The item discriminates well and has appropriate difficulty.

9C₁

VERSION IV				VERSION V			
8th				12th			
11	0000	1000	1100	All	0000	1000	1100
				0	NOT APPLICABLE		
6	12	0	11	0			
8	13	8	6	0			
69	35	85	72	96			
8	18	8	6	4			
6	6	0	6	0			

12th r_{bis} = .4376

Masters T = 7.9959

Imagine that frosting had been spread out ¼ inch thick on top of a small 6" x 6" cake. Predict what the thickness would be if the same amount of frosting were spread out over a 12" x 12" cake?

No response

A. More than ¼ inch because it covers less cake 1

B. Less than ¼ inch because it looks that way. 3

[✓ C.] Less than ¼ inch because it covers more cake 4

D. More than ¼ inch because there is more cake 2

E. I have no answer 0

CHANGE

Illustration added

VERSION VI

8th

All	0000	1000	1100
1	1	0	2
5	9	6	3
6	12	4	2
72	48	80	84
12	20	8	8
4	9	1	2

8th r_{bis} = 0.5677

T = 14.2150

Imagine that frosting has been spread out ¼ inch thick on top of a small 6" x 6" cake. Predict what the thickness would be if the same amount of frosting were spread out over a 12" x 12" cake?

No response

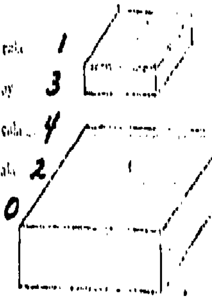
A. More than ¼ inch because it covers less cake 1

B. Less than ¼ inch because it looks that way 3

[✓ C.] Less than ¼ inch because it covers more cake 4

D. More than ¼ inch because there is more cake 2

E. I have no answer 0



REASON

Success on this problem for the C₁ level student should be possible without abstractly viewing what the area change demands.

LEVEL I

4C$_1$

VERSION IV				VERSION V				VERSION VI			
8th				12th				8th			
All	0000	1000	1100	All	0000	1000	1100	All	0000	1000	1100
				0	NOT APPLICABLE			6	15	3	1
4	12	0	6	3				6	10	3	5
14	35	0	0	3				13	21	6	3
6	18	0	6	3				8	19	6	3
62	6	92	72	91				59	25	80	74
13	0	8	17	0				8	9	3	11

The "O" rod here crosses 8 lines. The "Y" rod crosses 5 lines. The "O" rod, when turned, crosses 6 lines. How many lines would the "Y" rod cross if it were at this angle?

No response

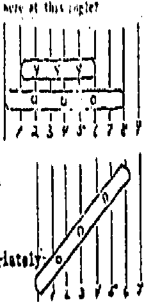
A. About 8 because it should get longer 3

B. About 5 because the "Y" rod is that long 1

C. About 6 because the "O" rod was 6 2

[✓ D.] About 4 because the "Y" rod is shorter 4

E. I have no answer 0



CHANGES

12th Masters r_{bis} = .5975 T = 12.2423

8th r_{bis} = .5794 T = 14.6567

None

REASON

The item appears to discriminate appropriately.

2C$_1$

VERSION IV				VERSION V			
8th				12th			
All	0000	1000	1100	All	0000	1000	1100
			0	0	NOT APPLICABLE		
66	41	69	89	93			
10	33	0	6	1			
5	18	0	0	3			
4	6	0	0	0			
14	0	31	6	3			

A student's desk measures about three textbook lengths or 5 pencil lengths wide. If a teacher's desk is 4 textbook lengths wide, how wide is a teacher's desk measured in pencil lengths?

No response

[✓ A.] More than 5 pencils because it is bigger than a student desk 4

B. Less than 5 pencils because it seems that way 1

C. About 4 pencils because it was 4 textbooks 2

D. 5 pencils because that is what the student desk measured 3

E. I have no answer 0

CHANGES

12th Masters r_{bis} = .5762 T = 11.5831

Added matrix with integer values

VERSION IV			
8th			
All	0000	1000	1100
3	3	10	0
64	41	65	73
5	11	6	2
6	9	3	8
7	17	3	2
15	18	14	16

A student's desk measures about three textbook lengths or 5 pencil lengths wide. If a teacher's desk is 4 textbook lengths wide, how wide is a teacher's desk measured in pencil lengths?

No response

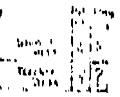
[✓ A.] More than 5 pencils [illegible] 4

B. Less than 5 pencils because it [illegible] 1

C. About 4 pencils because it was 4 textbooks 2

D. 5 pencils because that is what the student desk [illegible] 3

E. I have no answer 0



REASON

8th r_{bis} = .4407 T = 10.1212

Wished to more broadly suggest the proportion answer.

1C$_1$

VERSION IV				VERSION V			
8th				12th			
All	0000	1000	1100	All	0000	1000	1100
			0	0	NOT APPLICABLE		
6	5	8	6	3			
14	33	9	11	6			
19	37	15	17	0			
55	6	69	61	91			
5	19	0	6	0			

A ring is held between a table and a light bulb. The light bulb casts a shadow of the ring. If a smaller ring is held in the same place the shadow of the smaller ring would

No response

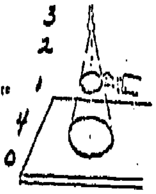
A. Be smaller because the light would change 3

B. Be larger because it is different 2

C. Be the same size because the ring is in the same place 1

[✓ D.] Be smaller because the ring is smaller 4

E. I have no answer 0



12th Masters r_{bis} = .6680 T = 12.7484

CHANGES

Added the "shadow" of a smaller ring.

VERSION IV			
8th			
All	0000	1000	1100
	0		
10	16	7	3
16	33	8	3
11	22	10	5
57	17	70	81
6	11	4	8

A ring is held between a table and a light bulb. The light bulb casts a shadow of the ring. If a smaller ring is held in the same place the shadow of the smaller ring would

No response

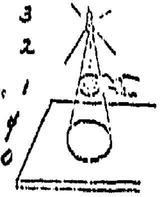
A. Be smaller because the light would change 3

B. Be larger because it is different 2

C. Be the same size because the ring is in the same place 1

[✓ D.] Be smaller because the ring is smaller 4

E. I have no answer 0



REASON

8th r_{bis} = 0.6468 T = 17.5747

This is more difficult than other items for the level. Wished to give a model for the suggested change. Success of ring size at the C$_1$ level should not demand that the student abstract what the change would look like.

LEVEL II

$14C_2$

	VERSION IV				VERSION V				VERSION VI			
	8th				12th				8th			
	All	1000	1100	1110	All	1000	1100	1110	All	1000	1100	1110
No response					0	NOT APPLICABLE		0	2	6	0	0
A.	77	62	94	93	93			93	67	54	92	94
B.	8	8	0	0	4			7	10	11	2	1
C.	13	23	66	7	1			0	14	15	6	3
D.	3	8	0	0	1			0	4	6	0	1
E.	0	0	0	0	0			0	3	8	0	0

These nature hunt groups are chosen for a nature hike.

Mrs. Andrews	- 5 students
Mr. Denton & Mrs. Felk	- 8 students
Mr. Holt	- 6 students

The teacher with the most students to help is:

No response 4

[✓ A.] Mr. Holt because $\frac{6}{1}$ is longer than $\frac{5}{1}$ is larger than $\frac{8}{2}$. 2

B. Mr. Denton & Mrs. Felk because $\frac{2}{8}$ is larger than $\frac{1}{5}$ is larger than $\frac{1}{6}$ 5

C. Mr. Denton & Mrs. Felk because they have the most students 1

D. Mrs. Andrews because she has fewer students 0

E. I have no answer

12th Masters r_{bis} = .5633 T = 11.2036

8th r_{bis} = .5606 T = 13.9561

CHANGES

None

REASON

The item matches well the appropriate difficulty for this level and discriminates well.

$11C_2$

	VERSION IV				VERSION V				VERSION VI			
	8th				12th				8th			
	All	1000	1100	1110	All	1000	1100	1110	All	1000	1100	1110
No response					0	NOT APPLICABLE		0	1	6	0	0
A.	51	15	67	93	86			86	55	31	81	73
B.	13	31	0	0	3			0	9	[illegible]	5	1
C.	10	23	0	0	4			7	11	15	2	3
D.	14	8	22	0	7			7	17	27	11	21
E.	12	23	11	7	0			0	6	11	2	1

Four cars have different speeds: Car A is the fastest, Car B the next fastest, Car C the next fastest, and, Car D the next fastest. The fastest car takes the least time to go 200 miles, the next fastest car the next least time and so on. Which car is the third fastest and takes the third least time to go 200 miles?

No response

[✓ A.] Car C because:	1st fastest CAR A 1st least time	2nd fastest CAR B 2nd least time	3rd fastest CAR C 3rd least time	4
B. Car B				1
C. No car because they don't match up				2
D. Car C because:	1st most fast CAR A 1st most time	2nd most fast CAR B 2nd most time	3rd most fast CAR C 3rd most time	3
E. I have no answer				0

12th Masters r_{bis} = .5895 T = 11.9909

8th r_{bis} = 0.5451 T = 13.4045

CHANGES

None

REASON

The item has excellent discrimination and appropriate difficulty.

$10C_2$

	VERSION IV				VERSION V				VERSION VI			
	8th				12th				8th			
	All	1000	1100	1110	All	1000	1100	1110	All	1000	1100	1110
No response					0	NOT APPLICABLE		0	0	0	0	0
A.	5	8	0	7	0			0	2	6	0	0
B.	16	38	22	0	7			7	13	25	3	1
C.	9	15	0	0	1			0	10	20	2	0
D.	68	31	78	93	91			93	60	41	94	97
E.	3	8	0	0	0			0	6	8	2	1

Here are some recipes for Kool Aide

	1 quart	4 quarts	5 quarts
Kool Aide Powder	½ pkg	2 pkg	?
Sugar	½ c	1 c	
Water	1 qt	4 qts	

How much powder is needed for 5 quarts of Kool Aide

No response

A. ½ pkg because it is the same mixture 1

B. 3 pkg because 4 qts + 1 qt = 5 qts and 2 pkg + 1 pkg = 3pkg 2

C. About 3 because it would have to be more 3

[✓ D.] 2½ pkg because 4 qts + 1 qt = 5 qts and 2 pkg + ½ pkg = 2½ pkg 4

E. I have no answer 0

12th Masters r_{bis} = .5801 T = 11.7011

8th r_{bis} = 0.6045 T = 15.6438

CHANGES

None

REASON

The item has good discrimination and a good difficulty level.

LEVEL II

5C2

VERSION IV				VERSION V			
8th				12th			
All	1000	1100	1110	All	1000	1100	1110
				0	NOT APPLICABLE		0
9	8	11	0	6			10
68	69	76	93	81			83
8	8	6	0	1			0
5	8	0	0	3			5
9	6	6	7	9			3

12th Masters r_{bis} = .4913 T = 9.2681

CHANGES

Ratios made less apparently proportional.

Trial I 7 people on side "A" balance 4 of the same size people on side "B"
Trial II 4 people on side "A" balance 8 of the same size people on side "B"
Trial III 6 people on side "A" should balance how many on side "B"?

No response

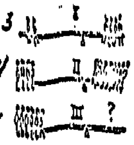
A. About 10 because 2 more on "A" should balance two more on "B" 3

(B.) About 12 because it goes up 4 and 8 + 4 = 12. 4

C. About 10 because it takes 2 more and 8 + 2 = 10 2

D. About 11 because it should be more 1

E. I have no answer 0



VERSION VI			
8th			
All	1000	1100	1110
4	3	2	1
14	23	16	9
35	8	55	63
20	34	10	12
16	23	15	7
11	10	3	7

8th r_{bis} = .4909 T = 11.6166

Trial I 2 people on side "A" [illegible]
Trial II 4 people on side "A" [illegible]
Trial III 6 people on [illegible]

No response

A. About 10 [illegible] on "B" 3

(B.) About 12 [illegible] 12 4

C. About 10 [illegible] 10 2

D. About 11 because it should [illegible] 1

E. I have no answer 0

REASON

The item did not discriminate well between Level I and Level II.

3C2

VERSION IV				VERSION V				VERSION VI			
8th				12th				8th			
All	1000	1100	1110	All	1000	1100	1110	All	1000	1100	1110
				0	NOT APPLICABLE		0	2	6	2	0
6	15	6	6	3			0	7	13	2	1
17	23	6	21	1			0	22	25	16	12
13	15	0	7	10			3	15	20	8	4
60	38	89	69	86			97	50	30	69	81
4	8	0	7	0			0	3	7	3	1

CHANGES

None

12th Masters r_{bis} = .3699 T = 6.5416

8th r_{bis} = .5037 T = 11.7257

This person sliding down a hill looks at her watch Each second she puts a stick in the snow. What most likely would be the pattern of these sticks?

No reponse

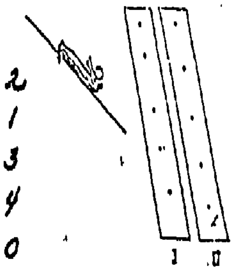
A. I because she moves each second 2

B. II because it is a steep hill 1

C. I or II because she is moving 3

(D.) I because her speed is changing 4

E. I have no answer 0



REASON

The item work appropriately for 8th graders. It lacks discrimination as expected for masters.

1C2

VERSION IV				VERSION V				VERSION VI			
8th				12th				8th			
All	1000	1100	1110	All	1000	1100	1110	All	1000	1100	1110
				0	NOT APPLICABLE		0	1	3	0	0
21	15	26	0	7			7	25	30	19	13
69	67	74	97	84			93	53	41	65	79
2	0	0	0	1			0	5	3	3	0
7	13	0	3	6			0	13	20	6	6
2	6	0	0	1			0	3	4	6	1

12th Masters r_{bis} = .5208 T = 10.0244

8th r_{bis} = .4944 T = 11.7257

CHANGES

None

A ring is held between a table and a light bulb. The light casts a shadow of the ring onto the table. If the ring is moved closer to the table, the shadow may:

No response

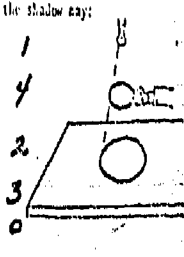
A. Become larger because the shadow spreads out 1

(B.) Become smaller because the light rays don't spread as much 4

C. Stay the same because it's the same ring 2

D. Become larger because the bulb is father away 3

E. I have no answer 0



REASON

The item has good discrimination although it is harder than many in the set.

LEVEL III

18F1

	VERSION IV 8th				VERSION V 12th				VERSION VI 8th				
	All	1100	1110	1111	All	1100	1110	1111	All	1100	1110	1111	
No response					0	NOT APPLICABLE	0	0	0	0	0	0	
A.	16	44	7	0	4		3	4	19	31	4	4	2
B.	58	50	93	100	84		86	89	46	26	81	91	4
C.	12	6	0	0	1		0	4	14	16	9	0	1
D.	6	0	0	0	6		7	0	11	13	1	4	3
E.	6	0	0	0	4		3	4	11	15	4	0	0

A movie projector lens spreads its light out over a 3' x 3' screen 9 feet away. To make the image spread over a 5' x 5' screen, how far back must the screen be moved?

No response

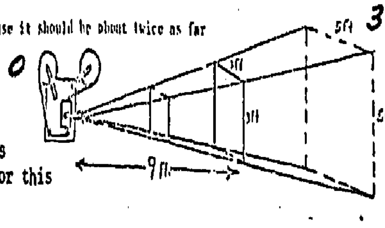
A. About 11 feet. The 5 foot image is 2 more than the 3 foot one and 11 feet is 2 more than 9 feet 2

(B.) About 15 feet because 3/9 = 5/15 4

C. About 12 feet because 9 + 3 = 12 1

D. About 18 feet because it should be about twice as far 3

E. I have no answer 0



CHANGES

None

12th Masters r_{bis} = .6116 T=12.7015

8th r_{bis} = .5863 T = 14.9215

REASON

The item works appropriately for this level.

17F1

	VERSION IV 8th				VERSION V 12th				VERSION VI 8th				
	All	1100	1110	1111	All	1100	1110	1111	All	1100	1110	1111	
No response					0	NOT APPLICABLE	0	0	1	2	0	0	
A.	9	5	0	0	0		0	0	9	2	3	0	2
B.	23	44	14	0	3		0	0	15	18	6	4	3
C.	34	28	71	67	70		83	64	34	16	61	65	4
D.	27	22	14	33	25		10	36	30	52	25	26	1
E.	4	0	0	0	3		7	0	11	11	4	4	0

Jane is weighing out apples on this supermarket scale. What will 14 apples weigh if 6 apples weigh 1½ lbs?

No response

A. 9½ lbs because 6 + 8 = 14 so 1½ + 8 = 9½ 2

B. 3 or 4 lbs because it is more 3

(C.) 3½ lbs because $\frac{1\frac{1}{2}}{6}$ x 14 = 3½ 4

D. 3½ because 1½ + 1½ + ½ = 3½ 1

E. I have no answer 0

CHANGES

None

12th Masters r_{bis} = .5240 T = 10.1080

8th r_{bis} = .4609 T = 10.7061

REASON

The problem although difficult does discriminate.

11F1

	VERSION IV 8th				VERSION V 12th				
	All	1100	1110	1111	All	1100	1110	1111	
No response					0	NOT APPLICABLE	0	0	
A.	13	42	0	5	3		0	4	2
B.	4	0	0	0	4		0	0	1
C.	65	52	90	97	87		100	90	4
D.	7	3	3	0	3		0	4	3
E.	5	3	7	0	3		0	4	0

A car moving at a constant 45 mph travels 198 ft. in 3 seconds. How far will it have traveled by the end of 5 seconds?

No response

A. More than 198 feet because it is still moving 2

B. Less than 400 feet because it is only 2 seconds more 1

(C.) 330 feet because $\frac{198}{3}$ x 5 = 330 4

D. 200 feet because 3 sec. + 2 sec. = 5 sec. 198 ft. + 2 ft. = 200 ft. 3

E. I have no answer 0

CHANGES

12th Masters r_{bis} = .5812 T = 11.7351

Numbers in the problem were changed.

	VERSION VI 8th				
	All	1100	1110	1111	
No response	0	0	0	0	
A.	19	15	10	9	2
B.	7	5	1	0	1
C.	55	53	79	83	4
D.	7	3	4	4	3
E.	12	24	6	4	0

A car moving at a constant 30 mph travels 88 ft. in 2 seconds. How far will it have traveled by the end of 5 seconds?

No response

A. About 264 feet because 3 x 88 = 264. 2

B. About [illegible] feet because it is only 3 seconds more 1

(C.) 220 feet because [illegible] = 220 4

D. 91 feet because [illegible] and 88 ft. + 3 ft. = 91 ft. 3

E. I have no answer 0

8th r_{bis} = .5346 T = 13.0426

REASON

The sampler integers were intended to be more readily identified as proportional or additive.

LEVEL III

10F$_1$

VERSION IV 8th — VERSION V 12th

All	1100	1110	1111	All	1100	1110	1111
				0	NOT APPLICABLE	0	0
49	22	93	67	84		79	96
23	33	0	33	3		3	0
4	0	0	0	0		0	0
1	6	0	0	3		0	4

Jim uses 2 heaping teaspoons of Tang powder with an 8 oz. glass of water. How much Tang is needed for the same mixture with 27 oz. of water?

No response

(A.) About 7 teaspoons because $\frac{27}{8} \times 2$ tsp. = 6 3/4 tsp 4

B. About 21 teaspoons because 27 oz -8 oz $\overline{19 \text{ oz}}$ and 2 tsp. + 19 tsp. = 21 tsp. 3

C. More than 2 teaspoons because there is more water 2

D. 2 teaspoons because it is the same mixture 1

E. I have no answer 0

CHANGES

1. Simplification of number ratios.
2. Distractor "B" changed to an addition type.

12th r_{bis} = ..5926
Masters T = 12.0891

VERSION VI

8th

All	1100	1110	1111
1	2	0	0
57	58	93	87
17	13	3	9
14	21	1	4
6	3	0	0
5	3	3	0

Jim uses 4 heaping [illegible] glass of water. How much [illegible] is needed for the same mixture [illegible]

No response

(A.) [illegible] 4

B. About 6 teaspoons because 8 oz. + 4 oz. = 12 oz. and 4 tsp. + 4 tsp. = 8 tsp. 3

C. More than 4 teaspoons because there is more water 2

D. 4 teaspoons because it is the same mixture 1

E. I have no answer 0

8th r_{bis} = .5793
T = 14.6514

REASON

1. The item overall is too difficult.
2. This is a more appropriate distractor for Level II.

8F$_1$

VERSION IV 8th — VERSION V 12th — VERSION VI 8th

All	1100	1110	1111	All	1100	1110	1111	All	1100	1110	1111
				0	NOT APPLICABLE	0	0	1	0	0	0
37	33	57	33	62		79	57	39	27	67	83
5	0	7	0	3		3	0	12	8	3	0
14	11	7	0	0		0	0	11	10	1	0
26	39	29	67	30		17	36	18	21	13	13
16	17	0	0	4		0	7	18	34	15	4

A model airplane wing made from the 2 cm. pattern shown measures 19 cm. long. What would be the length of such a wing made from a pattern with squares that are 6 cm.?

No response

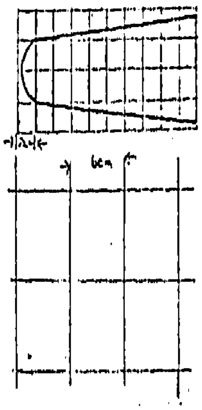
(A.) 57 cm, because 6/2 x 19 = 57 4

B. 18 cm, because it looks that way 3

C. 22 cm, because 19 + 3 = 22 2

D. 19 cm, but the squares would be larger 1

E. I have no answer 0



CHANGES

None

12th r_{bis} = .0750
Mastors T = 1.2352

8th r_{bis} = .5280
T = 12.8176

REASON

The item seems sound - wish to have a larger group tested with it.

2F$_1$

VERSION IV 8th — VERSION V 12th

All	1100	1110	1111	All	1100	1110	1111
				0	NOT APPLICABLE	0	0
13	6	6	0	3		0	4
17	23	3	0	0		0	0
43	16	57	100	90		97	93
22	53	3	0	3		3	0
5	6	0	0	4		0	4

Here is sketch #1 of an airplane. Sketch #1 is 7 pencil widths or 3 pennies high. Sketch #2 of this airplane is not shown. Sketch #2 looks the same but is 12 pencil widths high. How high must sketch #2 be in pennies?

No response

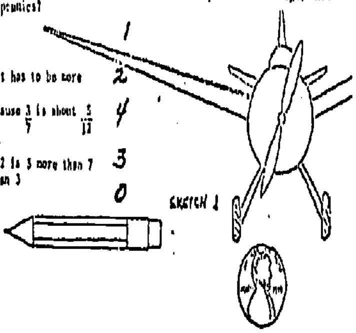
A. Seems to be 6 1

B. About 7 because it has to be more 2

(C.) About 5 because $\frac{3}{7}$ is about $\frac{5}{12}$ 4

D. About 8 because 12 is 5 more than 7 and 8 is 5 more than 3 3

E. I have no answer 0



CHANGES

12th r_{bis} = 0.5132
Masters T = 9.8243

Replace the problem with one that is less abstract.

VERSION VI

8th

All	1100	1110	1111
0	0	0	0
3	2	0	0
11	8	3	0
60	60	91	100
17	16	3	0
9	15	3	0

A lunchroom is 60 ceiling tile or 35 chairs wide. If a classroom is 12 chairs wide, how wide is this classroom measured in ceiling tiles?

No response

A. Seems to be 50. 1

B. About 40 because it has to be less. 2

(C.) About 20 because $\frac{60}{35}$ is about $\frac{20}{12}$. 4

D. About 47 because 60 is 25 more than 35 and 47 is 25 more than 12. 3

E. I have no answer. 0

8th r_{bis} = .5739
T = 14.4468

REASON

The item has some good characteristics but may be having the student pull together too many things.

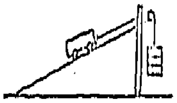
LEVEL IV

19F$_2$

	VERSION IV 8th			VERSION V 12th			VERSION VI (IN MASTERS) 8th				
	All	1110	1111	All	1110	1111	All	1110	1111		
No response				0	0	0	0	1	0		
A. 133 because $\frac{100}{300} \times 400 = 133$	22	29	0	29	41	18	16	19	4	3
B. 150 because it is more	9	0	0	6	14	0	12	6	4	1
C. 177 because it goes up 17 for every 100	25	29	0	4	7	0	17	12	0	2
[✓ D.] 133 because $\frac{100}{300} = \frac{133}{400}$	21	36	100	52	24	75	32	36	91	4
E. I have no answer	22	7	0	9	14	7	22	25	0	0

On the ramp illustrated the cart and its weight is balanced by weights on the string. What amount of weight is needed to balance 400 g of cart weight at 20° ?

	Weight	
Angle	Cart	String
10°	200g	35
10°	300g	52
20°	300g	100
20°	400g	?



CHANGES 12th ^{r}bis = 0.5095 8th ^{r}bis = .5591 REASON

Masters T = 9.7293 T = 13.9015

None

The item appeared to be operating appropriately.

15F$_2$

	VERSION IV 8th			VERSION V 12th			VERSION VI 8th				
	All	1110	1111	All	1110	1111	All	1110	1111		
No response				0	0	0	0	1	0		
[✓ A.] About 13 minutes because $\frac{4 \text{ min.}}{3 \text{ miles}} = \frac{10 \text{ min.}}{7.5 \text{ miles}} = \frac{13\ 1/3 \text{ min.}}{10 \text{ miles}}$	18	17	30	74	66	86	37	39	91	4	3
B. About 13 minutes because 10 - 7½ = 2½ miles and 10 + 2½ = 12½ min.	13	3	5	1	3	0	16	13	0	2	1
C. About 13 minutes because $\frac{4}{3} \times 10 = 13\ 1/3$	25	20	0	22	28	14	16	25	0	3	2
D. About 14 because 7½ + 3 = 10½ and 10 + 4 = 14	38	50	65	3	3	0	19	16	9	1	[illegible]
E. I have no answer	5	10	0	0	0	0	12	4	0	0	[illegible]

A freeway driver keeps track of the distance he travels. He finds that in 4 minutes he travels 3 miles/ in 10 minutes 7½ miles. If he continues at this speed, how long will it take him to travel 10 miles?

Distance	Time
3 miles	4 min
7½ miles	10 min
10 miles	? min

CHANGES 12th ^{r}bis = 0.5429 8th ^{r}bis = .5148 REASON

Masters T = 10.0235 T = 12.3790

None

The item discriminates well. ^{r}bis is excellent.

10F$_2$

Here is a recipe for 4 cups of cocoa: Heat to near boiling 4 c. milk
Add with stirring 6 T. sugar
5 T. Cocoa

How many tablespoons of sugar would be needed to make 12 cups of this cocoa?

	VERSION IV 8th			VERSION V 12th			VERSION VI 8th			
	All	1110	1111	All	1110	1111	All	1110	1111	
No response				0	0	0	1	0	0	
A. 18 tablespoons because $\frac{6}{4} \times 12 = 18$	26	50	0	36	69	7	24	28	9	3
B. More than 6 tablespoons because there is more cocoa	12	0	0	3	0	0	10	1	0	1
[✓ C.] 18 tablespoons because $\frac{6}{4}$ equals $\frac{18}{12}$.	38	36	100	57	28	89	45	63	83	4
D. 14 tablespoons because 4 c. + 8 c. = 12 c. so 6 T. + 8 T. = 14 T.	16	14	0	1	0	0	12	3	4	2
E. I have no answer	8	0	0	3	3	4	8	4	4	0

CHANGES 12th ^{r}bis = .5230 8th ^{r}bis = .5053 REASON

Masters T = 10.0820 T = 12.0709

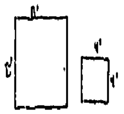
None

The item works well. ^{r}bis is appropriate.

LEVEL IV

9G$_2$

VERSION IV			VERSION V			VERSION VI				
8th			12th			8th				
All	1110	1111	All	1110	1111	All	1110	1111	Imagine that concrete has been mixed to make a patio 4 ft. x 4 ft. and ½ a foot thick. How thick will this concrete be if it is instead spread out over an 8 ft. x 8 ft. area?	
			0	0	0	3	0	0	No response	
19	7	67	74	69	86	16	13	57	[✓ A.] $\frac{1}{8}$ ft. thick because $\frac{16}{64} = \frac{\frac{1}{8}}{\frac{1}{2}}$	4
38	71	33	23	28	14	39	43	35	B. $\frac{1}{4}$ ft. thick because $\frac{4}{8} = \frac{\frac{1}{4}}{\frac{1}{2}}$	3
19	14	0	0	0	0	23	36	4	C. $\frac{1}{4}$ ft. thick because $\frac{1}{4}$ is less than $\frac{1}{2}$	2
8	0	0	[illegible]	0	0	14	4	0	D. $\frac{1}{4}$ ft. thick because it should be l[illegible]	1
12	0	0	[illegible]	[illegible]	0	5	3	4	E. I have no answer	0



CHANGES

None

12th Masters r_{bis} = 0.3807 T = 6.7655

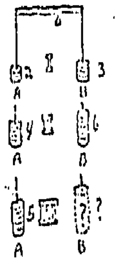
8th r_{bis} = 0.3876 T = 8.6678

REASON

The item seemed to discriminate but have high difficulty. I wished to see how it would work with the 12th grade masters.

5F$_2$

VERSION IV			VERSION V			VERSION V1				
8th			12th			8th				
All	1110	1111	All	1110	1111	All	1110	1111	Trial I - Two weights on side "A" balance three of the same weights on side "B" Trial II - Four weights on side "A". Six on side "B" Trial III- Five weights on side "A" then should balance how many weights on side "B"?	
			0	0	0	2	1	0	No response	
17	13	0	32	55	7	19	34	26	A. About 8 because $\frac{6}{4} \times 5 = 7.5$	3
34	23	95	54	21	86	25	22	57	[✓ B.] About 8 because $\frac{6}{4} = \frac{7.5}{5}$	4
26	13	0	3	3	4	22	19	13	C. About 7 because $4 + 1 = 5$ & $6 + 1 = 7$	1
15	47	5	12	21	4	13	9	4	D. About 7 because $\frac{6}{4}$ is less than $\frac{8}{5}$	2
6	3	0	0	0	0	18	13	0	E. I have no answer	0



CHANGES

None

12th Masters r_{bis} = .4005 T = 7.1822

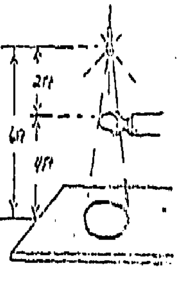
8th r_{bis} = .4066 T = 9.1762

REASON

The item appeared to be working appropriately.

IF$_2$

VERSION IV			VERSION V			VERSION VI				
8th			12th			8th				
All	1110	1111	All	1110	1111	All	1110	1111	A ring 3 inches across is 2 feet from the light and 4 feet from the table. The 3" ring has a 9" shadow. Where should a 4" ring be placed to make the same size shadow?	
12	0	0	3	3	4	0	0	0	No response	
14	14	33	4	0	4	10	6	0	A. The shadow will be larger than 9" wherever the ring. is placed.	1
29	43	33	28	7	36	26	16	70	[✓ B.] About 3 ft. from the lamp because $\frac{6}{9} = \frac{2}{3} = \frac{2.7}{4}$	4
25	21	0	35	52	29	24	28	30	C. About 3 ft. from the lamp because $\frac{2}{3} \times 4 = 2.7$	3
19	21	33	12	28	0	21	31	0	D. About 3 ft. from the lamp because the ring is 1" larger 3 + 1 = 4 and 2ft. + 1 ft. = 3 ft.	2
			19	10	29	19	18	0	E. I have no answer	0



CHANGES

None

12th Masters r_{bis} = .2896 T = 4.9718

8th r_{bis} = .4764 T = 11.1701

REASON

Wished to test with a larger sample.

LEVEL V

(F_2 - Level IV items)

16F_2

VERSION V			VERSION VI		
12th			8th		
All	1110	1111	All	1110	1111
0	0	0	1	0	0
0	0	0	10	4	0
13	3	21	33	39	22
47	66	43	25	37	30
39	31	36	19	18	43
0	0	0	13	1	4

r_{bis} = .6741 T = 12.7840 r_{bis} = .4336 T = 9.9206

Here is a listing of some metric and English measures: 2 gal. = 6.5 liters
5 gal. = 21.2 liters
8 gal. = ? liters

What is the value in liters of 8 gal.?

No response

A. About 24 liters because 8 gal. is 3 more than 5 gal. and 24 is about 3 more than 21.2 1

B. About 31 liters because 3 {2 gal. = 6.5 liters / 5 gal. = 21.2 liters} 12.7, 3 {5 gal. = 21.2 liters / 8 gal. = 33.9 liters} 12.7 2.

C. 26 liters because $\frac{6.5}{2}$ x 8 = 26 .3

(D) 34 liters because $\frac{2}{6.5} = \frac{5}{21.2} = \frac{8}{34.0}$ 4

E. I have no answer 0

COMMENT

This item should be considered. It has promise of good discrimination. It is now too difficult. Possible the 2-5-8 gal. could be just a 2-5 comparison and the distractors then simplified.

14F_2

VERSION V			VERSION VI		
12th			8th		
All	1110	1111	All	1110	1111
0	0	0	2	1	0
6	7	15	23	15	4
59	65	33	21	37	35
1	0	7	12	10	0
25	21	23	22	26	43
4	3	4	16	6	17

r_{bis} = .0838 T = 1.3819 r_{bis} = .5593 T = 13.9089

Skip has 2 pair of red socks and 4 pair of blue socks in his sock drawer.
Jim has 3 pair of red socks and 5 pair of blue socks in his sock drawer.
Steve has 1 pair of red socks and 3 pair of blue socks in his sock drawer.

Which boy has the best chance of grabbing a pair of red socks when reaching in the dark into his sock drawer?

No response

A. Jim because he has the most red socks 1

(B) Jim because $\frac{3}{8}$ = .375 $\frac{2}{6}$ = .333 $\frac{1}{4}$ = .250 4

C. Jim because each has 2 more blue than red socks 2

D. Jim because $\frac{3}{8}$ is more than $\frac{2}{6}$ is more than $\frac{1}{4}$ 3

E. I have no answer 0

COMMENT

Masters do not react appropriately to this item. The subtloty between distractor D and B - the key is probably too fine.

12F_2

VERSION V			VERSION VI		
12th			8th		
All	1110	1111	All	1110	1111
0	0	0	1	0	0
16	17	14	15	10	4
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
14	17	11	19	24	17
12	7	14	18	9	17
7	7	4	20	15	9

r_{bis} = [illegible] T = [illegible] r_{bis} = .4877 T = 11.5168

Books on top of this air spring compress the spring. For 2 books the spring is 8 cm. long. For 9 books it is 1.6 cm. long. What should be the spring length for 5 books?

No response

A. About 3 cm. to 4 cm. because it has to be about half between 1.6 cm. and 8 cm. 2

(B) About 3 cm. because if $\frac{2 \text{ books}}{9 \text{ books}} = \frac{1.6 \text{ cm.}}{8.0 \text{ cm.}}$ then $\frac{2 \text{ books}}{5 \text{ books}} = \frac{3.2 \text{ cm.}}{8.0 \text{ cm.}}$ 4

C. About 3 cm. because $\frac{2}{5}$ x 8 = 3.2 3

D. About 5 cm. because 5 books - 2 books = 3 books and 8 cm. - 3 cm. = 5 cm. 1

E. I have no answer 0

COMMENT

This question should be substituted for one of the poorer ones used in level IV.

9F_2

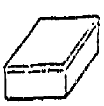
VERSION V			VERSION VI		
12th			8th		
All	1110	1111	All	1110	1111
0	0	0	2	0	0
78	72	89	20	16	57
6	7	0	42	61	39
3	0	0	19	18	0
6	7	7	11	4	4
7	14	4	6	0	0

r_{bis} = .3732 T = 6.6091 r_{bis} = .4761 T = 11.1607

Frosting has been spread out ¼ inch thick on top of a 6" by 6" cake. The same amount of frosting is spread out over a 12" by 12" cake. Predict the new frosting thickness.

No response

(A) $\frac{1}{16}$" because $\frac{36}{144} = \frac{\frac{1}{16}}{\frac{1}{4}}$ 4



B. $\frac{1}{8}$" because $\frac{6}{12} = \frac{\frac{1}{8}}{\frac{1}{4}}$ 3

C. $\frac{1}{8}$" because $\frac{1}{8}$ is less than $\frac{1}{4}$ 2

D. $\frac{1}{8}$" because it should be less 1

E. I have no answer 0

COMMENT

This problem involves inverse as the square variation. It is difficult and probably of an other level.

3F_2

VERSION V			VERSION VI		
12th			8th		
All	1110	1111	All	1110	1111
0	0	0	5	1	0
43	45	46	22	24	43
12	14	11	12	9	0
25	24	18	20	25	26
20	17	25	31	35	22
0	0	0	9	4	9

r_{bis} = .7097 T = 16.5518 r_{bis} = .3936 8.8266

A "stop-action" picture is taken of a diver. The picture shows where the diver was each one tenth second of his dive. Which of the pictures (I,II,III,IV) is the most likely "stop action" picture?

No response

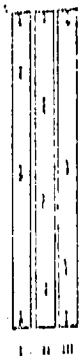
(A) I because each second he goes faster and travels further 4

B. I because each second he goes faster 3

C. I or III because his speed is changing 2

D. II because he travels each second 1

E. I have no answer 0



REASON

This needs some editing. Possibly distractors "D" and "E" should be changed. The item has some possibilities. For "C" the I or III should be on the same line.

2F_2

VERSION V			VERSION VI		
12th			8th		
All	1110	1111	All	1110	1111
0	0	0	1	1	0
1	0	0	17	6	4
[illegible]	[illegible]	7	35	57	[illegible]
41	55	18	23	24	13
3	0	0	13	7	4
3	0	4	12	4	4

r_{bis} = [illegible] T = [illegible] r_{bis} = .5401 T = 13.2289

Sketch #1 of a house is 5 pencil widths or 2 pennies high. Sketch #2 of this house is not shown. Sketch #2 looks the same but is 8 pencil widths high. How high must sketch #2 be in pennies?

No response

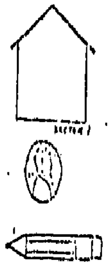
A. About 5 because 8 - 5 = 3 2

(B) About 3 because $\frac{2}{5} = \frac{3.2}{8}$ 4

C. About 3 because $\frac{2}{5}$ x 8 = 3.2 3

D. About 3 because it has to be more 1

E. I have no answer 0



COMMENT

This question should be substituted for one of the poorer ones used in level IV.