AN APPLICATION OF THE HYBRID MODEL TO A STATE COMPETENCY TEST IN MATHEMATICS

by

Laura Alvarez

A dissertation submitted to the Graduate Faculty in Educational Psychology in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

1996

UMI Number: 9618042

UMI
300 North Zeeb Road
Ann Arbor, MI 48103

This manuscript has been read and accepted for the Graduate Faculty in Educational Psychology in satisfaction of the dissertation requirements for the degree of Doctor of Philosophy.

1/17/96
Date

Chair of Examining Committee

1/17/96
Date

Executive Officer

Professor David Rindskopf

Professor Roger Milsap

Professor Burt Flugman

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

AN APPLICATION OF THE HYBRID MODEL TO A STATE COMPETENCY TEST IN MATHEMATICS

By

Laura Alvarez

Adviser: Professor Carol Kehr Tittle

The present study models a procedure to extract instructional information from a minimal competency test in mathematics, using Yamamoto's (1989) Hybrid model (a combination of latent class models (LCM) and an item response theory model (IRT)). This procedure is applied to the data of 1600 ninth grade students from three inner-city schools, who took the June 1991 examination of the New York State Regents Competency Test (RCT) in mathematics. The test is examined in terms of mathematical content and cognitive processes proposed by the National Council of Teachers of Mathematics (NCTM). Two analyses were made: a two parameter with 14 latent classes Hybrid model was compared to a two parameter IRT model in terms of fit. Next, the classes identified by the Hybrid model were compared with the classes identified by a sub-score method, using Cohen's coefficient of agreement to assess if the same students were being identified by each method. Nine teachers with a minimum of 10 years experience evaluated the latent class groupings of items obtained from the Hybrid model to determine instructional relevance. The Hybrid model provided a better fit then the IRT model (χ^2 diff (42) = 770.6); it identified different students from those in the sub-score method (κ = -.220); it identified 18% of the subjects into one of the latent classes; lastly, these classes were judged instructionally relevant by educators (M = 3.90). The Hybrid model provides instructional information not obtainable with traditional measures and offers a wide range of classification possibilities. Applying the Hybrid model to more sophisticated tests, especially in the area of math and science is suggested.

ACKNOWLEDGMENTS

I am grateful to the measurement faculty at the Graduate Center for their support and instruction. I am particularly grateful to Professor Carol Kehr Tittle, a unique professional, whose nurturing, guidance and encouragement has helped me more than she knows.

Special thanks to my family, in particular my mother, for her constant readings, discussions, faith and love. Thank you, Marianne for making it possible and thank you Mitchel for making it worthwhile.

TABLE OF CONTENTS

Appendixes

LIST OF TABLES

INTRODUCTION

Assessment is an important part of the education process. It is used for a variety of purposes: to determine mastery, to determine the need for remediation, to determine course placement, and to evaluate school instructional programs. The Regents Competency testing program was introduced by New York State to certify a student has obtained a minimal level of mastery. If students fail these exams they are placed in a remediation program. The problem of interest here is, when tests are designed for one purpose (to demonstrate minimal competency) can they be used for another (to identify areas of weakness for instructional planning)?

Current Views on Student Learning and Assessment

Although characterizing ability as a single continuous variable is appropriate for prediction and/or selection purposes, it is inadequate for instruction and placement. In fact, it is the belief of educators that a test score should do more than indicate whether a performance is adequate; it should also aid in the determination of what to do when it is not (Burton, 1978). Traditionally, teachers have been provided with reports on a student's standing by total test score or narrowly defined objectives defined by the content covered by a test. These results are typically interpreted as norm-referenced or criterion referenced. This information does not aid in the planning of instruction at the classroom level because in any given classroom, student performances vary not only in the area but also in the type of weakness. To be instructionally useful, a test score must identify cognitive commonalties rather than one general "underlying" ability. It can be argued that even criterion referenced tests are limited in their instructional usefulness because the scores reflect items which are clustered by surface/logical characteristics. A new test theory should move beyond the concept of providing low to high ability information and identify states of competence that are instructionally relevant.

Most standardized tests do not provide information relevant to instructional decisions and teachers need information that contributes to instructional planning (Glaser, 1981). Individualization of instruction is an important concept in education and teachers may benefit from the identification of specific class types that can be used to plan instruction. Currently, the method most often used is grouping.

The grouping of students is not a new idea, but the emphasis here is not on selection by ability level but on the identification of cognitive commonalties that are relevant for instruction. A classification system based on the analysis of the intelligent systematization of student errors can guide instruction for the further improvement of skills (Glaser, 1981), i.e., a student's responses can be used to infer qualitative information about that individual's cognitive processing. Such a system would require new models of assessment and should also reflect current thinking on psychological constructs that are relevant to instruction. Until such systems are developed, information about an individual's performance, information that is not currently available using conventional methods of analysis, can be obtained using newly developed statistical techniques (Gitomer & Yamamoto, 1991). Research can fruitfully examine current assessment tools for their potential use in identifying cognitive commonalties and the meaningfulness of such commonalties to teachers for their instructional planning. At present these statistical models are not used in any significant way in the school system.

Statistical Models

Standard models of educational measurement focus on overall proficiency. However, recent developments in educational measurement introduce models which focus on patterns of correct answers. These statistical models are described by Mislevy, Yamamoto & Anacker (1990). A brief outline of these models follows.

The Partial Credit model (Master and Mislevy, 1989; Wilson, 1989) builds on the one parameter Rasch model. This model is suggested when an item's solution is viewed as a series of steps. An examinee's understanding is inferred from the nature of the approach (completion of steps) rather than from the correctness of the response. The Saltus model (Wilson, 1989) posits stages of conceptual development and extends the Rasch model to include stage membership parameters.

The rule space model (Tatsuoka, 1983) was designed to link rule-assessment methods with the personal index approach in order to diagnose student's misconceptions when learning a task. This model maps all response patterns into a set of ordered pairs which are based on latent ability (Θ) and a parameter (ζ) introduced by Tatsuoka based on IRT caution indices. This model is suggested for application in areas where all the possible response patterns are known (i.e., in narrowly defined areas of mathematics such as solving simple equations).

Latent class models are suggested when competencies are viewed as qualitatively different and non-ordered, and when distinctions within class are unnecessary. A type of latent class model, the Binary Skill model (Falmagne, 1989; Haertel, 1984b) is proposed for use when competence in an area can be described by elements of skill or knowledge. To use this model, observational situations are devised that require specific skill combinations for solution.

Lastly, Yamamoto's (1987) Hybrid model combines latent class and IRT models to optimally describe student responses. This model has a broader application than the partial credit or the rule space model because it can be applied to items where the steps to solution are not clearly defined. Its use is recommended where instructional decisions are based on particular response patterns which are known to be responsive to certain educational activities.

In this study a criterion referenced test that has been in use for the past fifteen years is analyzed. The Hybrid model was chosen for this analysis because it does not place undue restrictions on item construction (i.e. that items must be written to elicit steps to solutions or that all possible response patterns are known). The use of the Hybrid model permits separating out those students for whom a continuous ability model is appropriate, and for whom the total score is the most information that the test can provide. Furthermore, the latent portion of the Hybrid model can identify classes which reflect response patterns that then can be translated into teaching practices.

Psychological Constructs

A framework for conceptualizing psychological constructs relevant to instructional assessment is outlined by Snow (1989). Snow states that improved assessment will include conative as well as cognitive constructs. However, to date, these constructs have been demonstrated singularly in small-scale situations. Research is needed on the validity of these constructs to produce a more "coherent and parsimonious system of distinguishable, understandable, and measurable constructs " (Snow, 1989, p.11). These constructs include conceptual structures of declarative knowledge; procedural skills involved in learning, thinking and reasoning; learning strategies, styles and tactics; self-regulatory functions; and motivational orientation. The first two constructs are considered cognitive and the last two conative while learning strategies are considered a combination of both. These constructs are each referenced to states in instructional development which include: initial states (aptitude constructs), transitions (learning-developmental constructs) and desired states (achievement constructs), creating a 3 by 5 matrix.

This matrix is intended as a "sketch" of those constructs that are sensitive to instructional design and therefore meaningful. Snow stresses the importance of research on teacher use of assessment procedures since "no progress can be made in educational practice related to assessment without it."(1989 p.14)

To understand a student's test performance more fully, psychological constructs as well as content areas should be examined. However, the assessment tool analyzed in this study (the New York State Regents Competency Test (RCT) in mathematics) was not constructed to report psychological constructs. While the nature of Snow's constructs are too broad and too refined to impose on this measure, incorporation of his cognitive constructs and emphasis on teacher use of assessment procedures is possible. A classification system which is more appropriate for this study is the cognitive framework proposed by the National Council of Teachers of Mathematics (NCTM). This system was chosen because its cognitive constructs are applicable to the RCT.

Purpose of the Study

The present study, using newly developed statistical techniques, models a procedure to extract instructionally relevant information from existing tests. The Hybrid model (Yamamoto, 1989), a combination of latent class models (LCM) and an item response theory model (IRT), was chosen because it could identify latent classes, in addition to providing ability classifications. The New York State Regents

Competency Test (RCT) in mathematics, a minimal competency test in mathematics, was analyzed to determine if the latent classes reveal instructionally relevant information.

In the high schools, the RCT is instrumental in several important educational decisions: minimal competency requirements for graduation, and the allocation by the state of compensatory funding for remediation programs. The use of cut scores for graduation assumes a continuous latent ability for which IRT models are useful. Its use in the above areas was not of concern here. However, these tests have also been used to determine instructional activities and course placement. It is this classroom-level use that is being examined, where the application of a latent class model may provide information to improve instructional planning.

The application of an overall proficiency score for planning instruction is limiting. For example, the grouping of students in all-encompassing categories such as "below grade level" or "the sixtieth percentile," fails to distinguish among the areas of weakness and the levels of skill proficiency of students. Even in the case of criterion referenced measures, teachers are "over-whelmed" with what to do with a student who needs remediation in all sub-areas. From a user standpoint, the formation of remedial classes using the Hybrid model, which combines IRT and latent class models, reduces information to a point where it is instructionally relevant. The Hybrid model, unlike traditional sub-score methods, follows a sequential process and can be designed to identify smaller, educationally significant samples. The present study identifies types of students for instructional purposes and examines the validity-related evidence for the groupings obtained from the Hybrid model, using statistical models and expert teacher judgment.

To improve the RCT's use as an instructional tool, the test is examined in terms of current learning and assessment models which include mathematical content and cognitive constructs. In particular, The National Council of Teachers of Mathematics (NCTM) has recommended a framework for assessment which incorporates many of the ideas advanced by Snow. The framework includes content areas, process areas and levels of knowledge. For the purposes of this study the content areas outlined and described in the RCT blueprint (Appendices B and D) and the process areas described by NCTM (Appendix E) are used to characterize the items. The Hybrid model was chosen to analyze the data because it can be applied to in-use tests (i.e. it does not require items to be specifically written for analysis). Furthermore, it is ideal in describing a population where it is hypothesized that there are two types of people: an IRT portion appropriately characterized for promotional decisions and a latent class portion ideal for instructional planning for remediation decisions.

In order to determine if the information provided by the Hybrid model is qualitatively different from that obtained from more traditional measures, two analyses were made. The Hybrid model was first compared to an IRT model in terms of fit. Next, a sub-score method was used to identify the classes used in the latent portion of the Hybrid model. A statistical comparison was made to assess if the same students were being identified by the sub-score and the latent class portion of the Hybrid model. It is worth noting that the Hybrid model offers a flexibility in defining classes not afforded by the sub-score method. Using the Hybrid model a school district, school, or teacher could easily create a latent profile based on any meaningful grouping of items. These diagnostic profiles could be used in a variety of educational context such as peer-tutoring, computer assisted instruction and tutorials. Furthermore the Hybrid method actively involves teachers in the process of test use.

In summary, in order to examine the instructionally relevant information of the RCT mathematics test, this study applied the Hybrid model. Latent classes were defined in terms of mathematical content and psychological constructs. Validation for using the Hybrid model was determined by comparing fit with alternate statistical procedures (the IRT model or the sub-score method) and by teacher judgment. Teachers evaluated the latent class groupings of items obtained from the Hybrid model to determine instructional relevance. The use of both statistical comparisons and teacher judgments provided validity-related evidence for the use of the RCT mathematics for instructional purposes.

Hypotheses

Four hypotheses were examined. It was hypothesized:

1) that the Hybrid model would provide a better fit for the RCT

data than an IRT model alone. Fit was determined by using a chi-square statistic. This hypothesis tested whether there was the possibility of identifying instructionally relevant information (latent classes) in the RCT mathematics.

2) that the latent classes developed using the Hybrid model would identify different sets of students from the sets of students identified by using a sub-score method. A comparison of the sets of students identified by the Hybrid model and the sub-score method was made using Cohen's (1960) κ of agreement. This hypothesis tested whether the information obtained from the Hybrid model differed from the information provided by the sub-scores.

3) that the latent class portion of the model would describe an educationally significant percentage of the sample of students. Educational significance was defined by the percentage of students that equaled one teacher's program. In each school, one teacher's remediation program consists of approximately one hundred students. This hypothesis tested the practical, as opposed to the statistical, significance of the use of the hybrid model with a competency test such as the RCT mathematics.

4) that the latent classes would be judged to be instructionally relevant units. The relevance of the classes to instruction was evaluated by experienced mathematics teachers. This hypothesis examined whether teachers of mathematics could use the information developed by the Hybrid model.

REVIEW OF LITERATURE

The following section first reviews classifications systems that are used to describe cognitive processes. Next, this section provides a description of the statistical model and its components used in this study. Lastly, it presents a review of statistical models proposed for educational research and their applications.

Category Systems for Subject Matter and Cognitive Processes

When constructing an item, a writer may make decisions about factors such as the content, descriptions of the problem situation, characteristics of the correct response and, when using a multiple choice format, characteristics of the incorrect response. There are numerous approaches and category systems used to classify a test item. At the very minimum, an item's classification should include the subject matter content and the cognitive process(es) it is intended to elicit.

Superficially, the content or subject matter topics a test will cover are relatively easy to define. The State Education Department defines the item content on the Regents Competency Test in mathematics according to the first seven units listed in the General High School Mathematics text (1978, pp. 5-8). The units include integers, rational numbers, graphing, measurement of geometric figures, ratio, proportion and percent, probability and statistics, and consumer and job related mathematics. These units are defined by the seven topics which are described in Appendix D, Description and

Examples of Item Content.

However, the cognitive processes used by an examinee are more difficult to define. The RCT math test plan does not specifically define the cognitive processes or skills used in this examination. Therefore, in order to impose a category system on the RCT math, an appropriate system must be chosen from those currently available.

Various cognitive process classifications have been used when developing tests. Perhaps the most widely known classification system is Bloom's (1956) Taxonomy of Educational Objectives: Cognitive Domain. According to Bloom et al, educational objectives are classified as knowledge, comprehension, application, analysis, synthesis and evaluation. This taxonomy has been useful in education because of its clarity as well as its broad application. However, for the RCT math, classification systems specifically designed for mathematics items may be more appropriate and informative.

A classification system created by the Educational Testing Service suggests that in answering an item a student may recall factual knowledge, perform mathematical manipulations, solve routine problems, demonstrate comprehension, solve non-routine problems requiring insight or ingenuity, and apply higher mental process to mathematics. However, this classification system appears too complex for the more elementary nature of the RCT math items.

Another classification system, proposed by the School

Mathematics Study Group, defined an item by processes such as knowing, translating, manipulating, choosing, analyzing, synthesizing, and evaluating. The simplest classification system and the one most suitable to the cognitive processes used in the RCT, are the process levels proposed by the National Council of Teachers of Mathematics(NCTM)(1989). These processes include:

knowledge - This item requires the recall or recognition of mathematical ideas, figures, or symbols. Knowledge items include such tasks as identifying common geometric figures or recalling a number fact.

Ex. Which of the following is a prime number?

(1) 13	(3) 21
(2) 15	(4) 4

skill - This item requires routine mathematical manipulations that have been learned or practiced. The item does not require the student to decide what operation to use. The item assesses proficiency in using an algorithm rather than the understanding of how it works. Such items might require the student to make a measurement, multiply fractions, solve an equation, or read a table.

Ex. Solve for x: $3/4 = x/12$

(1) 3	(3) 9
(2) 48	(4) 36

routine application - This item involves knowledge or skill with which the student is presumed to have had experience. Although the mathematical operation is not specified, the identification of the proper procedure is almost automatic.

Ex. Michael bought orange juice for $.89, a bagel with cream cheese for $.65 and a math book for $5.95. How much did he pay for the items. (Assume there is no tax on any of the items)

(1) $7.49	(3) $7.39
(2) $21.35	(4) $7.89

understanding and comprehension - This item assesses understanding focused on basic mathematical concepts and principles. It frequently requires the student to identify or establish relationships between different representations, such as identifying which model represents a given number sentence.

Ex. Which mathematical sentence represents five less than a number is 14.

(1) $x-5=14$	(3) $5x=14$
(2) $x-5>14$	(4) $x-14=5$

problem solving - This item requires higher-order thinking that involves the integration of concepts and skills to solve problems for which there is no clear method of solution. This is an item that can not be solved by a routine application of a concept or skill.

Ex. A stereo can be purchased with a downpayment of $125 and 10 monthly installments of $25 each. What will be the total cost of the stereo?

(1) $250	(3) $160
(2) $375	(4) $390

The classification system of the NCTM was used because it is the system that most reflects the nature of the actual questions used on the RCT, and because of the teacher's familiarity with the NCTM system. Written descriptions of the content blueprint used in constructing the RCT and the NCTM category system for cognitive processes were provided to the teachers who were asked to classify the items according to these schemes. (These procedural steps are described in the methods section.) Because the RCT is not a cognitive measure, it does not utilize fully the improved assessment procedures outlined by Snow (1989). However, as Snow suggests, the RCT can be viewed as a diagnostic measure of the initial state of the learner for the purpose of suggesting alternative "next steps" to the teacher.

Statistical Models

The Hybrid model was developed in response to a need for a model that could, when appropriate, rank individuals according to their ability as well as to identify the "cognitive structure" of individual examinees (Yamamoto, 1987). Thus, the Hybrid model was designed to characterize a population in which there exists two types of examinees. For one type, ability classifications are appropriate and the examinees' responses fit the IRT modeling. For the other, different classes of examinees exist and their responses are best characterized by latent class models. It is this latter situation that is being examined in the present study. The advantage of the Hybrid model is that it improves on IRT modeling and latent class analysis used separately. When these models are used in conjunction, the Hybrid model has the ability to utilize the strengths of both statistical models. The Hybrid model can represent cognitive structure, draw upon test users' informal error analysis, and incorporate IRT so that inter-item equivalency can be assessed and differential item analysis performed (Yamamoto, 1987).

In order to understand the Hybrid model in more detail a brief description of IRT and latent class theory is given.

Item Response Theory

Item Response Theory (IRT) proposes a mathematical relationship between an unobservable ability and an examinee's performance. This theory states that as a latent ability increases so does the probability of an examinee's responding correctly on test items. Two assumptions are central to the IRT model. The first assumption is local independence, which states that at a given ability level the responses of an examinee on individual items are independent of one another. In other words, the joint probability of responding correctly to any two items is equal to the product of the item's individual probabilities. The second assumption, unidimensionality, states that the statistical dependence among test items can be accounted for by a single latent trait.

The relationship between ability and probability of responding correctly is modeled by the item characteristic curve (ICC). The ICC is a plot of the proportion of persons who respond correctly to an item as a function of a latent trait. This function increases monotonically i.e. examinees are more likely to respond correctly to an item as the ability measure increases. The ICC has been mathematically defined by three different logistic models. It is the inclusion of the parameters for difficulty, discrimination and guessing that distinguishes the models. The general form is

$$P_g(\Theta) = \frac{e^x}{1 + e^x}$$

where $p_g(\Theta)$ is the proportion who respond correctly to an item g , e is the base of the natural log and x is defined by the number of parameters that are allowed to vary in the model. For the two parameter model $x = Da_g(\theta - b_g)$. Where D is a constant and a_g and b_g are defined as discrimination and difficulty parameters respectively. The one-parameter or Rasch model is a special case of the two-parameter model where the discrimination parameter is held constant and only the difficulty parameter is estimated (i.e. the variable a_g becomes the constant a). The three-parameter model contains the discrimination and difficulty parameter and adds a guessing parameter, c_g,and a multiple of $1-c_g$.

Pg(Θ) is usually interpreted as the probability of a correct response among examinees at a given ability level rather than the probability of a correct response of a given individual with ability theta (Hulin, Drasgow and Parsons, 1983). A continuum of latent ability is appropriate for characterizing the population. However, if ordering is the only result obtained for the individual, this information is difficult to use at the classroom level and creates a problem for the interpretation and use of an individual's scores. According to Yamamoto (1987), for the individual, a continuous increase in both theta and probability of responding correctly is unlikely. Rather, discrete levels of probability which reflect the occurrence of learning are more accurate. IRT can appropriately model two discrete ordered levels of probability. Once the number is greater than two, unordered latent class is the appropriate model and provides more meaningful information for both classifying learners and for making instructional decisions(Yamamoto, 1987).

IRT is useful in the evaluation of tests, equating multiple forms, item bias studies and ordering the achievement of examinees. At the classroom level, however, it is not as useful. IRT is unable to identify cognitive patterns, to incorporate informal error analyses at the time of parameter estimation and has difficulty in incorporating single event learning (Yamamoto, 1987). Furthermore, unidimensionality, one of the central assumptions of IRT modeling may be difficult to meet. Lord (1980) states that meeting this assumption is unlikely, even in a highly specific content area. This makes the application of the IRT model in content areas as diversified as high school mathematics questionable. Although corrective methods exist for this particular problem, in view of all the educational limitations, other solutions are preferable.

Latent Class Models

Although many psychological theories have employed a continuous measuring approach to behavior, some theories are best explained by typologies (Rindskopf, 1987). When it is theorized that in a sample there exist distinct classes or types of individuals, latent class models are used. These models, which previously were difficult to use, have become more accessible due to the statistical developments of Goodman and the computer developments of Clogg (Paulson, 1985; Rindskopf, 1983).

Conceptually, the simplest case of a latent class model (LCM) describes only two types of people: for example, those who have mastered a concept and those who have not. Suppose a group of students were given a four item test which perfectly measured a skill. Two distinct response patterns should be observed: a master and a non-master. Ideally, those who have mastered the skill should get all the items correct and those who have not mastered the skill should get all the items incorrect. However, because items are not perfectly constructed and people do not always respond according to their ability (i.e. due to factors such as fatigue, boredom, or guessing) we might observe a master getting an item wrong and a non-master getting an item right. In fact, any one of 16 response patterns is possible with a four item, two outcome test.

Latent class analysis allows for other than ideal response patterns by assuming that the true class to which a person belongs (a latent class) is not directly observable but inferred by response patterns that contain errors of measurement. Latent class analysis postulates a statistical model which predicts how response patterns should be distributed. These predictions are then compared to observed responses to test if a model is tenable.

Statistically, latent class models are used with discrete data where the purpose is to identify the unobserved variable(s) which account(s) for the statistical association among observable measures. Like IRT it also assumes that within a latent class the item responses are independent. In latent class analysis responses of each member of a population conform to a given number of mutually exclusive and exhaustive latent classes. In order to define a latent model it is necessary to define the number of latent classes, i.e. to specify the model. The unconditional probability of falling into each latent category (v_k) is defined as the probability of a subject being in class or state k. Next, the conditional probability of responding correctly to item j, given class membership k is specified (p_{kj}). The key assumption of this model is local independence, the assumption that there is no relationship between the observed measures once the latent variable has been controlled. In other words, the unobserved variable is the variable of interest and explains the relationship among the observed measures.

Parameters for the specified are then estimated, typically using the method of maximum likelihood. These estimates are generated by computer programs (i.e. MLLSA (Maximum Likelihood Latent

Structure Analysis), Clogg (1977)). Expected cell frequencies are obtained from the product of the conditional probability of a response pattern given class membership and the unconditional probability of that latent class. These joint probabilities are then summed across classes to obtain the probability associated with each response pattern. Expected counts are calculated by multiplying response pattern probabilities by the sample size. The fit of the model is evaluated by comparing the observed cell frequencies with the expected cell frequencies by means of either a Pearson or likelihood ratio chi-square statistic.

The obtained chi-square value is compared to a critical value from the chi-square distribution. The degrees of freedom are calculated by the number of independent cell proportions minus the number of independent parameters estimated. In the four item-two outcome example mentioned above, the number of independent cell proportions is one less than the possible outcomes (16-1=15). The number of independent unconditional probabilities, for K classes, is K-1, or in this example one (2-1=1). In each of the two latent classes there are four independent conditional probabilities of getting an item correct (one for each item). The total number of independent parameters is then nine (1+4+4=9) and the degrees of freedom are six (15-9=6). A small chi-square statistic indicates that the model may fit whereas a large chi-square indicates that the model is incorrect.

However, even though a model may fit the data it is possible that other models also fit the data. In general, the most parsimonious model is preferred. Models may be restricted (where parameters are fixed at certain values) or non-restricted (where parameters are free increases as the number of items increases.

Although the ability to estimate parameters has greatly increased due to the advances of Goodman (1974), his procedure is impractical to implement with large numbers of items (Paulson, 1985). This difficulty arises because the estimation procedure takes as its data the frequency counts of all the possible cells in the n-way contingency table. Paulson developed an algorithm which sums over subjects instead of cells, greatly increasing the number of items the analysis can handle. Similar modifications were made by Bock and Aitkin (1981).

Latent class models are appropriately applied when it is theorized that there are qualitatively different types of abilities rather than a continuous underlying ability. However, with latent class models it is difficult to determine the number of latent classes necessary to define the data especially if the classes are ordered. That is, it is difficult to tell when one class ends and another class starts.

Hybrid Model

The Hybrid model, developed and conceptualized by Yamamoto (1987), follows a probabilistic modeling approach that combines IRT and LCM. Yamamoto's impetus for creating this model was the desire to utilize the strengths of each model and, at the same time, to accommodate the differences. This model incorporates IRT procedures to attach scaled values to examinees and latent class theory procedures to identify qualitatively different classes. The value of using either model is dependent upon the planned usage. When

Hybrid is viewed from an IRT perspective, the addition of a latent class part to the model is used for cases where examinees systematically deviate from IRT modeling (i.e., they are considered to guess randomly, have learned the material incorrectly or incompletely) and do not fit IRT modeling. From a latent class perspective the addition of the IRT part of the model represents classes where examinees are ordered and the ordering is the most salient aspect about their responses. According to Yamamoto, assumptions and usage of the two models have been incompatible; a model that can represent patterns of responses and the relationship among items is needed.

To illustrate the Hybrid model an example is offered. In the simplest case, assume there exists a data set of dichotomous responses from a population composed of two distinct groups whose responses are modeled by IRT or LCM respectively. Within the latent class part of the model there may be several latent classes, each of which represents a unique cognitive pattern. Each latent class is characterized by an idealized response pattern. The deviations of the responses of subject's within a class from the idealized response pattern are identified as execution errors and appear as parameters in the Hybrid model. Each class can be thought of as a specific erroneous rule class with a consistent application of the particular misunderstanding (Yamamoto, 1987).

The Hybrid model assumes the following:

1) each examinee belongs to the IRT group or to one of the latent class groups; the IRT group and the classes in the latent class group are mutually exclusive and together exhaustive.

2) class membership is unknown; hence the data cannot be separated into multiple subsets of data prior to parameter estimation.
3) within the IRT group, both of the standard assumptions of IRT apply: a) local independence, and b) unidimensionality.
4) responses are conditionally independent given a latent class; this is equivalent to the local independence of IRT.
5) classes are to be expressed by particular response vectors. (Yamamoto, 1987, p.19).

Yamamoto also notes that the ability distribution of the IRT group is independent of the mixture proportions of IRT and latent-class groups.

To use the Hybrid model, the number of latent classes, the idealized response pattern, and the constraints of latent class parameters must first be specified. The Hybrid model then estimates the distribution of examinees in each part of the model. It also estimates IRT parameters for each item and each examinee and the conditional probabilities of each latent class (Yamamoto, 1987). The Hybrid model can also be used in an exploratory fashion, if the parameter constraints are relaxed (Gitomer & Yamamoto, 1991). Hybrid also can classify each examinee into either the IRT part of the model or one of the latent classes using a likelihood ratio test (Yamamoto, 1987).

The model includes the usual IRT and latent class parameters in addition to a parameter that specifies the proportion of examinees in each class.

The notation for the parameters in Yamamoto's (1987) Hybrid model are as follows:

X_{ij} = 1 if subject j is correct on item i
= 0 if subject j is incorrect on item i

I = number of items i=1,...,I

J = number of subjects j=1,...,J

K = number of groups (including the IRT group and all classes in the latent-class group)

$P(\gamma = k)$ = probability of a randomly sampled subject being in a class k
k = 1 for IRT
k = 2,3,...,K for latent classes

ζ_i = item parameters (a_i , b_i), and latent class parameters

Θ_j = ability parameters of subject j

$f(\Theta)$ = probability density function of Θ in the IRT group

Γ_{jk} = class membership matrix; for given subject j, only the Γ_{jk} with k standing for the class to which he belongs is 1 and all other Γ_{jk} , are 0

Q_t = number of discrete values of theta within the IRT group, used to approximate the integration over Θ by the summation across discrete Θ values, q=1,...,Qt

$A(\Theta_q)$ = quadrature weight associated with Θ_q (IRT)

R_{1iq} = number of correct responses on item i among IRT subjects who are at a fixed theta Θ_q

N_{iq} = number of attempts on item i among IRT subjects who are at a fixed Θ_q

R_{1ki} = number of correct responses on item i among subjects in a latent class k

N_{ki} = number of attempts on item i among subjects in a latent class

Logistic IRT (2-parameter)

$$P_{1ij} = P(x_{ij} = 1 \mid \Theta_j, \zeta_i)$$

$$= \frac{1}{1 + \exp(-Da_j(\Theta_j - b_i))}$$

$$Q_{1ij} = Q(x_{ij} = 0 \mid \Theta_j, \zeta_i)$$

$$= 1 - P_{1ij}$$

Latent Class

$$P_{ki} = P(x_j = 1 \mid \Gamma_k = 1)$$

$$Q_{ki} = P(x_j = 0 \mid \Gamma_k = 1) = 1 - P_{ki}$$

(pp. 19-20)

For an examinee selected from the mixture population (i.e., the unsorted or hybrid population), the conditional probability of a response pattern $\underline{x}_j$, given ability level, item parameter values, and class membership is expressed by

$$P(\underline{x}_j|\Theta_j, \zeta, \Gamma) = \prod_{i=1}^{I} P_{1i}(\Theta_j, \zeta_i)^{x_{ij}} Q_{1i}(\Theta_j,\zeta_i)^{1-x_{ij}} \quad \text{if } (\Gamma_{j1} = 1) \qquad (1)$$

$$= \prod_{i=1}^{I} P_{kij}^{\ x_{ij}} Q_{kij}^{\ 1-x_{ij}} \quad \text{if } (\Gamma_{jk} = 1, k \geq 2)$$

P_{kij} and Q_{kij} remain constant over j for all j's with $\Gamma_{jk} = 1$. This expression can be thought of as the likelihood of an ability level, item parameter values, and class membership conditional on responses $\underline{X}$.

Marginal Likelihood Function

The conditional probability of observing response pattern $\underline{x}_j$ conditional on ability and item parameters is given by

$$P(\underline{x}_j | \Theta_j, \zeta) = \prod_{i=1}^{I} P_{1i}(\Theta)^{x_{ij}} Q_{1i}(\Theta)^{1-x_{ij}}$$

The conditional probability of observing response pattern $\underline{x}_j$ from a randomly sampled subject from a population with a density function $f(\Theta)$ conditional on known item parameters can be obtained by integrating the above equation over the IRT population to obtain:

$$P(\underline{x}_j|\zeta) = \int_{\Theta} P(\underline{x}_j | \Theta, \zeta) f(\Theta) d\Theta \qquad (2)$$

This is the marginal probability of response pattern xj, given item parameters and f(Θ). In the mixture population, Γ has a density function p(γ), since the probability function of classes can be used in place of Γ if there is random selection of subjects. Let p(γ =1) represent the proportion of subjects who are in the IRT portion of the sample and p(γ=k) where k≥2 represent the proportion of subjects who are in latent class k. If $P(\underline{x}_j|\zeta)$ is rewritten as $P(\underline{x}_j|\zeta,\gamma=1)$, and by letting ζ include within latent-class parameters as well as item parameters of the IRT group, then the conditional probability of $\underline{x}_j$ given class parameters is:

$$P(\underline{x}_j|\zeta) = \sum_{k=1}^{K} P(\underline{x}_j|\zeta, \gamma=k)\bullet P(\gamma=k)$$

$$= L(\zeta|\underline{x}_j)$$

Therefore

$$L(\zeta|\underline{X}) = \prod_{j}^{J} L(\zeta|\underline{x}_j) \tag{3}$$

$$= \prod_{j}^{J} \sum_{k=1}^{K} P(\underline{x}_j|\zeta, \gamma=k)\bullet P(\gamma=k)$$

The above expression can be thought of as a likelihood function of parameters based on observed responses. Normally, to maximize this likelihood function, the parameter value which sets the first derivative of the function equal to zero is calculated (marginal maximum likelihood estimation). Because this process is burdensome, Hybrid estimates parameters using a method based on the EM algorithm of Dempster, Laird, and Rubin (1977). Additionally, the idea of using the EM algorithm with probit-analysis inner cycles, advanced by Bock and Aitkin (1981), is used. In this method, integration is achieved by replacing continuous theta with discrete theta points represented as convenient quadrature points. The IRT theta is "converted" to a discrete theta (Yamamoto, 1987).

Hybil Program

In order to understand the hybrid model (Yamamoto, 1987) and the functioning of the Hybil program (Yamamoto, 1991) the following example is offered. A sixteen item test is constructed to analyze mathematical proficiency. For the purposes of this example, mathematical proficiency will be defined by four areas such as algebra (a), geometry (g), trigonometry (t), and calculus (c). Thus, the test items are arranged as follows:

agtcagtcagtcagtc

Furthermore, assume that the ranking of the item difficulties is 1 for the first four items (i.e. the easiest), 4 for the next four items (i.e. the most difficult), 2 for the next set of four and 3 for the last four items. A data set of twenty unique cases was constructed and then duplicated to a sample size of 640.

The data set was created to reflect a latent class population in which five percent were deficient only on algebra items, five percent only on geometry, five percent only on trigonometry and five percent only on calculus (cases 1-4). Twenty-five percent of the hypothesized population responded in accordance with the item difficulty levels (cases 5-9). Ten percent of the data set was constructed to reflect a deficiency in two or three content areas (cases 10-11). Ten percent of the data reflected a testing situation where students became fatigued and did not answer the final items (cases 12-13). The remaining thirty-five percent of the data were random , i.e., the responses were coded following no particular pattern (cases 14-20). The data set is as follows:

Case Number	**Pattern**	**Description**
1	0111011101110111	Content Deficiency
2	1011101110111011	
3	1101110111011101	
4	1110111011101110	
5	1111111111111111	Item Difficulty Levels
6	0000000000000000	
7	1111000000000000	
8	1111000011110000	
9	1111000011111111	
10	0011001100110011	Two and Three area Deficiency
11	0001000100010001	

12	1111111111110000	Fatigue
13	1111111100000000	
14	1010101010101011	Random Responses
15	1101100001010101	
16	1101000100101010	
17	0011010111010010	
18	1100101010100101	
19	0000100010001001	
20	1010101011111010	

In order to apply the Hybil program, the user must have a theory about the structure of the latent classes. In the simplest case the user might hypothesize that there are four classes of people, each of which has a deficiency in a single content area.

There are three main sections of the Hybil Program's output: the IRT output, the latent class output, and the subject output.

IRT output.

PRIOR AND POSTERIOR THETA DISTRIBUTION OF THE IRT GROUP PROPORTION OF SUBJECTS IN IRT GROUP =.5524

RESCALED A AND B VALUES ACCORDING TO SUBJECTS STANDARDIZED DISTRIBUTION SD OF THETA ESTIMATE

ITEM ORDER	ITEM	A	B
1	1	4.32519	-.41262
2	2	4.32519	-.41256
3	3	4.32519	-.42035

ITEM ORDER	ITEM	A	B
4	4	4.32519	-.96495
5	5	.28162	1.39367
6	6	.40330	1.00326
7	7	.55744	1.39367
8	8	.22265	-.62371
9	9	1.04709	-.26371
10	10	4.32519	-.18728
11	11	4.32519	-.19507
12	12	1.09961	-.28325
13	13	.30442	1.29943
14	14	4.32519	1.06755
15	15	.45101	.86219
16	16	.23313	1.57407

The predicted item difficulties (b parameter) were fairly constant with the intended values except for item 8. However, once the Hybil program has apportioned subjects into either the latent class or the IRT group, the nature of the remaining data set changes. Once the latent subjects have been identified, item 8 does appear easier for the remaining subjects.

Latent class output.

PROPORTION OF SUBJECTS IN LATENT CLASSES

1	.05000	0111011101110111
2	.14601	1011101110111011
3	.10002	1101110111011101
4	.15159	1110111011101110

The Hybil program placed forty-five percent of the hypothesized sample in the latent portion of the model as opposed to the intended twenty percent. An additional twenty-five percent of the data set was classified as belonging to one of the idealized latent classes. A review of the data set revealed that four (twenty percent) of the "random" response patterns were very similar to the idealized latent class response pattern (see cases numbered 14, 15, 18 and 20). Additionally, upon inspection, the response pattern for one subject which reflected a need for algebra and geometry remediation, (see subject output, case 10) exhibited 75% agreement with the latent class reflecting a subject in need of geometry remediation only.

Subject output.

The subject data first prints the theta estimate for each subject in the IRT population. To designate latent class membership a code of 99.000000 is used in place of the theta estimate. Additionally, the posterior probabilities for the twenty quadrature points, followed by the latent class probabilities are printed (see Appendix K).

Application of Statistical Models in Educational Research

Models Proposed for Educational Assessment

In education, for the purposes of diagnosis, remediation, and curriculum revision, knowledge of how students solve problems could be more valuable than how many they solve (Messick, 1984; Mislevy and Verhelst, 1990). However, conventional use of IRT models yields no information about how an item's difficulty is influenced by the particular strategy used. Consequently, Mislevy and Verhelst (1990) adapted the IRT model for use on response data when examinees use a fixed set of different solution strategies to solve problems.

Building on the general IRT model, item parameters are expressed by a smaller, more defined set of parameters associated with the number or nature of solution strategies employed. Subject parameters describe strategy use and proficiency. The probability of response pattern x_i, conditional on the subject parameters ϕ_i and θ_i and item parameter α is expressed by:

$$p(x_i|\phi_i,\theta_i,\alpha) = \prod_k \left\{ \prod_j \left[f_k(\theta_{ik}, \beta_{jk}) \right]^{x_{ij}} \left[1 - f_k(\theta_{ik}, \beta_{jk}) \right]^{1-x_{ij}} \right\}^{\theta_{ij}}$$

where β_{jk} gives the item parameters for item j under strategy k and θ_{ik} gives the proficiency of subject i under the k strategies. To properly use this adapted IRT model, items must be constructed that maximize different strategy use. The Mislevy and Verhelst (1990) adaptation of the IRT model requires a strong cognitive foundation which is not available in any meaningful way in many school subject matter areas.

Masters (1985) compared IRT and latent class analysis (LCA) using Likert-type data. Although the data used were attitudinal and directly related to the type of data used in this study, some general observations about the differences and similarities of the models are worth noting. According to Masters, the major difference in the models is that IRT expresses the relationship of the probability of a correct response as a function of a latent ability. Latent class models do not model any such relationship, but rather simply estimate parameters. These model assumptions influence the interpretation of lack of fit. Lack of fit in IRT indicates that items do not represent a single latent variable (multidimensionality). In LCA a lack of fit would indicate the need to test another model (i.e. a different number of classes).

Similarities of the Mislevy & Verhelst (1990) and Masters (1985) methods include treating items as indicators of an underlying variable and assigning respondents to a particular class. Both are probabilistic models which assume local independence. In both models fit is determined by comparing observed and expected response frequencies.

Masters (1982) developed a Rasch model for the analysis of partial credit data. In previous ordered response category models, person and item parameters could not be separated. In the Partial Credit model parameters appear additively in the exponent and allow for parameter separability. The ability to separate person and item parameters results in sufficient statistics for the person and item. Ability is estimated free of the particular items and the estimation of item difficulties is free of the particular sample. This model is applied when the solution is viewed as a series of steps. The difficulties of the steps are assigned subscripts, i.e., parameterized. The probability of person n completing the k'th step of item i is

$$\phi_{kni} = \frac{\pi_{kni}}{\pi_{k-1,\, ni} + \pi_{kni}} = \frac{\exp(\beta_n - \delta_{ik})}{1 + \exp(\beta_n - \delta_{ik})}$$

where π_{kni} is the probability of person n responding in category k to item i, β_n is the ability of person n and δ_{ik} is the difficulty of the k'th step in item i.0

Based on the above expression and the assumption that a person must fall into one of the possible categories, the partial credit model states

$$\pi_{kni} = \frac{\exp \sum_{j=0}^{k} (\beta_n - \delta_{ij})}{\sum_{h=0}^{m} \exp \sum_{j=0}^{h} (\beta_n - \delta_{ij})} \qquad k = 0,1,...m_i$$

In the Partial Credit model the steps in an item's solution need not be ordered in difficulty which makes the application of this model to education promising.

In summary, although several newer models have been proposed for educational assessment, these models are not useful at present. The Mislevy & Verhelst (1990) model has stringent item development requirements, as does the Masters (1982) partial credit model.

Applications of Statistical Models to Achievement Tests

There have been only a few studies that have used achievement data with latent class models to identify groups of students for diagnostic or instructional purposes. These include studies by Haertel (1984a, 1984b & 1989), Bock (1972), Tatsuoka (1983), Birenbaum, Tatsuoka & Gutvirtz (1992), Birenbaum, Kelly & Tatsuoka (1992), Birenbaum, Kelly & Tatsuoka (1993), Bergen, Stone and Feld (1984), and Gitomer & Yamamoto (1991).

Haertel (1984a) applied latent class models, specifically a two-class state mastery model, to multiple-choice reading comprehension items. In this model, there were two types of examinees: masters, who can solve all of the items and nonmasters, who can solve none of the items.

The item content determines the reasonableness of applying this model. The state model has been successfully used in areas such as arithmetic computation where item solution requires the application of a learned procedure. However, in areas such as history, where the solution of items does not require the application of a procedure but rather requires knowledge of a particular fact, the state model is not appropriate. Haertel (1984a) chose to analyze reading comprehension items because they represented an area in between the two extremes. Reading comprehension items require a common set of skills for solution while being more complex than a simple two-column subtraction problem.

Haertel (1984a) analyzed the responses of 2,089 fourth-grade students to the reading comprehension subtest of the Metropolitan Achievement Test (MAT). He sought to determine if: 1) a set of items could be assembled that the two state model would fit; 2) a domain of items could be defined by stimulus features such that any item drawn from that domain with the same features would also give an acceptable fit; and 3) whether different subsets of items drawn from such a domain would give consistent classifications of masters and nonmasters.

Items were analyzed in sets of six because this was the maximum number of items that could yield full information maximum likelihood estimates and still provide sufficient degrees of freedom. The number of items analyzed has no effect on the length of the mastery test which can be constructed from the model.

Of the eighteen subtests constructed from the forty-five item test, five subtests yielded a satisfactory fit for the state model. Therefore, it was concluded that a set of items could be assembled that would fit the state model. Based on these analyses twenty-four items that appeared to conform to the state model were identified. Regardless of how these items were combined, the state model fit. Visual inspection of the items revealed that those conforming to the model were inference items while those not conforming to the model were literal items. To confirm this distinction, items were independently classified both by the author and by graduate students. The proportion of items consistently classified by pairs of raters averaged 0.84. Item difficulty was also found to relate to domain conformity, which was revealed here by a visual inspection of the stem-and-leafs of difficulty of conforming and non-conforming items. For the easy items (over 70% correct) only 17% conformed to the model while for the difficult items, 88% conformed to the model. When combining these two stimulus features, 95% of the difficult inference items conformed to the model.

To address the question of consistent classification, two subtests of six items were constructed. Seventy-five percent of the examinees were consistently classified. To determine if a longer test would result in better agreement, a procedure similar to correcting a correlation for attenuation due to unreliability was employed. This procedure estimated that ninety-five percent of the examinees would be consistently classified if the test could be extended indefinitely.

Because it is useful in making classification decisions, Haertel concluded that the state model was of practical significance. The model describes an examinees' performance in terms of a defined standard and is not dependent on the particular item used on a test. Further, the state model provides a classification for each examinee and an estimate of the accuracy of that classification. Lastly, Haertel suggested that the state model may provide better, more defensible test interpretation for mastery classification than continuum models. This hypothesis was not tested empirically with experienced teachers, as it is in the present study.

Haertel (1989) re-analyzed the MAT data from his 1984a study using a binary skills model. In this model the relationship between an observable response and a latent response is probabilistic and characterized by two classification parameters: a) the conditional probability of a correct observable response when the latent class predicts incorrect, (the items false positive probability, π_{1i}) and, b) the conditional probability of a correct observable response when the latent class predicts correct (the item's true positive probability, π_{2i}). For each item there is a latent difficulty parameter, δ_i, which is defined as the proportion of examinees whose latent response to that item is one.

Haertel sought to determine the skills required by a set of items and to create a map of the skill structure that would account for the observable response patterns to those items. He analyzed the items in sets of six as he did in previous studies. To increase the probability of identifying distinct item skills, the subsets of item were constructed to include items with contrasting features (i.e. easy versus difficult, literal versus inference). This approach resulted in 18 six item subsets. Several binary skills model were fit to the data beginning with a two state model and adding additional classes based on an analysis of the residuals. A chi-square difference was calculated to determine if the addition of classes improved model fit.

Once the final models for each of the subsets were identified, the items were compared, pairwise, based on the item estimates of π_{1i}, π_{2i}, δ_i. For each pair, the fraction of examinees who could solve both items, neither item, or one, but not both, was calculated. This information was used to identify those items that require a common skill. Next, the information on the item pairs was used to construct a skill map by identifying clusters of items that require a common skill. These clusters were determined statistically, much like exploratory factor analysis. Inspection of the clusters revealed similarities among items that were consistent with Haertel's previous findings. The author concludes that although neither continuous nor discrete models offer complete pictures of an examinee's abilities, binary skills models applied along with continuous models may offer a new perspective.

Bock (1972) proposed an IRT model for use with scoring multiple choice items. This method utilizes information from incorrect responses when making ability estimates. Bock analyzed the responses of 557 students to four items taken from the vocabulary section of the Cooperative Reading Tests. These items were chosen because they showed better than chance responding to at least one of the four incorrect alternatives. He compared unconditional and conditional methods of item parameter estimation (i.e. whether or not the ability of the subjects influences the estimation of the item parameters). Bock found that as the number of items increases the unconditional method became impractical computationally. However, he recommended the conditional method as both accurate and computationally practical. An information analysis revealed that for students below mean ability, multiple scoring increased the scoring precision comparable to doubling the test's length. For those at or above mean ability the method was equally precise.

The rule space model (Taksouka, 1983) was designed to diagnose student's responses that are produced by the consistent application of an erroneous rule. The model is capable of providing an exact description of an erroneous rule, predicting its likelihood and diagnosing the particular error pattern. Analysis of an erroneous rule response pattern and the nature of the student's misconception is potentially useful in evaluating and planning instruction.

In the rule space model all response patterns, based on erroneous rules, are mapped into a set of coordinate pairs based on latent ability (Θ) and an IRT-based caution index (ζ) introduced by Tatsuoka (1984). The parameters for the model are estimated from the entire sample of examinees. The resulting "rule space" is much like graphing in a Cartesian plane. A particular student's response pattern is plotted and an error classification is made based on calculating the distance from the student's response pattern and the nearest centroids. Tatsuoka illustrated the model with the addition and subtraction of signed numbers.

Birenbaum, Tatsuoka & Gutvirtz (1992), using a bug analysis and a rule space analysis, examined the effects of response format on diagnostic assessment. They analyzed the responses of 231 eighth and ninth graders (14-15 years old) of high and low achievement groupings. The students were given a 48 item (thirty-two open-ended items, and 16 multiple choice) diagnostic test of linear algebra with one unknown developed by Gutvitz (1989) based on detailed task analysis. The test was analyzed in subsets of eight, creating four open ended subtests and two multiple choice subtests.

The bug analysis is based on a detailed examination of the procedural errors students make in solving an item. After these errors are identified, the solutions, according to these mal-rules (i.e.

incorrect or incomplete conceptions on how to solve an item), are created. The students' responses are then matched to the mal-rules (matched correct, matched bug, one correct one error, non matched bug or error).

For the rule space analysis, it was first necessary to identify the task attributes necessary for an item's solution. The BUGLIB program was used in order to generate the ideal response patterns corresponding to an attribute mastery pattern.

The results indicated that for either analysis the open ended subsets were more similar to each other than to the multiple choice subtests with respect to percent matched (matched correct, matched bug or matched mastery, matched non mastery). Furthermore, the open ended items were more similar with respect to knowledge state classifications as well as in comparisons of mastery and non mastery classifications of single attributes.

Birenbaum, Kelly & Tatsouka (1992), using the same subjects and the 32 open-ended items, sought to compare the stability of the diagnosis of student errors obtained by a bug analysis with that obtained from a rule space analysis. The bug analysis is based on identifying the mal-rules or bugs a student might use and then constructing the item's solution based on these misconceptions. The students' responses are then matched to those created by the bugs and are coded accordingly .

The rule space model was used to classify students into knowledge states by calculating the distance from a given student's response pattern to that of the nearest ideal centroid.

The results indicated that for the bug analysis, 64.58% of the total matched responses were matched correct and 10.07% were matched incorrect. For the rule space analysis, 63.38% of the responses were matched mastery and 16.80% were matched non mastery. The authors concluded that the rule space analysis of task attributes provided a relatively stable representation of student errors. The advantages of using the rule space model is that task attributes are known entities which are subject to remediation. These attributes are integral subcomponents of the task; hence, failure on the task is traceable to subskill deficiencies. With the rule space model, remediation prescriptions are possible, whereas it is often difficult to understand the nature and conditions of bugs.

Birenbaum, Kelly and Tatsouka (1993), using the same data set and instrument as in their 1992 study, applied the rule space model with the intent of providing a diagnostic profile of the learner for the purposes of instruction. Two methods for the solution of the items were identified along with the task attributes for a particular method. The solution strategy used by method I subjects included 14 task attributes and for method 2 subjects, 13 task attributes. The BUGLIB program was used to determine the ideal item response pattern that corresponded to a particular attribute mastery pattern.

For method 1, 461 mastery attribute patterns were identified and for method 2, 453 mastery patterns were identified. The method most likely used by each student was determined by calculating the distance between the actual mapping of a student's response pattern and calculating the distance to the nearest ideal pattern centroid.

With method 1, at least one student was classified into 55 of the 461 patterns and with method 2, at least one student appeared in 51 of the 453 patterns. Differences existed between methods with respect to the item difficulties as well as with the mastery level achieved for the attributes common to both groups. Based on the attributes that comprise these groups a tree diagram was created. The results are displayed as branches from a larger set of task attribute deficiencies leading to subsets of those same attributes [i.e. attributes numbered (1,2,3,4,5,6,7,8,) →(2,3,4,5,6,7)→(3,4,5)] suggesting paths to follow in the sequencing of remediation . The use of the rule space model requires a strong cognitive framework and its usefulness with remediation prescriptions remains to be tested empirically.

The rule space model is effective in identifying misconceptions in similar areas where the rules for solving a problem are clearly defined and limited in number. In real life teaching of mathematics, these areas are few. Furthermore, the model is only capable of analyzing the response components two at a time. For students with severe gaps in understanding, artificially dichotomizing their responses is not useful.

A direct application of latent structure modeling to mathematics learning was conducted by Bergan, Stone and Feld (1984). They used latent class analysis to examine the nature of the development of rule replacement in simple counting skills. Four variants of the counting task (counting on by one, starting with one; counting on by one, starting with numbers other than one; counting down by one, from a number greater than one; and counting on by a number greater than one), each assumed to be subordinate to the next, were used to determine if rule replacement followed a restrictive knowledge perspective or a transition knowledge perspective. The restrictive knowledge perspective states that children learn in a stair-step fashion with complex rules substituting as a whole for simpler rules. The transition knowledge perspective states that children learn in three stages: a nonmastery state, a transition state, and a mastery state.

Four hundred eighty-five, three to eight and one-half year-old children of various ethnic and cultural backgrounds participated in the study. The math items used for this study were part of an instrument created to provide cognitive measures for a Head Start organization. Six hypotheses were tested by imposing restrictions on the latent class model. A subordinate and superordinate skill were compared in each model.

The first three models reflected the restrictive perspective. Model one tested the two state mastery model; model two added a third class where only the subordinate skill was mastered. Model three added two classes to the first model: in one class there were masters of only the subordinate items and in the other class there were masters of only the superordinate items. The last three models reflected the transition perspective. Model four is like model two except that it added a fourth class who mastered the subordinate item but responded inconsistently to superordinate items. Model five added to model four an additional class with inconsistent responding to the subordinate items and nonmastery of the superordinate items. The last model, model six, is a derivative of model three where there is an additional transitional class who has mastered the subordinate items and responded inconsistently to the superordinate items.

Models were statistically compared by means of a chi-square statistic. Statistical comparisons are made by subtracting L2 values of hierarchical models and their associated degrees of freedom. Models are hierarchical if one model contains all the restrictions of the first model with an additional constraint. The result is then compared to the chi-square distribution. If the chi-square value is significant then the additional constraint improved the fit of the model. If the chi-square value is not significant then the simpler model is preferred. All the preferred models were consistent with the restrictive knowledge perspective model.

In another study, Haertel (1984b) used exploratory latent class analysis on the National Assessment of Educational Progress (NAEP) mathematics data. He felt that one of the limitations of the results was that they were pooled into broad topics, spanning many objectives and that the reader is "left to imagine what generalizations from these statements are appropriate" (p.333). Furthermore, he felt that many researchers were "... ill-equipped to draw useful conclusions from hundreds of diverse items, each of interest in its own right" (p.333). Consequently, he sought to determine (1) how many latent classes needed to be assumed to account for the patterns underlying examinee performance, (2) whether these classes formed a Guttman scale, and (3) the prevalence of each class in the population.

The analysis was restricted to those items classified as algebraic and that relied on a common skill. A skill was defined as the basic unit of ability and was assumed to be dichotomous and determined by curriculum organization. Twelve different booklets were available during the test administration, with a single booklet for each student. Of the seventy-five algebraic items identified, twenty were rejected as not relying on a common skill. This left six of the booklets with at least six acceptable exercises.

The first stage of the analysis consisted of fitting the simplest model, the two class model, to each of the individual exercise booklets. In order to determine whether a more complicated model should be tried, the residuals were examined. When a new model was tried it was compared to the old model by calculating a difference chi-square. If the value was statistically significant the new model was preferred. All best fitting models formed a Guttman scale (the skills the items required formed a cumulative scale of levels of content mastery) except for the last booklet. For this booklet no model could be found that gave a non-significant chi-square. This booklet was the only booklet to contain non-routine problems and was removed from subsequent analysis.

The next step in the analysis was to identify a single comprehensive scale that could be applied to all five remaining booklets. For each of the booklets only the best fitting models were considered. When two choices existed the simpler model was chosen, except for one booklet where the more complex model accounted for an additional 15% of the sample and corresponded to a class found in three of the other booklets. Although it would have been ideal to re-estimate all parameters and latent class proportions, there was no software available to accomplish this and so a two stage estimation procedure was used. Five latent classes emerged with χ^2 (7, N>1000) = 10.73, p>.15

In addition to fitting the data, for the comprehensive model to be useful it had to define meaningful and coherent levels of content mastery. Although the specific content of items in each class was somewhat diverse, the latent classes all revealed increasing content complexity. This impression was confirmed as the sequencing of classes corresponded to that found in a widely used textbook. According to this model, 43% of the population were unable to solve any of the items. The remaining 57% could solve at least class 1 problems (Prealgebra), 42% could solve at least class 2 problems (Translation), 31% could solve at least class 3 problems (Linear) and 19% could solve at least class 4 problems (Quadratic). Haertel concluded that the use of latent class models to group items together in an empirical manner provided information that could be useful in abstracting policy-relevant generalizations about examinee abilities.

The Hybrid model was used by Gitomer and Yamamoto (1991) to analyze the responses of electronics technicians on two types of logic gates. A logic gate is a subset of symbols that represent the type of voltage input to a circuit, i.e. high or low. There are two types of logic gates: the exclusive OR(XOR) and the exclusive NOR(XNOR). Both types of gates have two inputs (HH or LL or HL or LH). In the XOR gate, when the inputs are the same (HH or LL) the output is L and when the inputs are mixed (HL or LH) the output is H. In the XNOR gate the output is simply the negation of the XOR output. Knowledge of the operation of inputs and negations are required to correctly interpret a gate.

For some technicians, the error patterns observed were qualitatively different; however, for other technicians a nonsystematic error pattern was observed. Therefore, the application of the Hybrid model seemed ideal. The study compared the Hybrid model with a two and a three parameter IRT model to determine: 1) which model better described test performance and 2) whether the level of diagnostic information obtained from the Hybrid model was superior to that of the IRT methods.

Two hundred and fifty-five technicians in the United States Air Force participated in the study. The sample was divided into two groups: 119 took a paper-and-pencil version of the test; the remaining subjects solved the problems on computer. For both groups the test was untimed and consisted of 288 logic gate items. Inspection of the first group's responses suggested that certain error patterns were common across subjects. The five most frequent error patterns were used to define the ideal classes in the latent class portion of the Hybrid model. Membership to a particular class was assigned if 80% of all responses were consistent with the idealized pattern. Three IRT models with one, two and three parameters were fit to the data and compared using Bock's chi-square statistic. Application of Akaike's Information Coefficient (AIC) as well as a monotonic increase for all but one item, indicated that the two-parameter IRT was the most appropriate for these data.

The two and three parameter IRT models were compared with the Hybrid model. The Hybrid model was a three parameter IRT model with a fixed c parameter and five latent classes. The c parameter was fixed at .5 because with only two response choices, pure guessing would result in a correct answer one half of the time. The Hybrid model fit the data significantly better than either of the IRT models. More than 25% of the response patterns were clearly modeled by one of the five hypothesized classes. In addition, when the ICC's for the Hybrid model were compared with the ICC's of the IRT model, the Hybrid's ICCs were much steeper because those individuals with below-chance probabilities of responding correctly (i.e. those who are influenced by an incorrect domain knowledge) were placed in the latent class portion of the model.

Correlations of the estimated conditional probabilities of a subject being assigned to one of ten quadrature points or one of the five latent classes were computed. The probability of belonging to any latent class was unrelated to having an ability score at one of the quadrature points. Likewise, the probability of belonging to any latent class was unrelated to the probability of belonging to another class. In addition, the correlation at neighboring quadrature points was higher than the association between more separated quadrature points.

The authors conclude that the use of a mixed model enables the researcher to link the identified latent classes to particular misunderstandings, thus giving increased potential for improved instructional decision making. In addition, the latent class portion of the model does not require a unidimensional assumption. Lastly, the Hybrid model provided a statistically better fit. The Hybrid model revealed that there are individuals with perfect understanding; there are those who are guessing, and there are those with partial understandings. These partial understandings are not hierarchical and a continuous model is not adequate to fit the data.

The Gitomer and Yamamoto (1991) study focused on a domain that appeared relatively easy to analyze. The results were promising but the domain was simple. The Hybrid model needs to be tested further in more complex domains. Such analyses have already been undertaken by Gitomer and Rock (1989) and in the present study. The RCT mathematics items used in this study cover a wider domain than do the logic gates used by Gitomer and Yamamoto. Furthermore, the present study combines and expands on Haertel's (1984a, 1984b) and Gitomer and Yamamoto's work by: 1) defining classes by item content and by instructional commonalties as determined by math experts; and 2) using the Hybrid model to try to establish practical categories for teaching purposes and then drawing upon experienced math teachers to evaluate the usefulness of the results for instructional purposes.

METHODS

The first section describes the pilot study and how items were chosen for analysis in the main study. The second section describes the main study, in particular the two procedures used to identify students in need of remediation: the development of the classes used in the latent portion of the Hybrid model and the sub-score method.

Pilot Study and Latent Class Development

The following section describes the Regents Competency Test (RCT) in mathematics and why only certain items were chosen for inclusion in the analysis. In addition, the procedure and rationale for the definition of latent classes used in the statistical analysis are described. Latent classes were developed for the full or unrestricted population (those students who scored 0% to 100%) and a restricted population (those students who scored 55% to 68%). It is this latter, targeted segment of the population that is likely to be responsive to specific remediation, as they scored close to the passing score for the RCT (see below).

Pilot Study: Regents Competency Test

The RCT in mathematics is an untimed 60 item test covering the skills described in the first seven units of the State Education Department's publication, General High School Mathematics (1978) (for a list of the topics and test emphasis, see Appendix B). At the beginning of each term a listing of the total raw score achieved by each student is available to teachers. Approximately six months into the year a summary chart listing the percentage of items correct for each of the seven topic units is generated for every student in a remediation class. This information is seldom useful for the teacher because in a remediation class, most students are weak in every area. Furthermore, no distinction is made for those students who are "almost passing" (those who score between 55% and 63%). Thus, the information provided to teachers in these reports is seldom useful for instructional planning. One of the purposes of this study was to obtain information that could assist in instructional planning. In order to provide useful information it was first necessary to determine what part of the RCT should be included in the analysis.

The first 20 items on the RCT mathematics are free response; the next 40 are multiple choice. Although the free response items tend to be simple skill problems, the incorrect responses often contain information about a student's thinking. However, this information is not available to the teacher because only the correctness of a free response is reported.
For example,

> What is the median of the following set of numbers?
>
> 40, 50, 50, 45, 70, 45, 50

A student who responds 50 is probably confusing the median with the mean or with the mode. A student who responds 45 is most likely forgetting to arrange the numbers in numerical order. A student who responds 70 is probably guessing. It is reasonable to infer that the student who chose 45 "seems to know more" than the one who chose 70. Since the actual responses are never seen by the teacher it is impossible to judge the quality of the incorrect response. These free response items are included in the total score and percent correct for the seven unit types, further obfuscating interpretation of the data.

The free response items contain more information than can be utilized by statistical models that use dichotomous data. Consequently, it was decided to investigate whether or not the removal of these items affected the data. The particular focus of the investigation was to determine if any of these items were exceptionally easy or exceptionally difficult since the removal of these items might affect the score interpretation. A preliminary Rasch analysis (Alvarez, 1990) of sixty-nine students who took the June 1989 RCT indicated that for the first 20 (free response) items, all but three fell within two standard deviations of the mean item difficulty (see Appendix C, p.3, Map of items). Item two, an easy item, required the student to read a picture graph and had an ability rating of -2.16. Item 18 required a student to change a decimal to its equivalent fraction and item 20 required a student to add signed numbers. These items were difficult and had ability estimates of 3.05 and 2.37 respectively. Since there were only three extreme items in the twenty free response items, the removal of the free response items should not jeopardize the general interpretation of the test. For the purpose of this study the free response items were not used in further analysis.

The passing grade for the RCT is 39 correct or 65%. However, taking into account the deletion of the twenty free response items, the pass score for this study was determined by taking 65% of the remaining 40 items or 26 correct.

Main Study

Subjects

The sample consists of 1600 ninth grade students who took the June 1991 RCT at three inner-city high schools. Two schools are approximately one-half minority (black and Hispanic) and one-half white (N=900, N=350). The third school is approximately ninety-five percent minority and five percent white (N=350). Students in all three schools are in the lower to middle income socio-economic status. Socio-economic status is determined by the percentage of free lunch recipients in each school.

Nine expert mathematics teachers, three from each school, participated in the study. Experts were defined as teachers with a minimum of ten years teaching experience, which included at least four years teaching the tenth year Regents mathematics curriculum (see Appendix A, Teacher information form).

Instrument

The forty multiple choice RCT mathematics items were used in the analysis. These items are designed to cover the first seven units of the New York State Education Department's publication, General High School Mathematics. The topics include: set of integers; rational numbers; graphing; measurement of geometric figures; ratio, proportion and percent; probability and statistics; and consumer and job related mathematics (see Appendix B).

Data Collection

The RCT data of the 1600 students were collected in June of 1991 during the marking of the exams. In all three schools the data are scored by a scantron machine and stored on tape. The data were copied from these tapes, omitting the students' names and identification numbers. The data were then converted to 1,0 coding.

Procedure Development for the Latent Classes

This section describes the procedures and rationale for the development of the latent classes used to describe the 1991 RCT sample used in the present study. First, the items were classified by mathematical content and psychological construct as determined by the teachers. Next, an IRT analysis was performed on the data to identify the item difficulties. The classes were developed using items that were in the middle difficulty range for the sample.

Item Judging Procedures for the Item Classifications

The experts for this study classified the forty RCT items according to the seven unit areas provided by the state education department as explained in the instrument section. For a complete description and examples see Appendix D. These items were also classified using the cognitive processes described by the National Council of Teachers of Mathematics(NCTM) (1989). These are problem solving, routine application, understanding and comprehension, skill, and knowledge. For a complete description and examples see Appendix E.

Teachers were trained in one workshop which began with the categorizing of items. These items were pre-categorized by the author and verified by a high school mathematics chairperson (a person licensed by the City as an expert in the field). First, definitions and examples of the content and cognitive processes categories were distributed to the teachers who familiarized themselves with the examples and the definitions (see Appendix D and E). Once these definitions and examples were reviewed, the teachers, as a group, sorted the forty RCT items from an earlier test version (January, 1991; see Appendix F).

The forty RCT items (June, 1991) used in the present statistical analysis were distributed and the teachers individually classified these items (see Appendix G). Their classifications were collected, xeroxed, and returned. Agreement of seven of the nine math teachers on both content and cognitive process was required for accepting an item classification. For the content and cognitive processes the percent agreement was 73% and 60% respectively. In a second workshop, the teachers in each school, as a group, completed the final analysis sheet (see Appendix H). Before completing the final analysis sheet the teachers discussed their results in order to achieve consensus. For those items not classified in the first workshop, agreement of two of the three schools was necessary for accepting an item classification; otherwise the item would have been removed from the analysis. This criterion was met.

Latent Class Development

The latent classes were defined in two ways; by the content area and by the cognitive processes of the item. Furthermore, classes were developed for the full and for the restricted population. The full population was defined as those students who scored anywhere from 0% to 100% on the RCT. The restricted population was defined as those students who were "close to passing" the RCT, (i.e. who would be close to the passing score of 26). This group of students was likely to include those who would benefit most from a targeted instructional program and are of most interest to teachers. These procedures resulted in four groups of latent classes: 1) full content, 2) full cognitive, 3) restricted content and 4) restricted cognitive.

For either "population" the ideal latent class response pattern was defined according to the content or the cognitive description of the item using 1 (correct), and 0 (incorrect) coding. For example, the ideal response pattern for the measurement latent class was coded as missing all the measurement items (0) and passing all the other items (1). The latent classes were hypothesized based on at most two content areas or two cognitive processes because it was judged that latent classes of more than two areas would lack meaning and become difficult to interpret.

In order to create the "close to passing" latent response patterns for the restricted classes, it was determined that 13-18 responses to the forty items could be wrong. This set of responses represented percent correct scores ranging from 68% to 55%. A score of 68% was chosen to compensate for the deletion of the two most difficult items on the entire test. (These items always appear in the free response section.) Furthermore, if percent correct scores of 63% to 55% had been used, the resulting range of 15-18 included too few students (N=315) for the analysis.

An IRT analysis was performed on the data to identify the item difficulties. Based on the item difficulties, the easiest items were coded 1, on the assumption that the student would get these answers correct and the most difficult items 0, on the assumption that the student would get these answers incorrect. To stay within the 68% to 55% range, and to try to include at least one item from each content and cognitive process category, the sixteen easiest items (ability estimates of -3.60 to -1.40) were coded 1 and the twelve most difficult items (ability estimates of -0.05 to 2.35) were coded 0. The questions at the beginning of the exam tended to be the easiest items and the questions at the end of the exam tended to be the most difficult. The remaining twelve "middle difficulty" items were coded 1 or 0 according to the category the item represented. For example, if the latent response pattern was intended to represent students who were weak in measurement, then the "middle difficulty" measurement items were coded zero and all the remaining "middle difficulty," non-measurement items were coded one.

This method of coding kept the number of items defining each latent class within the 13-18 range except for the set of integers content area. Because of the large number of items in this area, students falling into this latent class and the combinations with this class were expected to perform below 55%. Furthermore, with this method of coding all the content and cognitive process areas were represented except for graphing. Since there were no middle difficulty graphing items, all of the ideal response patterns for a restricted latent class combining graphing with another content area were identical to the single area latent class (i.e. the probability & statistics latent class had the same idealized response pattern as probability & statistics and graphing latent class). However, these classes were maintained in the statistical analysis should such a class exist in the sub-score method.

Statistical Analyses

The HYBRID model was applied to the June 1991 RCT data to determine whether its application improved the diagnostic assessment of student responses. The latent class portion of the model identifies the instructionally useful information which was previously confounded with the IRT portion. Therefore, it was hypothesized that the Hybrid model would provide a better fit to the data than the IRT model alone. Four initial Hybrid analysis were run (i.e. full content, full cognitive, restricted content, restricted cognitive). From these analysis, those classes that described .01 of the sample or higher were used for the final Hybrid model.

Two analyses of the model were made. First, the Hybrid model was compared to a two parameter IRT model in terms of fit. These two models were compared by means of a chi-square statistic. Second, to assess whether the latent class information developed by the cognitive and content analysis of the items by math teachers could be obtained using a simpler measure, a sub-score was calculated to identify the latent classes.

Sub-scores were calculated in each of the seven content and five cognitive processes areas by calculating a percentage correct sub-score. For example, in the integer area there are ten items. To calculate the integer sub-score, the total number of correct integer items was divided by ten. The sub-score measure was then compared to .65 (i.e. 65%). If the sub-score was greater than .65 the subject was assigned a one in this sub-area; if the sub-score was less than one the subject was assigned a zero. This procedure was followed to calculate each of the twelve sub-scores. A total score was also calculated by tallying the total number correct and dividing by forty. Subjects scoring over 65% were assigned a one, those scoring below 65% were assigned a zero. This resulted in thirteen measures (twelve sub-score and one total score) for each subject. The total score measure was used only for the restricted group analysis.

To make the sub-score classes comparable to the fourteen classes used in the Hybrid model the following procedure was followed. To be identified as belonging to the sub-score Integer/Rational Number area the subject had to receive a zero in these two areas (i.e. score less than a .65) and one (i.e. greater than .65) in all the other sub-areas.

This method of identifying classes was then compared with the latent portion of the Hybrid model to determine if the two methods identified the same students for the full and restricted population groups. However, to make the students identified by the sub-score method comparable to the students identified by the Hybrid model restricted group, the sub-score students were first selected on their total score (i.e. total score less than 65). The data of only the failing students were used because by restricting the range of total score, the sub-score group becomes more comparable to the "almost passing" group used in the latent portion of the Hybrid model. To determine whether the students identified by the latent portion of the Hybrid model were the same as those identified by the sub-score method, Cohen's (1960) κ of agreement was calculated. It was hypothesized that the Hybrid model would identify different students from those identified by the sub-score method.

The practical value of using this model was determined by the percentage of the sample described by the latent portion of the model. A minimum of 25%, the percent obtained by Gitomer & Yamamoto (1991), of the sample was considered educationally significant.

<u>Judgment of Instructional Relevance</u>

The results of the Hybrid model analyses of the content and cognitive processes areas were summarized and disseminated to the teachers. They were provided with a packet that included a) a description of each of the latent classes as defined by the test items and the percent of the items comprising that class; b) a class summary listing the names (fictitious) of the students in each latent class and c) an evaluation form (Appendix I, parts a-c).

The teachers were asked to evaluate the instructional relevance of these classes. They determined instructional relevance by evaluating 1) whether the topics characterizing each latent class could be taught as individual units; and 2) if they judged these units to be instructionally relevant. The teachers completed a short five-point Likert type evaluation form (ranging from "Strongly Agree" to "Strongly Disagree", where "Neutral" = 3) for both the content and cognitive processes latent classes (see Appendix I, part c) where a rating of 3.75 or higher was considered favorable. The teachers were asked to provide written examples of instructional activities based on the information provided and how the information would impact on their teaching practices.

RESULTS

Statistical Analyses

Item Classification

The analysis of the RCT mathematics first required the classification of the items. As indicated in the Methods section, the items were classified into the seven mathematical content and five cognitive processes areas (see Appendix J for the classification of the individual items). A summary of the results of the expert classification of the forty RCT are shown in Table 1. There are fewer items (10% versus 25%) dealing with rational numbers than given in the RCT specifications (see Appendix B); consequently there are more items in other areas (i.e. measurement; ratio, proportion and percent, consumer and job-related). In the cognitive areas, for which there are no RCT guidelines, the items are predominantly skill and knowledge.

Table 1.

Expert Classification of Items on the RCT Mathematics, in percent

Area	*% of Exam*
Mathematical Content	
1. Integers	25%
2. Rational Numbers	10%

(table continues)

Table 1. *(continued)*

Area	% of Exam
Mathematical Content	
3. Graphing	10%
4. Measurement	18%
5. Ratio, Proportion & Percent	15%
6. Probability & Statistics	2%
7. Consumer & Job-Related	20%
Cognitive Processes	% of Exam
1. Knowledge	22%
2. Skill	38%
3. Routine Application	15%
4. Understanding & Comprehension	10%
5. Problem Solving	15%

Latent Class Identification

Identification of the latent classes was governed by a theory driven, exploratory method. Four separate analysis were conducted. using the Hybil program to identify the ideal latent classes.

Firstly, the latent classes were constructed to reflect the mathematical content or the cognitive process characteristics of the items. For the content areas, twenty-eight latent classes were hypothesized; seven for each of the singular content areas and twenty-one for the combination of two classes. (The total number of two combination classes is obtained by multiplying seven by six and dividing by two.)

For the cognitive process areas, fifteen latent classes were hypothesized; five for each of the singular cognitive areas and ten for the combinations classes ((5 x 4)/2).

Secondly, in addition to the entire range of total test scores, latent classes were created to reflect the response patterns of those students in a restricted range of total test scores (i.e. a restricted group of students who were close to passing.) The restricted range of scores and the total range of scores are referred to as the restricted group and the unrestricted group respectively.

This classification system resulted in four groups: a restricted content, a restricted cognitive, an unrestricted content and an unrestricted cognitive analysis. The results of the four analyses showing the proportion of subjects in each of the latent classes are shown in Table 2. (Note: because there are no middle difficulty graphing items in the restricted class the proportion observed for this class is .0000. Therefore, any class combined with graphing will have the same proportion as the pure class.)

Table 2.

Proportion of Subjects in Latent Classes

Class	*Restricted*	*Unrestricted*
Mathematical Content		
Integers (I)	.00191	.00000
Rational Numbers (RN)	.00086	.00000
Graphing (G)	.00000	.00511
Measurement (M)	.00268	.00001
Ratio, Proportion & Percent (R)	.00039	.00064
Probability & Statistics (P)	.02079*	.03014*
Consumer & Job-Related (C)	.00366	.00214
IRN	.01262*	.00311
IG	.00191	.00141
IM	.02578*	.00497
IR	.00021	.00000
IP	.00575	.00000
IC	.00666	.00000
RNG	.00086	.00000
RNM	.00334	.00000
RNR	.00003	.00000
RNP	.00012	.00000
RNV	.00261	.03718*
GM	.00268	.00000

(table continues)

Table 2. *(continued)*

Class	Restricted	Unrestricted
	Mathematical Content	
GR	.00039	.01519*
GP	.02079*	.00555
GC	.00366	.00077
MR	.00520	.00328
MP	.01148*	.00168
MC	.00135	.00084
RP	.00502	.00000
RC	.02055*	.00004
PC	.00891	.00313
	Cognitive Processes	
Knowledge (K)	.00706	.00005
Skill (S)	.00453	.00135
Routine Application (RA)	.00393	.04978*
Understanding & Comprehension (U)	.00554	.00113
Problem Solving (PS)	.00110	.00290
KS	.01858*	.00063
KRA	.00590	.00061
KU	.02477*	.00000

(table continues)

Table 2. *(continued)*

Class	*Restricted*	*Unrestricted*
	Cognitive Processes	
KPS	.00312	.00335
SRA	.00904	.00125
RAU	.00327	.00473
RAPS	.00505	.00247
UPS	.01434*	.00059

* indicates educationally significant number of students in class (.01 or greater)

For each of the analyses, only those latent classes where the proportion of subjects in the latent class was .01 or higher were included in the subsequent analysis. For this data set, one percent of the sample represents sixteen students, which is the State-mandated register for a remediation class. In the restricted analysis, six of the content classes and four of the cognitive classes were included in the final Hybrid model. For the unrestricted analysis, three of the content classes and one of the cognitive classes were included in the final analysis. The resulting fourteen latent classes that were used in the Hybrid model, and the corresponding ideal response patterns, are shown in Table 3. (Note: For the restricted classes, the easiest items, 1,2,3,4,6, 7,8,9,10,11,13,14,16,17,19 and 24 are coded 1 and the most difficult items, 12,30,31, 32,33,34,35,36,37,38,39 and 40 are coded 0)

Table 3.

Hybrid Latent Classes and Response Patterns

Class	Ideal Response Pattern
Restricted	
1.Probability & Statistics	1111111111011111111110111111000000000000
2.Integers/Rational Numbers	1111111111011011111011110001000000000000
3.Integers/Measurement	1111011111011011110011110101000000000000
4.Graphing/Probability & Statistics	1111111111011111111110111111000000000000
5.Measurement/Probability & Statistics	1111011111011111110110111111000000000000
6.Ratio,Proportion & Percent/ Consumer & Job-Related	1111111111011111011101101110000000000000
7.Knowledge/Skill	1111011111011111010011110011000000000000
8.Knowlege/Understanding & Comprehension	1111011111011111010011111000000000000000
9.Skill/Understanding & Comprehension	1111111111011111111111101000000000000000
10.Problem Solving / Understanding & Comprehension	1111111111011111111111011000000000000000
Unrestricted	
11.Probability & Statistics	1111111111111111111111011111111111111111
12.Rational Numbers/ Consumer & Job-Related	1110101111101101111110101101001110011101
13.Graphing/Ratio,Proportion & Percent	0101110100110111000111111111111101111111
14.Routine Application	1111111110111011111100011111111111110111

The two parameter model with 14 latent classes model (-2*log-likelihood = 68799.455) fit the data significantly better compared to the two parameter IRT model (-2log-likelihood = 69570.104), χ^2 diff(42) = 770.6.

The ideal latent class response patterns modeled 18 % of the data. For the present sample this amounts to almost 300 students. In educational terms this would equal 15 classes or three teacher programs of five classes each.

The proportion of subjects in each latent class is presented in Table 4. Class 5 (measurement with probability and statistics (restricted)) describes the largest proportion of the sample, .041. Classes 6, 3 and 9 (ratio, proportion and percent with consumer and job-related (restricted), integers with measurement (restricted), skill with understanding and comprehension (restricted)) describe .029, .025, .021 respectively. Classes 8, 13 and 11 (knowledge with understanding and comprehension (restricted), graphing with ratio, proportion and percent (unrestricted), probability and statistics (unrestricted)) account for .016, .013, .010 respectively. The remaining seven classes collectively describe .03 of the sample.

Table 4.

Hybrid Latent Classes and Percents

Class	Proportion
Restricted	
1. Probability and Statistics	.000
2. Integers/Rational Numbers	.009
3. Integers/Measurement	.025
4. Graphing/Probability and Statistics	.000
5. Measurement/Probability and Statistics	.041
6. Ratio, Proportion and Percent/Consumer and Job-Related	.029
7. Knowledge/Skill	.007
8. Knowledge/Understanding and Comprehension	.016
9. Skill/Understanding and Comprehension	.021
10. Problem Solving/Understanding and Comprehension	.002
Unrestricted	
11. Probability and Statistics	.010
12. Rational Numbers/Consumer and Job-Related	.001
13. Graphing/Ratio, Proportion and Percent	.013
14. Routine Application	.007

In order to determine if the same information would have been obtained using simpler methods, Cohen's (1960) coefficient of agreement (κ), which determines the degree to which non-chance factors are the source of agreement between measures, was calculated. As a method for comparison, sub-scores were calculated for each subject. The sub-scores were calculated for the 14 latent classes identified in Table 4 according to the expert item classifications (see Appendix J). Four-hundred-twenty-six subjects were identified by either the latent portion of the Hybrid model (N = 288) or by the sub-score method (N = 174); only thirty-six subjects were identified by both methods. A 15 x 15 table was constructed (14 categories representing the latent classes used in the Hybrid model and 1 category for those students who were identified by one method but not the other). The results indicate that the two methods do not identify the same students ($\kappa = -.220$). It is estimated with 95% confidence that the value for κ is between -0.244 and -0.196. The significance of κ is determined by calculating σ_κ and dividing to determine a z score. The value for κ is significant in the direction of disagreement ($z = -8.677$).

Cohen's coefficient of agreement was also calculated for the thirty-six subjects who were identified by both methods resulting in a 14 x 14 table. For these students the methods do identify the same students ($\kappa = .471$). It is estimated with 95% confidence that the value for κ is between .278 and .664. The value for κ is significant in the direction of agreement ($z = 6.50$).

Judgment of Instructional Relevance

The results of the five point Likert inventory indicated that the nine teachers responded favorably (M = 3.90, SD = .037) in their assessment of the instructional relevance of the information provided by the Hybrid analysis.

The following is a summary of the teachers responses to the five open ended questions regarding their opinions of the information provided by the Hybrid model (see Appendix I, part c).

In response to the first question, "What instructional activities would you use based on this information ?" five of the teachers indicated that the information would be useful with cooperative learning and group work. They felt the information would allow them to allocate their time better. They also believed they would be able to concentrate on certain topics and include lessons that related to the daily lives of their students. They further indicated that the information provided would lend itself to using the buddy system . For example, if a teacher is aware that a student is weak in graphing but strong in integers, that student would be paired with a "buddy" who is weak in integers but strong in graphing. In this way each student gets to tutor and be tutored by the other. Furthermore, by being aware of their student weaknesses the teacher would be able to plan her lessons around the use of manipulatives.

In response to the second question, "For what instructional strategies do you think the information provided is best suited ?" these experts suggested cooperative learning (a technique where a group of students, all given specific "jobs", work together to complete a common task), the spiraling of lessons (a technique where each new lesson incorporates some part of a previous lesson), grouping, concept planning and personal tutoring as activities.

Regarding the question, "How could the information provided affect your time management ?" the teachers indicated it would allow them to plan their groups better. They explained that once the teacher is aware of her students' strengths and weaknesses she can accordingly allocate the time spent on each unit. More time could be spent on students' unique areas of difficulty and those areas that were crucial to exam success. Furthermore, students could also be paired more efficiently and would be able to help one another and to talk out their questions.

To the fourth question, "How could the information provided change your view of your students ?" the teachers felt that the information would give them better insight into why their students could not answer certain questions. It contributed to their understanding of the nature of their students' weaknesses (i.e. measurement, problem solving etc.). The teachers felt they would be better able to help students and to give them a sense of confidence that is crucial for exam success.

Lastly, to the question, "How would this information impact on your current teaching practices?" the teachers reiterated that they would do more cooperative learning and concept planning, topic concentration, real life application and mastery learning. They would do less theory and lecturing. They would do about the same spiraling and group work.

DISCUSSION

Research in the measurement field has shifted its view of the learner as a passive absorber of information to recognize the learner as active in the construction of his or her own learning. Proper assessment, according to this view will require new and imaginative tests items as well as statistical models that permit inferences about students understandings (Masters & Mislevy, 1989). However, such measures will take time in making their way into the education system. "If assessment and instruction can be integrated, education will improve."(Snow & Mandinach, 1991, p.1) Consequently, to explore the possibility of filling the gap in the interim, the present study was undertaken. The purpose of this study was to determine if newly developed statistical techniques would allow a test user to extract instructionally relevant information from in-use tests. To achieve this goal it was necessary to chose a statistical technique capable of extracting the desired information and an appropriate measure capable of producing the information, and to identify a population in need. Additionally, the technique had to also produce statistically significant results; and perhaps, more importantly, the information must be of value to educators.

It was postulated that the Hybrid model would provide a better fit for the RCT data than an IRT model alone. The results of the present study support the hypothesis that the Hybrid model improves the diagnostic assessment of student responses on the RCT mathematics by providing a better fit to the data than the IRT model alone.

To identify the final classes used in the latent portion of the Hybrid model, four separate Hybil program analyses were run. Eighty-six possible classes were explored (twenty-eight in the mathematical content and fifteen in the cognitive processes areas for both the restricted and the unrestricted groups.) Of the fourteen classes used in the latent portion of Hybrid analyses, ten were classes that describe the restricted (i.e. close to passing) group. From an educational standpoint this information is significant. These classes were constructed to reflect the ideal response patterns of students who were close to passing and deficient in an identifiable area. It is this group that is most in need of remediation and for which an understanding of their unique misconceptions would be most useful for an educator in terms of planning instruction. It could also indicate that something unusual is happening in the "almost passing" group. These children may have identifiable gaps in their understanding that could be remediated if they were identified, .i.e., not all failures are equal. At present, there is no distinction among students placed in a remediation class. Students who scored in the high fifties are in the same classes with those student who scored a ten. This grouping system creates a difficult instructional task for the teacher. Furthermore, diagnostic information specific to individual students was judged useful by the teachers who participated in the study.

The inclusion of the latent portion of the Hybrid model describes 18% of the data. For the present sample, this percentage translates to approximately 300 students. These students all have identifiable response patterns useful for instructional planning.

In the restricted content area, 4.1% of the students were identified by a latent class in measurement with probability and statistics (class 5). An additional 2.9% were identified in the ratio, proportion and percent with consumer and job-related class (class 6). Lastly, 2.5% were identified by a class in integers with measurement (class 3). In class 6, about half of the items that comprise the consumer and job-related area require ratio, proportion or percent strategies for their solution. Measurement appears as a remediation area in classes 5 and 3. This may be due to the students' lack of familiarity with the instruments used for measurement, instruments such as a ruler or a protractor.

In the restricted cognitive areas, 2.1% were placed in the skill with understanding and comprehension class (class 9) and 1.6% were placed in the knowledge/understanding and comprehension class (class 8). This is not surprising as it seems reasonable to conclude that understanding and comprehension items or what might be considered "deeper" cognition, are difficult for less knowledgeable math students.

For the unrestricted content analysis 1.3% of the subjects were in need of remediation in graphing with ratio, proportion and percent (class 13) and 1.0% in probability and statistics (class 11). When reviewing the items that comprise class 13, the emphasis on pictorial representations becomes apparent. Many of the items contain charts or figures with which the students may have been unfamiliar or which held little interest for them. Although there is only one item in class 11, and no serious conclusions can be made, the students falling into this class are getting almost every other item correct. The educators involved in this study hypothesized that (based on their experience) it is students who have recently immigrated that have the most difficulty with this class because they are not exposed to probability and statistics in their country.

The results of calculating Cohen's coefficient of agreement indicate that the Hybrid model and the sub-score method do not identify the same students in the latent classes. In fact, the results indicate that the disagreement between the two methods is highly significant. The disagreement between measures is attributable to the manner in which the two methods determine class placement.

For a student to be placed into a latent class by the sub-score method, he or she must get fewer than 65% of the items in that particular class correct. For example, if a student were identified as belonging to the Integers/Rational Numbers latent class then that student must fail in the integer and rational numbers areas only. According to the sub-score method, if the student also failed in another area (i.e. graphing), this student is not eligible for placement into the Integers/Rational Numbers latent class. In the sub-score method, if the student barely failed in that area, i.e. answered 3 of the 7 graphing items incorrectly, he or she would still be ineligible for the Integers/Rational Numbers latent class. In the Hybrid model, this same "barely failing" student's response pattern would be very similar to the ideal response pattern for the Integers/Rational Numbers latent class. However, the Hybrid model would place this student into the Integers/Rational Numbers latent class. Thus, the Hybrid method allows for more variation in the student's response pattern. Therefore, more students are classified into the latent classes by the Hybrid model. This is borne out by the statistics where, of those students not jointly classified, 65% were identified by the Hybrid model and 35% by the sub-score method.

The practical value of using the Hybrid model is determined by the percentage of the sample the latent portion describes. In this study, the latent class portion of the Hybrid model described 18% of the sample. This number fell short of the hypothesized percent required for significance. Although the 25% obtained by Gitomer & Yamamoto (1991) and required for educational significance was not realized in this study some findings are worth considering.

On a practical level, the 18% realized in this study equals approximately 300 students. This number amounts to three full teacher programs. In New York City high schools, interest expressed by as few as fifteen students is enough to program that class. A teacher is then required to develop a course of study for as few as fifteen students. Consequently, 300 students in need would indicate a justifiable mandate. Perhaps even more significant than the numbers, is that the Hybrid model provides a diagnostic profile of the learner for which an educational program could be devised. Furthermore, this information could be used in a variety of educational settings such as peer tutoring, computer assisted instruction and tutorials.

The RCT mathematics is the instrument used to determine remediation programming. However, with this special programming, there are no clear accompanying guidelines on how to approach the remediation problem. The results of the Hybrid model could be such a source of information. Particularly significant is that this information can be created by teachers for teacher use. Teachers could be trained in the processes of translating their teaching experience into formalized measurement practice. Actively involving teachers and utilizing real-life experience would contribute to the meaningfulness of a test score.

As originally stated, the intent of this study was to determine if instructionally significant information could be obtained from in-use tests, by applying newly developed statistical techniques and developing a model for such a procedure. Although the weaknesses of the RCT mathematics as a testing instrument were initially apparent, for the purposes of this study it was the best measure available. However, the measure is problematic due to the unbalanced distribution of items in the content area (i.e. one statistic item) and cognitive areas (60% either knowledge or skill items). The RCT is also extremely simple in its content and scope. It has been said that these minimum competency tests are more aptly described as the "most" minimum imaginable competency test (Popham, 1993). Be that as it may, it is these measures that are used to determine graduation and it is these students who do not pass this exam who are most in need of further instruction. The need to bridge the gap between testing and planning instruction is most acute in these remediation programs.

The shortcomings of the measure aside, not only does the Hybrid model improve the fit of a purely continuous model, it also provides additional information, information that is judged valuable by educators. As indicted in their responses, the teachers felt they could use the information to plan their instructional activities. The teachers reported that they could be prepared for their lessons by locating and choosing appropriate manipulatives. Additionally, by being aware of their students' strengths and weaknesses they could more appropriately pair certain students. They could plan their group lessons to specialize in the areas containing a high percentage of students. Special need areas could be reinforced in the "Do now" (a short problem placed on the board at the beginning of every period), in home work, and by exposing the students to real life applications.

They also reported that the information would allow them to understand their students better. For example, one teacher indicated that "by reviewing the information I might be able to determine what the students truly don't understand or if they just answer without caring." Finally, the teachers reported the information would affect how they allocated their time and in determining the topics they would specialize in.

What is promising about the procedure modeled is its flexibility. In this study, the basis for the classes explored is dictated by the test blueprint (content) and by educational literature (cognition). The test users are not restricted to classifying items by these characteristics, but could use any that they feel might give them insights into their students. For example, a teacher might suspect that certain questions are somewhat gender biased (i.e. Leslie earns $2.50 babysitting...etc.) and create a a latent class response pattern accordingly. If more complex measures are used then more complex information can be obtained.

Future research might include applying the Hybrid model to a variety of in-use tests. The procedure modeled could be applied to a number of exams in different content areas (Regents examinations , Pre-Scholastic Aptitude Test (PSAT), Scholastic Aptitude Test (SAT), achievement tests, etc.). As the level of the complexity of the measurement instrument increases, so would the information obtainable from the analysis of the exam. An examination, such as the English Regents, could be productively analyzed by a variety of criteria (reading comprehension, vocabulary, writing skills, analogies, inference items etc.) to obtain more meaningful scores.

The New York State Regents is a comprehensive examination in terms of both the topics covered and the variety of the questioning techniques. Students answer free response and multiple choice questions. This examination is distinguished by the fact that students are required to answer questions where the nature of their approach is evaluated. Application of the Hybrid model to this examination and examinations of this type (i.e. Regents exam in other subject areas) could provide instructionally useful information. By applying the Hybrid model to this examination, educators could obtain useful instructional information for teaching those students who do not pass the exam.

To assess whether the Hybrid model improves the instructional program of students with remediation needs, a comparison could be made of the passing percentages of a traditional remediation program and the passing percentages of the students identified and taught as a result of the Hybrid assessment.

The application of this approach would be particularly significant because of the current Math and Science initiative. As a result of this initiative all students, students who had previously been programmed for non Regents courses, are required to take and pass more rigorous academic math and science classes. In addition to passing the RCT's in these areas, students are also required to pass the Course 1 Regents (first year mathematics) in order to graduate from high school. As a result of this initiative the number of students requiring remediation has increased. It is apparent that there exists a large group of students who need additional assistance to pass this examination. The Hybrid model seems an ideal method for identifying the needs of these students with respect to the Regents exam required by the State.

In a system where students are constantly tested, the procedure modeled provides additional instructional information without unnecessary intrusion on the test taker or the test user.

Appendix A

Teacher Information Form

You have been asked to participate in a study designed to model a procedure for obtaining additional information from Regents Competency exams. The purpose of the study is to obtain information designed for classroom use to aid in instructional planning. As a teacher myself I realize that your time is limited and I thank you in advance for you assistance and participation in this study

Laura Alvarez

Date_______________ School______________________
Name____________________________________
License:___________(grade level)_______________(subject)
Number of years teaching high school mathematics____________
Number of years teaching Regents mathematics________________
Number of years teaching remedial mathematics_______________

If you would like a summary of the study when it is completed, please give your mailing address below.

____
____
____

Appendix B

The University of the State of New York
THE STATE EDUCATION DEPARTMENT
Division of Educational Testing
Albany, New York 12234

REGENTS COMPETENCY TEST IN MATHEMATICS

Test Description and Blueprint

The Regents competency test in mathematics is designed to measure competence in the skills described in the first seven units of the Education Department's publication, General High School Mathematics. This curriculum is recommended for one of the two required years of high school mathematics for students not enrolled in Regents mathematics courses.

Students may use as much time as they need to complete the test, which contains 60 questions. The passing grade is 39 questions answered correctly, or 65%.

The questions on the test are based on the topics listed on pages 5-8 of General High School Mathematics. These topics are summarized below.

- operations of addition, subtraction, multiplication, and division with whole numbers, fractions, decimals, including positive and negative numbers
- the comparison of numbers by ordering and by ratio
- the properties of numbers rising out of the consideration of divisibility of whole numbers, namely, primes, composite numbers, factors, divisors, factorization, greatest common factor, and least common multiple
- measurement, both direct, using the metric system, and indirect, using formulas to find perimeter, circumference, area, and the Pythagorean relationship
- proportions
- percent
- graphs, which include analysis and interpretation of common graphs, number lines, vertical and horizontal, and graphing ordered pairs
- probability, and statistics involving mean, median, and mode
- problem solving in job-related and consumer applications
- solution of simple equations

(over)

The test blueprint below shows the approximate percentage of questions for each unit. The percentages are approximate because items can be categorized appropriately in more than one unit.

Unit	Approximate %
I. Set of Integers	25%
II. Rational Numbers	25
III. Graphing	8
IV. Measurement of Geometric Figures	12
V. Ratio, Proportion, and Percent	10
VI. Probability and Statistics	5
VII. Consumer and Job-Related Mathematics	15
	100

Appendix C
Summary of Rasch Analysis

```
1 RCT DATA 1989                                  "BIGSCALE" RASCH ANALYSIS  VER. 1.73  Nov  6  15:37:29  1990
 INPUT:    69 PERSONS    60 ITEMS                ANALYZED:   69 PERSONS   60 ITEMS   2 CATEGORIES      TABLE 12
```

ITEMS STATISTICS -- ENTRY ORDER

NUM	NAME	COUNT	SAMPLE	CALIBRTN	ERROR	MNSQ	INFIT	MNSQ	OUTFIT
1	1= ADDWHOLE	58	69	-1.45	.33	1.1	.6	1.4	1.3
2	2= PIGRAPH	63	69	-2.16	.43	1.0	.0	.8	-.3
3	3= NUMERAL	58	69	-1.45	.33	1.1	.5	1.3	.9
4	4= ADDDEC	48	69	-.57	.27	1.1	.7	1.1	.5
5	5= SUBWHOLE	52	69	-.88	.29	1.0	.1	1.0	.2
6	6= SUBDEC	44	69	-.29	.26	1.0	.5	1.0	.3
7	7= IMPFRA	32	69	.48	.25	.8	-2.5	.8	-2.0
8	8= ROUND	56	69	-1.24	.31	.9	-.6	.7	-1.0
9	9= MULTWHOL	49	69	-.64	.27	1.0	-.2	.9	-.5
10	10=MULTDEC	40	69	-.03	.25	.9	-.7	.9	-.7
11	11=MULTFRAC	51	69	-.80	.28	1.1	.9	1.3	1.5
12	12=VAREQU	42	69	-.16	.25	.8	-2.2	.8	-2.1
13	13=DIVREM	52	69	-.88	.29	1.0	.0	1.0	.2
14	14=DIVDEC	34	69	.36	.25	.9	-.9	.9	-1.0
15	15=DIVSIGN	39	69	.04	.25	1.1	1.4	1.1	1.3
16	16=DIVWHOLE	39	69	.04	.25	1.1	.7	1.0	.5
17	17=AREAUND	28	69	.74	.25	1.0	.3	1.0	.3
18	18=DECTOFRA	5	69	3.05	.47	1.0	.2	1.3	.7
19	19=AREATRI	15	69	1.72	.38	.9	-.4	.9	-.5
20	20=ADDSIGN	9	69	2.37	.36	1.0	.0	.8	-.6
21	21=METRIC	66	69	-2.91	.59	1.0	.1	.7	-.3
22	22=PERWHOL	64	69	-2.36	.46	1.0	.1	1.0	.2
23	23=MAPSCALE	64	69	-2.36	.46	.9	-.2	.5	-.9
24	24=PICGRAPH	53	69	-.97	.29	1.0	-.1	1.3	1.3
25	25=DIGPLACE	45	69	-.36	.26	1.2	1.9	1.4	2.0
26	26=DIVFACT	51	69	-.80	.28	1.0	.0	1.0	.3
27	27=BANKWRD	43	69	-.22	.26	1.0	-.5	.9	-.8
28	28=FACTORS	42	69	-.16	.25	1.1	1.2	1.1	1.0
29	29=DECTOPERC	44	69	-.29	.26	1.1	1.2	1.1	.0
30	30=NUMLINE	37	69	.17	.25	.9	-1.0	.9	-.9
31	31=CHANGEWRD	40	69	-.03	.25	.9	-.8	.9	-.8
32	32=SQURT	52	69	-.88	.29	.9	-.7	.8	-.9
33	33=BUDGETWRD	48	69	-.57	.27	.9	-.5	.9	-.5
34	34=CIRTERMS	36	69	.23	.25	1.1	1.2	1.1	1.1
35	35=PURCHWRD	50	69	-.72	.28	1.0	-.1	.9	-.6
36	36=LCD	36	69	.23	.25	.9	-1.6	.9	-1.6
37	37=EXP	49	69	-.64	.27	.9	-.8	.8	-1.2
38	38=RATESPWRD	41	69	-.09	.25	1.0	.5	1.0	.2
39	39=GRAPH	31	69	.55	.25	1.0	-.3	1.0	-.5
40	40=PROBWRD	46	69	-.43	.26	.9	-1.2	.8	-1.1
41	41=PERCENT	38	69	.10	.25	1.1	1.2	1.1	1.3
42	42=ADDFRAC	32	69	.48	.25	1.0	.3	1.1	.0
43	43=DIVWRD	38	69	.10	.25	1.0	-.3	1.0	-.2

1"CONTINUED FROM PREVIOUS SHEET"

ITEMS STATISTICS -- ENTRY ORDER

NUM	NAME	COUNT	SAMPLE	CALIBRTN	ERROR	MNSQ	INFIT	MNSQ	OUTFIT
44	44-CIRCUMFRD	31	69	.55	.25	1.0	.1	1.0	-.2
45	45-MEDIAN	37	69	.17	.25	1.0	.0	1.0	-.1
46	46-CHART	33	69	.42	.25	1.0	.0	1.0	-.2
47	47-PRIME	33	69	.42	.25	1.1	1.7	1.1	1.4
48	48-TAXWRD	36	69	.23	.25	1.1	1.0	1.0	.6
49	49-RATIOWRD	20	69	1.30	.27	.9	-.6	.9	-.5
50	50-BARGRAPH	43	69	-.22	.26	1.0	.3	1.1	.6
51	51-TIMEWRD	32	69	.48	.25	1.0	-.6	.9	-.7
52	52-UNITCOST	29	69	.68	.25	1.0	-.5	1.0	-.1
53	53-PERCBUNIS	22	69	1.15	.27	1.1	1.1	1.2	1.5
54	54-COMPDEC	22	69	1.15	.27	1.1	.5	1.1	.9
55	55-CROSSEQU	19	69	1.38	.28	.9	-1.0	.8	-1.2
56	56-RULER	21	69	1.23	.27	1.1	1.1	1.1	1.0
57	57-TRIANBAS	38	69	.10	.25	1.0	.1	1.0	.2
58	58-TRDTRANSL	21	69	1.23	.27	.9	-.4	1.0	-.1
59	59-PYTHAG	15	69	1.72	.30	1.0	.3	1.1	.4
60	60-SUBNIT	15	69	1.72	.30	1.0	.2	1.1	.4

```
1 RCT DATA 1989                                    'BIGSCALE' RASCH ANALYSIS  VER. 1.73  Nov  6  15:37:29  1990
INPUT:   69 PERSONS   60 ITEMS               ANALYZED:   69 PERSONS   60 ITEMS   2 CATEGORIES        TABLE 6
-----------------------------------------------------------------------------------------------------------

                                        MAP OF ITEMS
     ------------------------------------------------------------------------------------------------------
    4.0
     |
     |
     |
     |
    3.0 18-DECTOFR
     |
     |
     |  20-ADDSIGN
Q    |
    2.0
     |  19-AREATRI  59-PYTHAG   60-SUBMII
     |
     |  55-CROSSEQ  49-RATIONR
S    | .56-RULER    58-WRDTRAN  53-PERCNUM  54-COMPDEC
    1.0
     |  17-AREAWRD
     |  52-UNITCOS  39-GRAPH    44-CIRCUMF
     |  7- INTRFRA  42-ADDFRAC  51-TIMEWRD  46-CHART    47-PRIME    14-DIVDEC
     |  34-CIRTERM  36-LCD      48-TAXWRD   30-NUMLINE  45-MEDIAN   41-PERCENT  43-DIVWRD   57-TRIAREA
M    .0 15-DIVSIGN  16-DIVWHOL  10-MULTDEC  31-CHANGEW  38-RATESPW
     |  12-VARSQU   28-FACTORS  27-BANKWRD  50-BARGRAP  6- SUBDEC   29-DECTOPE
     |  25-DIGPLAC  40-PROBWRD
     |  4- ADDDEC   33-BUDGETW  9- MULTWNO  37-EIP
     |  35-PURCHWR  11-MULTFRA  26-DIVFACT  5- SUBWHOL  13-DIVREM   32-SQURT
   -1.0 24-PICGRAP
S    |  8- ROUND
     |  1- ADDWHOL  3- NUMERAL
     |
     |
   -2.0
Q    |  2- PICRAPH
     |  22-PERWHOL  23-MAPSCAL
     |
     |
   -3.0 21-METRIC
     |
     |
     |
     |
   -4.0
     ------------------------------------------------------------------------------------------------------

1 RCT DATA 1989                                    'BIGSCALE' RASCH ANALYSIS  VER. 1.73  Nov  6  15:37:29  1990
```

Appendix D

DESCRIPTION AND EXAMPLES OF ITEM CONTENT

1. Set of Integers. Problems whose components and solution involve an integer.

Ex. What is the sum of +23 and -19?

(1) -4	(3) +42
(2) +4	(4) -42

2. Rational Numbers. Problems whose components or solution involve a rational number.

Ex. When listed in order from smallest to largest, which decimal would come first?

(1) 1234	(3) 3249
(2) 0547	(4) 0468

3. Graphing. Problems which include analysis and interpretation of common graphs, number lines, vertical and horizontal, and graphing ordered pairs.

Ex. What is the length of the line segment joining points G and H on the graph below?

(1) 4	(3) 12
(2) 8	(4) 15

4. Measurement. Problems requiring both direct, using the metric system, and indirect, using formulas to find the perimeter, circumference, area, and the Pythagorean relationship.

Ex. If a side of an equilateral triangle is 7cm, what is the perimeter of the triangle?

(1) 7cm	(3) 21cm
(2) 14cm	(4) 9cm

5. Ratio, Proportions, and Percent. Problems requiring the use of ratio, proportion or percent to which there in no consumer related solution.

Ex. 10% of 35 is what number?

(1) 35	(3) 350
(2) 3.5	(4) .35

6. Probability, and Statistics. Problems involving probability, mean, median, or mode.

Ex. What is the mode for the following four test scores: 70, 60, 90, 60?

(1) 70	(3) 60
(2) 120	(4) 90

7. Problem solving in job-related and consumer applications.

Ex. John earns $2.25 an hour. If he works 10 hours, how much will he earn?

(1) $225	(3) $20.50
(2) $2.25	(4) $22.50

Appendix E

DEFINITIONS AND EXAMPLES OF COGNITIVE PROCESSES(NCTM)

Knowledge - This item requires the recall or recognition of mathematical ideas, figures, or symbols. Knowledge items include such tasks as identifying common geometric figures or recalling a number fact.

Ex. Which of the following is a prime number?

(1) 13	(3) 21
(2) 15	(4) 4

Skill - This item requires routine mathematical manipulations that have been learned or practiced. This item does not require a student to decide what operation to use. It assesses proficiency in using an algorithm rather than the understanding of how it works. Such items might require the student to make a measurement, multiply fractions, solve an equation, or read a table

Ex. Solve for x   3/4 = x/12

(1) 3	(3) 9
(2) 48	(4) 36

Routine Application - This item involves knowledge or skill with which the student is presumed to have had experience. Although the mathematical operation is not specified, the identification of the proper procedure is "almost" automatic.

Ex. Michael bought orange juice for $.89, a bagel with cream cheese for $.65 and a math book for $5.95.

How much did he pay for the items. (Assume there is no tax on any of the items)

(1) \$7.49 (3) \$7.39
(2) \$21.35 (4) 7.89

Understanding and Comprehension - This item assesses understanding focused on basic mathematical concepts and principles. It frequently requires student to identify or establish relationships between different representations, such as identifying which model represents a given number sentence.

Ex. Which mathematical sentence represents, five less than a number is 14.

(1) $x-5=14$ (3) $5x=14$
(2) $x-5>14$ (4) $x-14=5$

Problem Solving - This item requires higher-order thinking and involves the integration of concepts and skills to solve problems for which there is no clear method of solution. This is an item that can not be solved by a routine application of a concept or skill. No item on the RCT really fits this definition. For the purposes of this study an item which requires two steps for solution, other than a routine change problem will be considered "problem solving".

Ex. A stereo can be purchased with a downpayment of \$125 and 10 monthly installments of \$25 each. What will be the total cost of the stereo?

(1) \$250 (3) \$160
(2) \$375 (4) \$390

Appendix F
Classification Instructions and RCT January 1991 Items

Name ________________ School__________________

Date_________ RCT form January 1991

Classify the forty items in the accompanying RCT booklet based on the classifications discussed earlier. First, classify each item according to cognitive processes by placing an X in the column that best represents the item; second, classify each item by mathematical content. When you are finished we will discuss the classifications. Write down any comments you may have using the black of the paper if needed.

COGNITIVE PROCESSES

ITEM #	Knowledge	Skill	Routine Application	Understanding/ Comprehension	Problem Solving	Comments
21						
22						
23						
24						
25						
26						
27						
28						
29						
30						

	Knowledge	Skill	Routine Application	Understanding/ Comprehension	Problem Solving	Comments
31						
32						
33						
34						
35						
36						
37						
38						
39						
40						
41						
42						
43						
44						
45						
46						
47						
48						
49						
50						
51						
52						
53						
54						
55						
56						
57						

	Knowledge	Skill	Routine Application	Understanding/ Comprehension	Problem Solving	Comments
58						
59						
60						

Place an X in the column that best describes the items mathematical content.

Item Content

ITEM # / Set of Integers	Rational Numbers	Graphing	Measurement	Ratio, Proportion Percent	Probability Statistics	Consumer Job-Related
21						
22						
23						
24						
25						
26						
27						
28						
29						
30						
31						
32						
33						
34						
35						
36						
37						
38						
39						
40						
41						
42						
43						
44						
45						
46						
47						
48						
49						

	Set of Integers	Rational Numbers	Graphing	Measurement	Ratio, Proportion Percent	Probability Statistics	Consumer Job-Related
50							
51							
52							
53							
54							
55							
56							
57							
58							
59							
60							

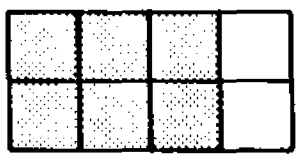
Part B

Answer all 40 questions in this part. Mark your answers in the rows of answer circles provided in PART B on the separate answer sheet. Use only a black lead pencil on the answer sheet.

21 Which expression represents 4 times a number, n, minus 3?

(a) $3n - 4$
(b) $3n + 4$
(c) $4n - 3$
(d) $4n + 3$

22 What percent of the rectangle below is shaded?

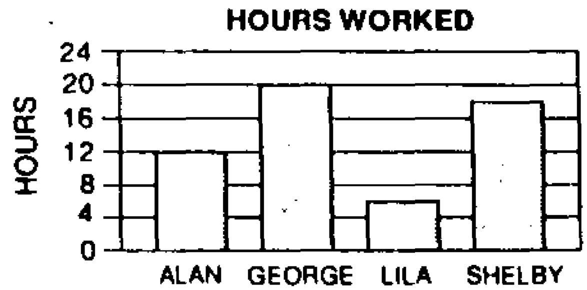


(a) 75%
(b) 2%
(c) 28%
(d) 25%

23 A bag contains 10 red marbles, 16 blue marbles, and 20 yellow marbles. A student picks one marble at random from the bag. What is the probability that a blue marble is picked?

(a) $\frac{10}{46}$
(b) $\frac{20}{46}$
(c) $\frac{30}{46}$
(d) $\frac{16}{46}$

24 The bar graph below shows the number of hours worked by four students in one week. How many more hours did George work than Shelby?



(a) 8
(b) 2
(c) 6
(d) 4

25 If tickets cost \$5 each, what is the *least* number of tickets that must be sold to collect \$198?

(a) 100
(b) 200
(c) 10
(d) 40

26 A case of soft drinks contains 12 cans and costs \$6.96. What is the price of one can?

(a) \$0.48
(b) \$0.58
(c) \$6.84
(d) \$7.08

27 Tony works 5 hours a day for four days during the week. If he is paid $4 an hour, what is his pay for the week?

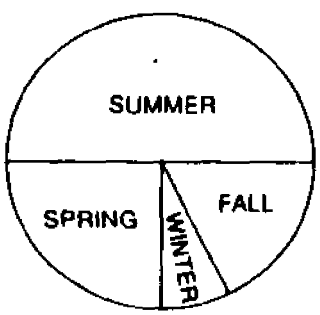
(a) $16
(b) $20
(c) $36
(d) $80

28 Solve for t: $\frac{2}{9} = \frac{10}{t}$

(a) 5
(b) 10
(c) 45
(d) 90

29 The circle graph below represents the amount of ice cream sold in each season of a year. Which two seasons have the *lowest* combined sales?

ICE CREAM SALES



(a) fall and winter
(b) summer and winter
(c) summer and spring
(d) spring and fall

30 Which is a measure of mass (weight)?

(a) gram
(b) liter
(c) meter
(d) Celsius degree

31 From 424.7 subtract 38.46.

(a) 40.1
(b) 382.34
(c) 382.36
(d) 386.24

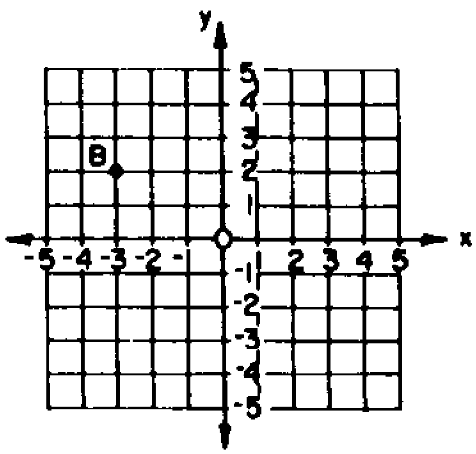
32 Which fraction is equal to $\frac{2}{3} \times \frac{5}{6}$?

(a) $\frac{4}{5}$
(b) $\frac{5}{9}$
(c) $\frac{7}{9}$
(d) $\frac{15}{12}$

33 What is the value of 4^3?

(a) 12
(b) 16
(c) 64
(d) 256

34 On the graph below, what are the coordinates of point B?



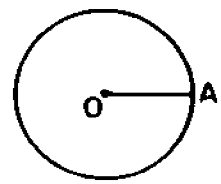
(a) (2,3)
(b) (2,–3)
(c) (–3,2)
(d) (3,2)

35 Which value of x makes the following sentence true?

$$-15 + x = 0$$

(a) 15
(b) –15
(c) $\frac{1}{15}$
(d) 0

36 If O is the center of the circle below, what is line segment OA called?



(a) an arc
(b) a diameter
(c) a chord
(d) a radius

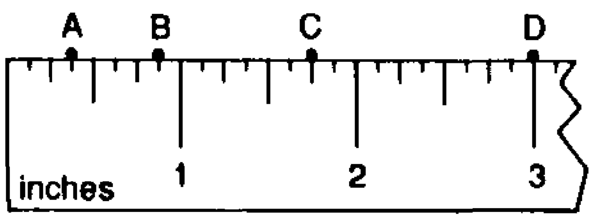
37 Andrea earns \$6.85 an hour for a 40-hour work week. If \$39 is withheld for Federal income tax and \$12.44 is withheld for Social Security tax, what is her weekly net pay?

(a) \$325.94
(b) \$247.94
(c) \$222.56
(d) \$222.06

38 Julian had a balance of \$600 in his bank account. He made a \$100 withdrawal and earned \$7.50 interest on his account. What is his new balance?

(a) \$492.50
(b) \$507.50
(c) \$592.50
(d) \$707.50

39 On the ruler below, which letter indicates $\frac{7}{8}$ inch?



(a) *A*
(b) *B*
(c) *C*
(d) *D*

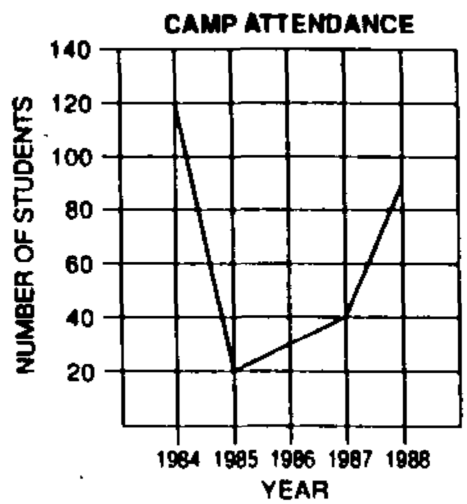
40 A stereo can be purchased with a downpayment of $125 and 10 monthly installments of $25 each. What will be the total cost of the stereo?

(a) $150
(b) $250
(c) $350
(d) $375

41 The cost of a telephone call is 40¢ for the first 3 minutes and 8¢ for each additional minute. What is the cost of a 9-minute call?

(a) 98¢
(b) 88¢
(c) 78¢
(d) 48¢

42 The line graph below shows the number of students attending a summer camp. How many more students attended the camp in 1988 than in 1986?



(a) 80
(b) 70
(c) 60
(d) 50

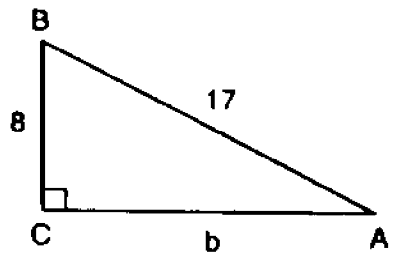
43 Which fraction has a value greater than 1?

(a) $\frac{4}{3}$
(b) $\frac{7}{8}$
(c) $\frac{9}{10}$
(d) $\frac{6}{11}$

44 What is the remainder when 8650 is divided by 23?

(a) 1
(b) 2
(c) 3
(d) 0

45 In the right triangle shown below, $a = 8$ and $c = 17$. Using the Pythagorean theorem, $c^2 = a^2 + b^2$, what is the value of b?



(a) 9
(b) $12\frac{1}{2}$
(c) 15
(d) 25

46 The sum of $\frac{1}{6}$ and $\frac{7}{9}$ is

(a) $\frac{17}{18}$
(b) $\frac{5}{15}$
(c) $\frac{8}{15}$
(d) $\frac{7}{54}$

47 Given the formula $Y = 3(2A + B)$, what is the value of Y when $A = 6$ and $B = 9$?

(a) 24
(b) 25
(c) 63
(d) 105

48 A plane leaves Albany, New York, at 7:30 p.m. and arrives in Washington, D.C., at 9:10 p.m. What is the total length of time of this flight?

(a) 1 hour 20 minutes
(b) 1 hour 40 minutes
(c) 2 hours 20 minutes
(d) 2 hours 40 minutes

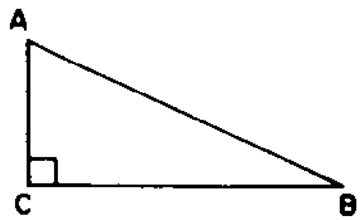
49 What is the perimeter of a rectangle with a length of 10 and a width of 4?

(a) 14
(b) 28
(c) 40
(d) 80

50 A camera was originally priced at \$64.80. If the camera is on sale for 25% off the original price, what is the sale price?

(a) \$16.20
(b) \$25.00
(c) \$39.80
(d) \$48.60

51 In right triangle *ABC* below, what is the measure of angle *A* if the measure of angle *B* is 25°?



(a) 50°
(b) 65°
(c) 115°
(d) 165°

52 At a basketball game, Ricky made 18 out of 36 baskets that he attempted. What percent of the baskets did Ricky make?

(a) 18%
(b) 20%
(c) 50%
(d) 65%

53 Sara jogged 3 miles one day and $4\frac{1}{2}$ miles on each of the next two days. How many miles did she average per day?

(a) $3\frac{3}{7}$
(b) $3\frac{3}{4}$
(c) $4\frac{1}{14}$
(d) 4

54 What is $\frac{5}{8}$ written as a decimal?

(a) 0.58
(b) 0.625
(c) 1.6
(d) 5.8

55 What is the median of the following test scores?

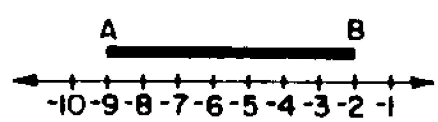
75, 60, 85, 60, 90

(a) 60
(b) 74
(c) 75
(d) 90

56 Which decimal has the *smallest* value?

(a) 0.07
(b) 0.08
(c) 0.006
(d) 0.004

57 What is the length of the line segment joining points *A* and *B* on the graph below?



(a) 7
(b) 8
(c) 9
(d) 11

58 Which is the closest estimate of $\sqrt{32}$?

(a) 64
(b) 16
(c) 6
(d) 5

59 Which is equal to 9?

(a) $(30 + 18) - 12 \div 4$
(b) $2 + 10 \div 2 + 3$
(c) $6 + 6 \div 2$
(d) $36 - 18 \div 2$

60 Which inequality is represented by the graph below?

(a) $x \geq -1$
(b) $x > -1$
(c) $x \leq 4$
(d) $x < 4$

Appendix G
Classifications Instructions and RCT June 1991 Items

Name ______________ School______________
Date ___________ RCT Form June 1991

Classification Instructions

Classify the forty items in the accompanying RCT booklet based on the classification instructions discussed earlier. First, classify each item according to cognitive processes by placing an X in the column that best represents the item: second, classify each item by mathematical content. After your classification is complete, the item classification sheets will be collected, xeroxed and returned to you. When the item sheets are returned, please discuss your analysis with your colleagues. When you have agreed upon a classification, complete the final classification sheet.

COGNITIVE PROCESSES

ITEM #	Knowledge	Skill	Routine Application	Understanding/ Comprehension	Problem Solving	Comments
21						
22						
23						
24						
25						
26						
27						
28						
29						
30						
31						
32						
33						
34						
35						
36						
37						
38						

	Knowledge	Skill	Routine Application	Understanding/ Comprehension	Problem Solving	Comments
39						
40						
41						
42						
43						
44						
45						
46						
47						
48						
49						
50						
51						
52						
53						
54						
55						
56						
57						
58						
59						
60						

Place an X in the column that best describes the items mathematical content.

Item Content

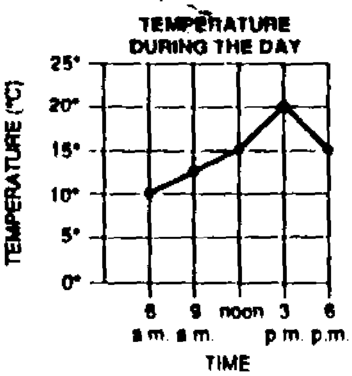
Set of Integers ITEM #	Rational Numbers	Graphing	Measurement	Ratio, Proportion Percent	Probability Statistics	Consumer Job-Related
21						
22						
23						
24						
25						
26						
27						
28						
29						
30						
31						
32						
33						
34						
35						
36						
37						
38						
39						
40						
41						
42						
43						
44						
45						
46						
47						
48						
49						

	Set of Integers	Rational Numbers	Graphing	Measurement	Ratio, Proportion Percent	Probability Statistics	Consumer Job-Related
50							
51							
52							
53							
54							
55							
56							
57							
58							
59							
60							

Part B

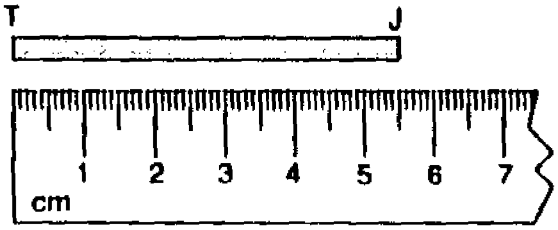
Answer all 40 questions in this part. Mark your answers in the rows of answer circles provided in PART B on the separate answer sheet. Use only a black lead pencil on the answer sheet.

21 The graph below shows the temperature for various times of day. At what time did the temperature reach 20°C?



(1) 6:00 a.m.
(2) noon
(3) 3:00 p.m.
(4) 6:00 p.m.

22 What is the length of the line segment $\overline{TJ}$ shown below?



(1) 55 cm
(2) 5.5 cm
(3) 0.55 cm
(4) 5 cm

23 The chart below shows the average monthly temperature for certain cities in the United States. What is the average monthly temperature for August in Miami, Florida?

AVERAGE MONTHLY TEMPERATURE (°F)

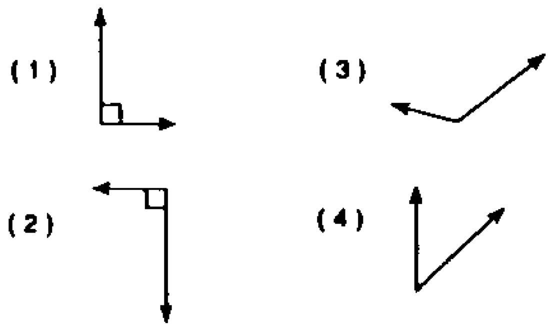
	Juneau, Alaska	Nashville, Tennessee	Miami, Florida	San Francisco, California
Jan	-11	42	68	50
Feb	0	44	68	53
Mar	9	53	71	54
Apr	29	62	74	56
May	47	70	77	57
June	58	78	80	59
July	60	81	82	59
Aug	55	80	82	59
Sept	44	74	81	62
Oct	27	63	78	61
Nov	3	52	72	57
Dec	-8	44	69	52

(1) 68
(2) 80
(3) 81
(4) 82

24 The fraction $\frac{19}{3}$ may be expressed as which mixed number?

(1) $5\frac{1}{3}$
(2) $6\frac{1}{3}$
(3) $7\frac{2}{3}$
(4) $16\frac{1}{3}$

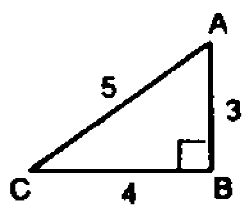
25 Which drawing shows an acute angle?



26 Jill has $282.15 in her checking account. If she writes checks for $25 and $46.75, how much money will be left in her account?

(1) $210.40 (3) $329.15
(2) $235.15 (4) $353.90

27 What is the ratio of *AB* to *BC* in the right triangle below?



(1) $\frac{3}{5}$ (3) $\frac{4}{5}$

(2) $\frac{3}{4}$ (4) $\frac{4}{3}$

28 Which number is equal to 5^3?

(1) 8 (3) 53
(2) 15 (4) 125

29 Which number is *not* equal to 25%?

(1) $\frac{1}{4}$ (3) 0.25

(2) 2.5 (4) $\frac{25}{100}$

30 Solve for *x*: $\frac{3}{4} = \frac{x}{16}$

(1) 15 (3) 12
(2) 14 (4) 9

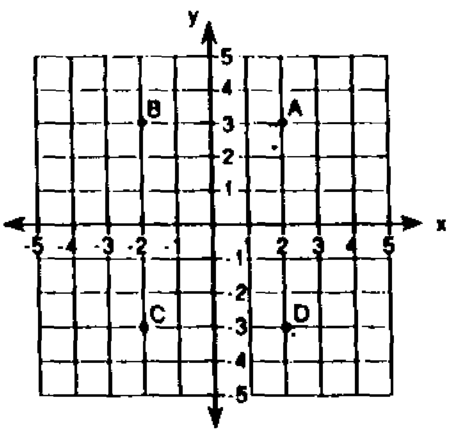
31 In the morning, the temperature was +5°C. By afternoon the temperature had risen to +31°C. How many degrees did the temperature increase during that day?

(1) 26° (3) 30°
(2) 28° (4) 36°

32 An egg truck carrying 200 dozen eggs was involved in an accident. Half of the eggs were broken. What is the *total number* of eggs broken in this accident?

(1) 6
(2) 100
(3) 400
(4) 1200

33 On the graph below, which point has coordinates (2,-3)?



(1) *A*
(2) *B*
(3) *C*
(4) *D*

34 Which is a prime number?

(1) 12
(2) 13
(3) 14
(4) 15

35 What is the largest number less than 50 that is divisible by 4?

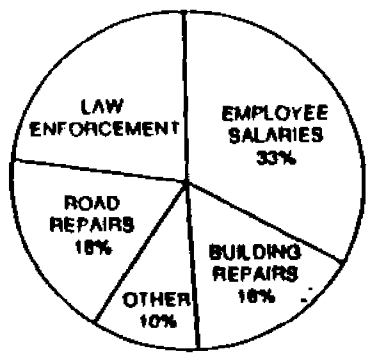
(1) 16
(2) 40
(3) 48
(4) 49

36 Tracy bought toothpaste for $3.73, a toothbrush for $2.48, and mouthwash for $5.39. If she paid for the items with a $20 bill, how much change should she receive?

(1) $11.60
(2) $9.60
(3) $9.40
(4) $8.40

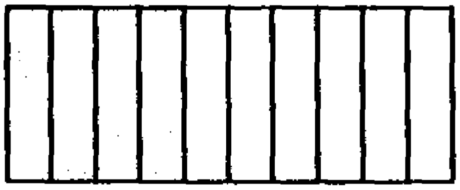
37 What percent of the city's property tax dollar is spent on law enforcement?

YOUR TAX DOLLAR



(1) 23%
(2) 33%
(3) 77%
(4) 283%

38 What percent of the rectangle below is shaded?



(1) 4% (3) 40%
(2) 14% (4) 140%

39 On a map, 1 inch represents 10 miles. If the actual distance between two cities is 25 miles, the distance on the map between the two cities is

(1) 20.5 in (3) 2.5 in
(2) 2 in (4) 25 in

40 Which is the shortest length?

(1) 1 meter
(2) 1 centimeter
(3) 1 kilometer
(4) 1 millimeter

41 Which set of integers is arranged from least value to greatest value?

(1) −1, −3, 1, 3 (3) −3, −1, 1, 3
(2) −1, 1, −3, 3 (4) 3, 1, −3, −1

42 Malcolm kept $1200 in the bank for 1 year. How much interest did he earn on that money if the bank pays an annual interest rate of 8%?

(1) $9.60 (3) $150
(2) $96 (4) $960

43 A bag contains two red marbles and three blue marbles. If one marble is drawn from the bag at random, what is the probability that the marble will be red?

(1) $\frac{2}{5}$ (3) $\frac{3}{5}$

(2) $\frac{1}{2}$ (4) $\frac{2}{3}$

44 Which is the best estimate of 57 × 33?

(1) 1200 (3) 2000
(2) 1800 (4) 2400

45 Leslie earns $2.50 per hour babysitting. If she babysits from 7:00 p.m. to 11:30 p.m., how much will she earn?

(1) $17.50 (3) $11.25
(2) $13.75 (4) $10.50

46 What is the value of

$3(4 + 5) - \frac{(7 - 3)}{2}$?

(1) 7 (3) 23
(2) 13 (4) 25

47 Which group of fractions is arranged in order from smallest to largest?

(1) $\frac{2}{7}, \frac{2}{5}, \frac{2}{3}$ (3) $\frac{2}{5}, \frac{2}{7}, \frac{2}{6}$

(2) $\frac{1}{8}, \frac{1}{5}, \frac{1}{7}$ (4) $\frac{1}{2}, \frac{1}{4}, \frac{1}{3}$

48 Which value of x will make the sentence $2x + 1 > 7$ a true statement?

(1) 1 (3) 3
(2) 2 (4) 4

49 Three gallons of fuel costs a total of $2.85. What is the cost of 10 gallons of fuel?

(1) $8.55 (3) $10.50
(2) $9.50 (4) $28.50

50 A stereo that is regularly priced at $280 is on sale for 25% off. What is the sale price of the stereo?

(1) $70 (3) $210
(2) $140 (4) $240

51 Which statement represents the sentence below?

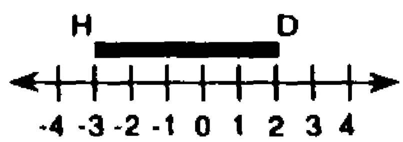
Two more than a number, x, is 20.

(1) $x + 2 = 20$ (3) $2x = 20$
(2) $x + 2 > 20$ (4) $x + 20 = 2$

52 Evaluate: $\sqrt{25} + \sqrt{100}$

(1) 15 (3) 55
(2) 25 (4) 125

53 What is the length of line segment HD on the graph below?



(1) 1 (3) 5
(2) 6 (4) 4

54 Amanda bought three pairs of socks. The total cost was $7.29, which included a $0.54 sales tax. What was the price of each pair of socks before taxes?

(1) $2.25
(2) $2.35
(3) $2.43
(4) $2.61

55 What is the value of $12 \div \frac{1}{2} + 1$?

(1) 7
(2) 8
(3) 18
(4) 25

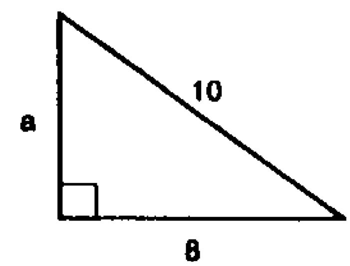
56 How many kilometers are equal to 10,000 meters?

(1) 1
(2) 10
(3) 100
(4) 1,000

57 Which number's value is closest to 6?

(1) $5\frac{1}{2}$
(2) $5\frac{3}{4}$
(3) $6\frac{1}{2}$
(4) $6\frac{3}{4}$

58 Using the formula $a^2 + b^2 = c^2$, what is the value of a in the right triangle below?



(1) 6
(2) 2
(3) 36
(4) 8

59 What is the difference between 29 and $2\frac{1}{3}$?

(1) $26\frac{1}{3}$
(2) $26\frac{2}{3}$
(3) $27\frac{1}{3}$
(4) $27\frac{2}{3}$

60 Using the formula $A = \pi r^2$, what is the area of a circle whose diameter is 6?

(1) 6π
(2) 9π
(3) 3π
(4) 36π

Appendix H

School________________

Date ____________ RCT Form June 1991

Final Classification Instructions

Please discuss your previous analysis with your colleagues. Once you have agreed upon an items classification, place an X in the column that best describes the item's cognitive processes and then the item's mathematical content.

COGNITIVE PROCESSES

ITEM #	Knowledge	Skill	Routine Application	Understanding/ Comprehension	Problem Solving	Comments
21						
22						
23						
24						
25						
26						
27						
28						
29						
30						
31						
32						
33						
34						
35						
36						
37						
38						
39						
40						
41						
42						
43						
44						

	Knowledge	Skill	Routine Application	Understanding/ Comprehension	Problem Solving	Comments
45						
46						
47						
48						
49						
50						
51						
52						
53						
54						
55						
56						
57						
58						
59						
60						

Place an X in the column that best describes the items mathematical content.

Item Content

Set of Integers ITEM #	Rational Numbers	Graphing	Measurement	Ratio, Proportion Percent	Probability Statistics	Consumer Job-Related
21						
22						
23						
24						
25						
26						
27						
28						
29						
30						
31						
32						
33						
34						
35						
36						
37						
38						
39						
40						
41						
42						
43						
44						
45						
46						
47						
48						
49						

	Set of Integers	Rational Numbers	Graphing	Measurement	Ratio, Proportion Percent	Probability Statistics	Consumer Job-Related
50							
51							
52							
53							
54							
55							
56							
57							
58							
59							
60							

Appendix I

Earlier this year you classified RCT questions according to the mathematical content and the cognitive processes of the items. Your classifications were used to define the content and cognitive areas used in this study. These areas were used to identify students who needed remediation in the different areas of content or cognitive processes. The data are from students (the actual names and identification numbers have been deleted) who took the RCT in June 1991 at three Brooklyn high schools.

A sample of a class listing of students has been developed, showing the proposed area(s) of remediation. I would like your help in evaluating the usefulness of these content and cognitive areas of remediation. The following packet contains:

1 the questions defining the content and cognitive areas.
2 a listing of students according to their area of weakness.
3 an evaluation form.

Please refer to this packet when completing the evaluation form. Lastly, please describe how this information would impact on your teaching practices.

AREA OF REMEDIATION: **INTEGERS**
QUESTION NUMBERS: 28, 31, 34, 35, 41, 44, 46, 48, 51, 52
PERCENT OF EXAM: 25%

28 Which number is equal to 5^3?

(1) 8 (3) 53
(2) 15 (4) 125

31 In the morning, the temperature was +5 C. By afternoon the temperature had risen to +31 C. How many degrees did the temperature increase during that day?

(1) 26 (3) 30
(2) 28 (4) 36

34 Which is a prime number?

(1) 12 (3) 14
(2) 13 (4) 15

35 What is the largest number less than 50 that is divisible by 4?

(1) 16 (3) 48
(2) 40 (4) 49

41 Which set of integers is arranged from least value to greatest value?

(1) -1,-3,1,3 (3) -3,-1,1,3
(2) -1,1,-3,3 (4) 3,1,-3,-1

44 Which is the best estimate of 57 x 33?

(1) 1200 (3) 2000
(2) 1800 (4) 2400

46 What is the value of $3(4 + 5) - (7 - 3)/2$?

(1) 7 (3) 23
(2) 13 (4) 25

48 Which value of x will make the sentence $2x + 1 > 7$ a true statement?

(1) 1 (3) 3
(2) 2 (4) 4

51 Which statement represents the sentence below?
Two more than a number, x, is 20

(1) $x + 2 = 20$ (3) $2x = 20$
(2) $x + 2 > 20$ (4) $x + 20 + 2$

52 Evaluate: $\sqrt{25} + \sqrt{100}$

(1) 15 (3) 55
(3) 25 (4) 125

AREA OF REMEDIATION: RATIONAL NUMBERS
QUESTION NUMBERS: 47, 55, 57, 59
PERCENT OF EXAM: 10%

47 Which group of fractions is arranged in order from smallest to largest?

(1) $\frac{2}{7}, \frac{2}{5}, \frac{2}{3}$ (3) $\frac{2}{5}, \frac{2}{7}, \frac{2}{6}$

(2) $\frac{1}{8}, \frac{1}{5}, \frac{1}{7}$, (4) $\frac{1}{2}, \frac{1}{4}, \frac{1}{3}$.

55 What is the value of 12 + 1/2 + 1?

(1) 7 (3) 18
(2) 8 (4) 25

57 Which number's value is closest to 6?

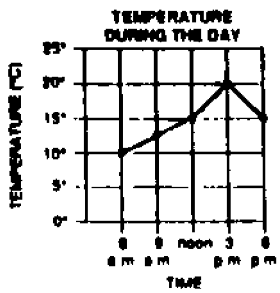
(1) 5 1/2 (3) 6 1/2
(2) 5 3/4 (4) 6 3/4

59 What is the difference between 29 and 2 1/3?

(1) 26 1/3 (3) 27 1/3
(2) 26 2/3 (4) 27 2/3

AREA OF REMEDIATION: GRAPHING
QUESTION NUMBERS: 21, 23, 33, 37, 53
PERCENT OF EXAM: 12%

21 The graph below shows the temperature for various times of the day. At what time did the temperature reach 20 C?



(1) 6:00 a.m. (3) 3:00 p.m.
(2) noon (4) 6:00 p.m.

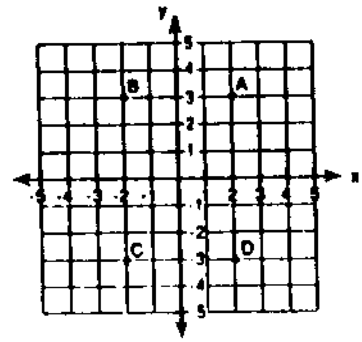
23 The chart below shows the average monthly temperature for certain cities in the United states. What is the average monthly temperature for August in Miami, Florida?

AVERAGE MONTHLY TEMPERATURE (°F)

	Juneau, Alaska	Nashville, Tennessee	Miami, Florida	San Francisco, California
Jan	-11	42	[illegible]	50
Feb	[illegible]	44	[illegible]	53
Mar	[illegible]	53	71	54
Apr	[illegible]	[illegible]	74	[illegible]
May	[illegible]	[illegible]	77	57
June	[illegible]	[illegible]	[illegible]	[illegible]
July	[illegible]	81	82	[illegible]
Aug	55	[illegible]	82	[illegible]
Sept	44	74	81	[illegible]
Oct	27	[illegible]	78	61
Nov	3	52	72	57
Dec	[illegible]	44	[illegible]	52

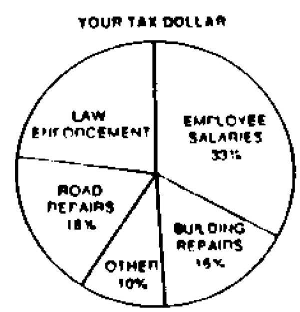
(1) 68 (3) 81
(2) 80 (4) 82

33 On the graph below, which point has coordinates (2,-3)?



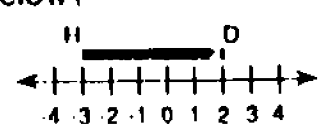
(1) A (3) C
(2) B (4) D

37 What percent of the city's property tax dollar is spent on law enforcement?



(1) 23% (3) 77%
(2) 33% (4) 283%

53 What is the lenght of line segment HD on the graph below?



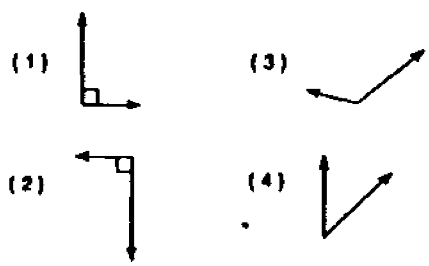
(1) 1 (3) 5
(2) 6 (4) 4

AREA OF REMEDIATION: **MEASUREMENT**
QUESTION NUMBERS: 22, 25, 40, 56, 58, 60
PERCENT OF EXAM: 15%

22 What is the length of the line segment TJ shown below?

(1) 55 cm (3) 0.55 cm
(2) 5.5 cm (4) 5 cm

25 Which drawing shows an acute angle?



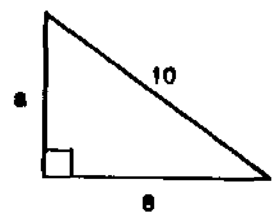
40 What is the shortest length?

(1) 1 meter (3) 1 centimeter
(2) 1 kilometer (4) 1 millimeter

56 How many kilometers are equal 10,000 meters?

(1) 1 (3) 100
(2) 10 (4) 1,000

58 Using the formula $a^2 + b^2 = c^2$ what is the value of a in the right triangle below?



(1) 6 (3) 36
(2) 2 (4) 8

60 Using the formula $A = \Pi r^2$, what is the area of a circle whose diameter is 6?

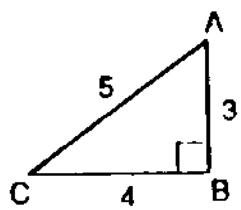
(1) 6 (3) 3
(2) 9 (4) 36

AREA OF REMEDIATION: **RATIO, PROPORTION AND PERCENT**
QUESTION NUMBERS: 24, 27, 29, 30, 38
PERCENT OF EXAM: 12%

24 The fraction 19/3 may be expressed as which mixed number?

(1) 5 1/3 (3) 7 2/3
(2) 6 1/3 (4) 16 1/3

27 What is the ratio of *AB* to *BC* in the right triangle below?



(1) 3/5 (3) 4/5
(2) 3/4 (4) 4/3

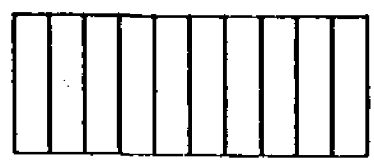
29 Which number is *not* equal to 25%?

(1) 1/4 (3) 0.25
(2) 2.5 (4) 25/100

30 Solve for *x*: 3/4 = *x*/16

(1) 15 (3) 12
(2) 14 (4) 9

38 What percent of the rectangle below is shaded?



(1) 4% (3) 40%
(2) 14% (4) 140%

AREA OR REMEDIATION: **PROBABILITY AND STATISTICS**
QUESTION NUMBERS: 43
PERCENT OF EXAM: 3%

43 A bag contains two red marbles and three blue marbles. If one marble is drawn at random, what is the probability that the marble will be red?

(1) 2/5	(3) 3/5
(2) 1/2	(4) 2/3

AREA OF REMEDIATION: **CONSUMER AND JOB RELATED**
QUESTION NUMBERS: 26, 32, 36, 42, 45, 49, 50, 54
PERCENT OF EXAM: 20%

26 Jill has $282.15 in her checking account. If she writes checks for $25 and $46.75, how much money will be left in her account?

(1) $210.40 (3) $329.15
(2) $235.15 (4) $353.90

32 An egg truck carrying 200 dozen eggs was involved in an accident. Half of the eggs were broken. What is the *total number* of eggs broken in this accident?

(1) 6 (3) 400
(2) 100 (4) 1200

36 Tracy bought toothpaste for $3.73, a toothbrush for $2.48, and mouthwash for $5.39. If she paid for the items with a $20 bill, how much change should she receive?

(1) $11.60 (3) $9.60
(2) $9.60 (4) $8.40

54 Amanda bought three pairs of socks. The total cost was $7.29, which includes a $0 54 sales tax. What was the price of each pair of socks before taxes?

(1) $2.25 (3) $2 43
(2) $2.35 (4) $2 61

42 Malcolm kept $1200 in the bank for 1 year. How much interest did he earn on that money if the bank pays an annual interest rate of 8%?

(1) $9.60 (3) $150
(2) $96 (4) $960

45 Leslie earns $2.50 per hour babysitting. If she babysits from 7:00 p.m. to 11:30 p.m., how much will she earn?

(1) $17.50 (3) $11.25
(2) $13.75 (4) $10.50

49 Three gallons of fuel costs a total of $2.85. What is the cost of 10 gallons of fuel?

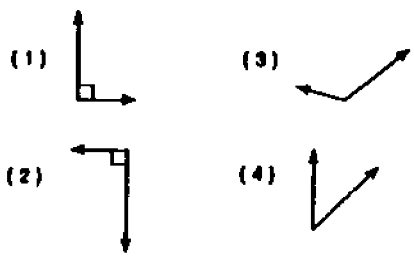
(1) $8.55 (3) $10.50
(2) $9.50 (4) $28.50

50 A stereo that is regularly priced at $280 is on sale for 25% off What is the sale price of the stereo?

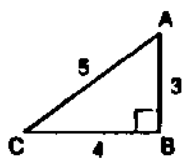
(1) $70 (3) $210
(2) $140 (4) $240

AREA OF REMEDIATION: **KNOWLEDGE**
QUESTION NUMBERS: 25, 27, 33, 34, 38, 40, 41, 47, 56
PERCENT OF EXAM: 22%

25 Which drawing shows an acute angle?

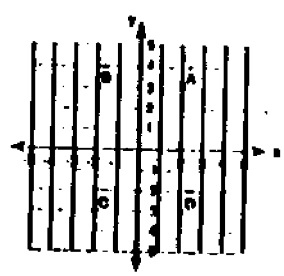


27 What is the ratio of AB to BC in the right triangle below?



(1) 3/5 (3) 4/5
(2) 3/4 (4) 4/3

33 On the graph below, which point has coordinates (2, -3)?



(1) A (3) C
(2) B (4) D

34 Which is a prime number?

(1) 12 (3) 14
(2) 13 (4) 15

38 What percent of the rectangle below is shaded?

(1) 4% (3) 40%
(2) 14% (4) 140%

40 Which is the shortest length?

(1) 1 meter (3) 1 kilometer
(2) 1 centimeter (4) 1 millimeter

41 Which set of integers is arranged from least value to greatest value?

(1) -1, -3, 1, 3 (3) -3, -1, 1, 3
(2) -1, 1, -3, 3 (4) 3, 1, -3, -1

47 Which group of fractions is arranged in order from smallest to largest?

(1) $\frac{2}{7}, \frac{2}{5}, \frac{2}{3}$ (3) $\frac{2}{5}, \frac{2}{7}, \frac{2}{6}$

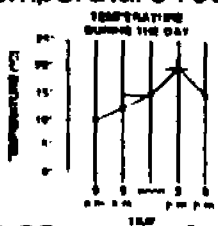
(2) $\frac{1}{8}, \frac{1}{5}, \frac{1}{7}$ (4) $\frac{1}{2}, \frac{1}{4}, \frac{1}{3}$

56 How many kilometers are equal to 10,000 meters?

(1) 1 (3) 100
(2) 10 (4) 1,000

AREA OF REMEDIATION: **SKILL**
QUESTION NUMBERS: 21, 22, 23, 24, 28, 29, 30, 37, 46, 52, 53, 55, 58, 59, 60
PERCENT OF EXAM 38%

21 The graph below shows the temperature for various times of the day. At what time did the temperature reach 20 C?

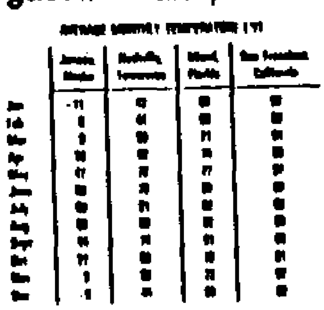


(1) 6:00 a.m. (3) 3:00 p.m.
(2) noon (4) 6:00 p.m.

22 What is the length of the line segment *TJ* shown below?

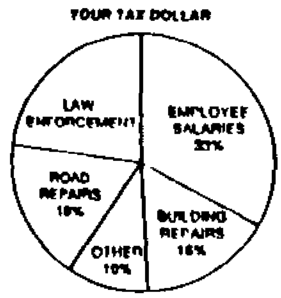
(1) 55cm (3) 0.55cm
(2) 5.5cm (4) 5cm

23 The chart below shows the average monthly temperature for certain cities in the United States. What is the average monthly temperature for August in Miami, Florida?



(1) 68 (3) 81
(2) 80 (4) 82

37 What percent of the city's property tax dollar is spent on law enforcement?



(1) 23% (3) 77%
(2) 33% (4) 283%

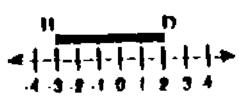
46 Which is the best estimate of 57 x 33?

(1) 1200 (3) 2000
(2) 1800 (4) 2400

52 Evaluate $\sqrt{25} + \sqrt{100}$

(1) 15 (3) 55
(2) 25 (4) 125

53 What is the length of line segment *HD* on the graph below?



(1) 1 (3) 5
(2) 6 (4) 4

55 What is the value of 12 - 1/2 + 1 ?

(1) 7 (3) 18
(2) 8 (4) 25

24 The fraction 19/3 may be expressed as which mixed number?

(1) 5 1/3 (3) 7 2/3
(2) 6 1/3 (4) 16 1/3

28 What number is equal to 5^3 ?

(1) 8 (3) 53
(2) 15 (4) 125

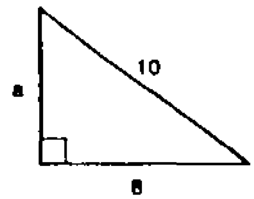
29 Which number is *not* equal to 25%?

(1) 1/4 (3) .25
(2) 2.5 (4) 25/100

30 Solve for x: 3/4 = x/16

(1) 15 (3) 12
(2) 14 (4) 9

58 Using the formula $a^2 + b^2 = c^2$ what is the value of a in the right triangle below?



(1) 6 (3) 36
(2) 2 (4) 8

59 What is the difference between 29 and 2 1/3?

(1) 26 1/3 (3) 27 1/3
(2) 26 2/3 (4) 27 2/3

60 Using the formula $a = \Pi r^2$, what is the area of a circle whose diameter is 6?

(1) 6 (3) 3
(2) 9 (4) 36

AREA OF REMEDIATION: **ROUTINE APPLICATION**
QUESTION NUMBERS: 31,35, 42, 43, 44, 57
PERCENT OF EXAM: 15%

31 In the morning, the temperature was +5 C. By afternoon the temperature had risen to +31 C. How many degrees did the temperature increase during that day?

(1) 26 (3) 30
(2) 28 (4) 36

35 What is the largest number less than 50 that is divisible of by 4?

(1) 16 (3) 48
(2) 40 (4) 49

42 Malcolm kept $1200 in the bank for 1 year. How much interest did he earn on that money if the bank pays an annual interest rate of 8%?

(1) $ 9.60 (3) $150
(2) $96 (4) $960

43 A bag contains two red marbles and three blue marbles. If one marbles is drawn at random, what is the probability that the marble will be red?

(1) 2/5 (3) 3/5
(2) 1/2 (4) 2/3

44 What is the best estimate 57 x 33?

(1) 1200 (3) 2000
(2) 1800 (4) 2400

57 Which number is closest to 6?

(1) 5 1/2 (3) 6 1/2
(2) 5 3/4 (4) 6 3/4

AREA OF REMEDIATION **UNDERSTANDING/COMPREHENSION**
QUESTION NUMBERS. 39, 48, 49 51
PERCENT OF EXAM: 10%

39 On a map, 1 inch represents 10 miles. If the actual distance between two cities is 25 miles, the distance on the map between the two cities is

(1) 20.5 in (3) 2.5 in
(2) 2 in (4) 25 in

48 What value of x will make the sentence $2x + 1 > 7$ a true statement?

(1) 1 (3) 3
(2) 2 (4) 4

49 Three gallons of fuel costs a total of $2.85. What is the cost of 10 gallons of fuel?

(1) $ 8.55 (3) $ 10.50
(2) $9.50 (4) $28.50

51 Which statement represents the statement below?
Two more than a number, is 20

(1) $x + 2 + 20$ (3) $2x + 20$
(3) $x + s > 20$ (4) $x + 20 + 2$

AREA OF REMEDIATION: PROBLEM SOLVING
QUESTION NUMBERS: 26, 32, 36, 45, 50, 54
PERCENT OF EXAM: 15%

26 Jill has $ 282.15 in her checking account. If she writes checks for $25 and $46.75, how much money will left in her account?

(1) $210.40 (3) $329.15
(2) $235.15 (4) $353.90

32 An egg truck carrying 200 dozen eggs was involved in an accident. Half of the eggs were broken. What is the *total number* of eggs broken in this accident?

(1) 6 (3) 400
(2) 100 (4) 1200

36 Tracy bought toothpaste for $3.73, a toothbrush for $ 2.48, and mouthwash for $5.39. If she paid for the items with a $20 bill, how much change should she receive?

(1) $11.60 (3) $9.40
(2) $9.60 (4) $8.40

45 Leslie earns $2.50 per hour babysitting. If she babysits from 7:00 p.m. to 11:30 p.m., how much will she earn?

(1) $17.50 (3) $11.25
(2) $13.75 (4) $10.50

50 A stereo that is regularly priced at $280 is on sale for 25% off. What is the sale price of the stereo?

(1) $70 (3) $210
(2) $140 (4) $240

54 Amanda bought three pairs of socks. The total cost was $7.29, which included a $0.54 sales tax. What was the price of each pair of socks before taxes?

(1) $2.25 (3) $2.43
(2) $2.35 (4) $ 2.61

CLASS SUMMARY FOR MR. SMITH'S MATH LAB

AREA OF REMEDIATION	STUDENT NAME	RCT FINAL SCORE
Integers 25% of exam	Mark Farkas	58%
	Lisa Gruled	58%
	Jill Maldonado	55%
	Hector Ortiz	58%
	Russel Pierce	58%
	Laura Reyes	60%
	Jesus Solo	55%
	Peter Williams	58%
Rational Numbers 10% of exam	Laura Baez	63%
	Alfred Coyle	65%
	Mitchel Fox	65%
	Mary Gross	63%
	Lisa Gruled	58%
	Jasmine Jones	63%
	Barbara Martin	60%
	Russel Pierce	55%
	Peter Williams	58%
Graphing 10% of exam		
Measurement 18% of exam	Mark Farkas	58%
	Jasmine Jones	63%
	Barbara Martin	60%
	Laura Reyes	60%
	Jesus Solo	55%
Ratio, Proportion and Percent 15% of exam	Alfred Coyle	65%
	Mary Gross	63%
Probability and Statistics 2% of exam	Laura Baez	63%
	Tim Collins	60%
	Robert Dunn	58%
	Mitchel Fox	65%
	Jill Maldonado	55%
	Hector Ortiz	58%
Consumer and Job-Related 20% of exam	John Anderson	63%
	Leslie Barbar	60%
	Tim Collins	60%
	Robert Dunn	58%

CLASS SUMMARY FOR MR. SMITH'S SEQUENTIAL MATH CLASS

AREA OF REMEDIATION	STUDENT NAME	RCT FINAL SCORE
Integers	Kris Delana	63%
25% of exam	Angel Fama	55%
	Jill Maldonado	55%
	Danny Lund	58%
	Rose May	60%
	Lou Soux	63%
	Mark Yoast	58%
Rational Numbers		
10% of exam		
Graphing	Miranda Seal	87%
10% of exam	Robin Smith	87%
	Patricia Wyle	90%
Measurement	Ronald Dux	79%
18% of exam	Carol Bunn	83%
	Melanis Katz	82%
	Joe Mack	82%
	Ellen Marks	80%
	Linda Moskos	71%
	Boris Taylor	68%
Ratio, Proportion and Percent	Kris Delano	63%
15% of exam	Rose May	60%
	Linda Moskos	71%
	Lou Soux	63%
	Boris Tarlo	68%
Probability and Statistics	Carol Bunn	83%
2% of exam	Alex Kowski	95%
	Oleg Krinsky	98%
	Ellen Marks	80%
Consumer and Job-Related	Angel Fama	55%
20% of exam	Danny Lund	58%
	Mark Yoast	58%

CLASS SUMMARY FOR MS. YOUNG'S MATH LAB

AREA OF REMEDIATION	STUDENT NAME	RCT FINAL SCORE
Knowledge	Pauline Ahrens	55%
22% of exam	Roseann Cordo	58%
	Maria Gray	58%
	Elkin Lopez	63%
	David Penn	53%
	Dianne Phillips	55%
	Lance Phox	53%
	Adam Rid	58%
	Aileen Sanchez	58%
Skill	Pauline Ahrens	55%
38% of exam	Maya Cave	63%
	Gary Chase	63%
	Vivian Glass	65%
	Maria Gray	58%
	Merlin Oppie	60%
	Adam Rid	58%
Routine Application	Carmen Blank	58%
15% of exam	Gary Chase	63%
	Debbie Harris	60%
	Elkin Lopez	53%
	Merl Oppie	60%
	David Penn	53%
	Lance Phox	50%
Understanding and Comprehension	Carmen Blank	58%
	Maya Cave	63%
10% of exam	Vivian Glass	65%
Problem Solving	Roseann Cordo	58%
15% of exam	Debbie Harris	60%
	Dianne Phillips	55%
	Aileen Sanchez	58%

CLASS SUMMARY FOR MS. YOUNG'S SEQUENTIAL MATH CLASS

AREA OF REMEDIATION	STUDENT NAME	RCT FINAL SCORE
Knowledge	Dinora Aquiler	40%
22% of exam	Grace Box	65%
	Wendy Key	62%
	Nancy Pez	62%
Skill	Dinora Aquiler	40%
38% of exam	Walter Girotto	55%
	Nancy Pez	43%
	Michael Romano	48%
	Adriane Sands	58%
	Amy Yearly	53%
	Sandy Young	53%
Routine Application	Theo Blot	88%
15% of exam	Sergio Dupree	83%
	Gigi Ezagui	75%
	Thomas Green	85%
	Dana Kubiak	78%
	Michael Romano	48%
	Amy Yearly	53%
	Sandy Young	53%
Understanding and Comprehension	Gigi Ezagui	75%
10% of exam	Walter Girotto	55%
	Dana Kubiak	78%
	Adriane Sands	58%
	Anna Troia	90%
	Jennifer Tsang	90%
Problem Solving	Grace Box	65%
15% of exam	Jessica Chan	88%
	Gerry Gonzalez	85%
	Wendy Key	62%
	Agnes Label	88%
	Vicky Volfson	85%

Please rate the degree to which you agree with the following statements.

	STRONGLY AGREE	AGREE	NEUTRAL	DISAGREE	STRONGLY DISAGREE
The remediation areas are taught as instructional units.	SA	A	N	D	SD
Please note any exceptions ________					
The remediation areas could be taught as instructional units.	SA	A	N	D	SD
Please note any exceptions ________					
Grouping students by the remediation areas would facilitate topic exploration.	SA	A	N	D	SD
Comments ________					
Grouping students by the remediation areas is *not* meaningful for instructional purposes.	SA	A	N	D	SD
Comments ________					
The benefits from implementing the necessary curriculum changes would *not* justify the time taken from other topics.	SA	A	N	D	SD
Comments ________					
Teaching the remediation areas as distinct units would *not* increase test scores.	SA	A	N	D	SD
Comments ________					
Teaching the remediation areas would increase student comprehension and understanding.	SA	A	N	D	SD
Comments ________					

For what instructional strategies do you think the information provided is best suited?

____

____

____

____

____

____

How could the information provided affect your time management?

____

____

____

____

____

____

How could the information provided change your view of your students?

____

____

____

____

____

____

How would this information impact on your current teaching practices?
Be specific:

a) I would do more:______________________________

____

____

b) I would do less:______________________________

____

____

c) I would do about the same:______________________

____

____

Other____

____

____

____

____

	STRONGLY AGREE	AGREE	NEUTRAL	DISAGREE	STRONGLY DISAGREE
Teaching the remediation areas in a group setting would increase student interest.	SA	A	N	D	SD
Teaching the remediation areas as instructional units would allow for in depth problem analyses.	SA	A	N	D	SD
I would use the information from the class summary sheet for instructional planning.	SA	A	N	D	SD
Teaching the remediation areas as units would serve as building blocks for successive units.	SA	A	N	D	SD
Teaching the remediation areas would *not* help to clarify math concepts.	SA	A	N	D	SD

Comments (after each item) ______________________________

When answering the following questions, please refer to the class summary sheet. Whenever possible be specific, select several student's names to illustrate your points.

What instructional activities would you use based on this information?

____
____
____
____
____
____

Appendix J

Results of Expert Item Classification

Area of Remediation	Percent of Exam	Question Numbers
	Mathematical Content	
Integers	25%	28, 31, 34, 35, 41, 44, 46, 48, 51, 52
Rational Numbers	10%	47, 55, 57, 59
Graphing	12%	21, 23, 33, 37, 53
Measurement	15%	22, 25, 40, 56, 58, 60
Ratio, Proportion and Percent	12%	24, 27, 29, 30, 30
Probability and Statistics	3%	43
Consumer and Job Related	20%	26, 32, 36, 42, 45, 49, 50, 54
	Cognitive Processes	
Knowledge	22%	25, 27, 33, 34, 38, 40, 41, 47, 56
Skill	38%	21, 22, 23, 24, 28, 29, 30, 37, 46, 52, 53, 55, 58, 59, 60
Routine Application	15%	31, 35, 42, 43, 44, 57
Understanding and Comprehension	10%	39, 48, 49, 51
Problem Solving	15%	26, 32, 36, 45, 50, 54

Appendix K

THETA VALUE FOR EACH STUDENT RESPONSE PATTERN
AND POSTERIOR PROBABILITY FOR IRT AND LATENT CLASSES
IRT QUADRATURE POINTS FIRST AND LATENT CLASSES

1 99.000000
.000 .000 .000 .000 .000 .000 .000 .000 .000 .000
.000 .000 .000 .000 .000 .000 .000 .000 .000 .000
1.000 .000 .000 .000
0111011101 110111
2 99.000000
.000 .000 .000 .000 .000 .000 .000 .000 .000 .000
.000 .000 .000 .000 .000 .000 .000 .000 .000 .000
.0001.000 .000 .000
1011101110 111011
3 99.000000
.000 .000 .000 .000 .000 .000 .000 .000 .000 .000
.000 .000 .000 .000 .000 .000 .000 .000 .000 .000
.000 .0001.000 .000
1101110111 011101
4 99.000000
.000 .000 .000 .000 .000 .000 .000 .000 .000 .000
.000 .000 .000 .000 .000 .000 .000 .000 .000 .000
.000 .000 .0001.000
1110111011 101110
5 1.860687
.000 .000 .000 .000 .000 .000 .000 .000 .000 .000
.000 .009 .000 .000 .000 .000 .000 .000 .000 .990
.000 .000 .000 .000
1111111111 111111
6 -2.071655
.894 .004 .000 .000 .000 .000 .000 .000 .102 .000
.000 .000 .000 .000 .000 .000 .000 .000 .000 .000
.000 .000 .000 .000
0000000000 000000

7 -.661576
.000 .000 .000 .000 .000 .000 .000 .000 .0001.000
.000 .000 .000 .000 .000 .000 .000 .000 .000 .000
.000 .000 .000 .000
1111000000 000000
8 .119089
.000 .000 .000 .000 .000 .000 .000 .000 .000 .102
.580 .319 .000 .000 .000 .000 .000 .000 .000 .000
.000 .000 .000 .000
1111000011 110000
9 1.263351
.000 .000 .000 .000 .000 .000 .000 .000 .000 .000
.004 .993 .000 .000 .000 .000 .000 .000 .000 .000
.000 .000 .000 .003
1111000011 111111
10 99.000000
.000 .000 .000 .000 .000 .000 .000 .000 .000 .061
.000 .000 .000 .000 .000 .000 .000 .000 .000 .000
.000 .939 .000 .000
0011001100 110011
11 -1.215619
.000 .000 .000 .000 .000 .000 .000 .000 .925 .075
.000 .000 .000 .000 .000 .000 .000 .000 .000 .000
.000 .000 .000 .000
0001000100 010001
12 .426744
.000 .000 .000 .000 .000 .000 .000 .000 .000 .008
.247 .744 .000 .000 .000 .000 .000 .000 .000 .000
.000 .000 .000 .001
1111111111 110000
13 -.631422
.000 .000 .000 .000 .000 .000 .000 .000 .000 .997
.000 .000 .000 .000 .000 .000 .000 .000 .000 .000
.000 .000 .000 .003
1111111100 000000

```
14          99.000000
     .000 .000 .000 .000 .000 .000 .000 .000 .000 .000
     .000 .000 .000 .000 .000 .000 .000 .000 .000 .000
     .000 .969 .000 .031
1010101010 101011
15          99.000000
     .000 .000 .000 .000 .000 .000 .000 .000 .000 .000
     .000 .000 .000 .000 .000 .000 .000 .000 .000 .000
     .000 .0001.000 .000
1101100001 010101
16          -.640633
     .000 .000 .000 .000 .000 .000 .000 .000 .000 .999
     .000 .000 .000 .000 .000 .000 .000 .000 .000 .000
     .000 .000 .000 .001
1101000100 101010
17          -.694301
     .000 .000 .000 .000 .000 .000 .000 .000 .000 .999
     .000 .000 .000 .000 .000 .000 .000 .000 .000 .000
     .000 .000 .000 .001
0011010111 010010
18          99.000000
     .000 .000 .000 .000 .000 .000 .000 .000 .000 .000
     .000 .000 .000 .000 .000 .000 .000 .000 .000 .000
     .000 .000 .0001.000
1100101010 100101
19          -1.636489
     .000 .000 .000 .000 .000 .000 .000 .000 .996 .001
     .000 .000 .000 .000 .000 .000 .000 .000 .000 .000
     .000 .003 .001 .000
0000100010 001001
20          99.000000
     .000 .000 .000 .000 .000 .000 .000 .000 .000 .000
     .000 .000 .000 .000 .000 .000 .000 .000 .000 .000
     .000 .009 .000 .991
1010101011 111010
```

REFERENCES

Alvarez, L. (1990). [Rasch analysis of June 1989 RCT]. Unpublished raw data.

Bergen J. R., Stone, C. A., & Feld, J. K. (1984). Rule replacement in the development of basic number skills. Journal of Educational Psychology, 76, 289-299.

Birenbaum, M., Kelly, A. E., & Tatsuoka, K. K. (1992). Towards a stable diagnostic representation of students' errors in algebra. ETS Research Report No. RR-92-58-ONR. Princeton, N. J. Educational Testing Service

Birenbaum, M., Kelly, A. E., & Tatsuoka, K. K. (1993). Diagnosing knowledge states in algebra using the rule-space model. Journal for Research in Mathematics Education, 24(5), 442-459.

Birenbaum, M., Tatsuoka, K. K., & Gutvirtz, Y. (1992). Effects of response format on diagnostic assessment of scholastic achievement. Applied Psychological Measurement, 16(4), 353-363.

Bloom, B. S. (1956). Taxonomy of educational objectives. New York: D. Mckay.

Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. Psychometrika, 37, 29-51.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: An application of an EM algorithm. Psychometrika, 46, 443-459.

Burton, N. W. (1978). Societal standards. Journal of Educational Measurement, 15, 263-271.

Clogg, C. C. (1977). Unrestricted and restricted maximum likelihood latent class analysis: A manual for users. University Park, PA: Population Issues Research Office.

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20(1), 37-46.

Dempster, A. P., Laird, N. M., & Rubin D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society (Series B), 39, 1-38.

Falmagne, J-C. (1989). A latent trait model via a stochastic learning theory for a knowledge space. Psychometrika, 54, 283-303.

Gitomer, D. H. & Rock, D. A. (1989) Addressing process variables in test analysis. In N. Fredriksen, R. J. Mislevy and I. Bejar (Eds.), Test theory for a new generation of tests. Hillsdale, N. J.: Erlbaum.

Gitomer, D. H., & Yamamoto, K. (1991). Performance modeling that integrates latent trait theory and class theory. Journal of Educational Measurement, 28, 173-189.

Glaser, R. (1981). The future of testing : A research agenda for cognitive psychology and psychometrics. American Psychologist, 36, 923-926.

Goodman, L. A.(1974). The analysis of qualitative variables when some of the variables are unobservable. Part I-A modified latent structural approach. American Journal of Sociology, 79, 1179-1259.

Heartel, E. (1884a). Detection of a skill dichotomy using standardized achievement test items. Journal of Educational Measurement, 21, 59-72.

Heartel, E. (1984b). An application of latent class models to assessment data. Applied Psychological Measurement, 8, 333-346.

Heartel, E. (1989). Using restricted latent class models to map the skill structure of achievement items. Journal of Educational Measurement, 26(4), 301-321.

Hulin, C. L., Drasgow, F., & Parsons, C. K. (1983). Item response theory: Application to psychological measurement. Illinois: Dow Jones-Irwin.

Lord, F.M. (1980). Applications of item response theory to practical testing problems. Hillsdale, N. J: Lawrence Erlbaum Associates

Masters, G. N. (1982). A Rasch model for partial credit scoring. Psychometrika, 47, 149-1874.

Masters, G. N. (1985). A comparison of latent trait and latent class analysis of likert-type data. Psychometrika, 50, 69-82.

Masters, G. N., & Mislevy, R. J. (1989) New views of student learning: Implications for educational measurement. In N. Fredriksen, R. J. Mislevy and I. Bejar (Eds.), Test theory for a new generation of tests. Hillsdale, N. J.: Erlbaum.

Messick, S. (1984). The psychology of educational measurement. Journal of Educational Measurement, 23, 147-156.

Mislevy, R.J. (1989). Foundations of a new test theory. In N. Frederiksen, R. J. Mislevy, & I.I. Bejar (Eds.), Test theory for a new generation of tests. Hillsdale, N. J.: Erlbaum.

Mislevy, R. J., & Verhelst, N. (1990). Modeling item responses when different subjects follow different solution strategies. Psychometrika, 55, 195-215.

Mislevy, R. J., Yamamoto, K., & Anacker, S. (1990). Toward a test theory for assessing student understanding (Tech. Rep. No. ETS-RR-91-32). Princeton, NJ: Educational Testing Service.

National Council of Teachers of Mathematics. (1989). News Bulletin. Virginia: The Council.

Paulson, J. A. (1982). A discrete latent-state approach to diagnostic testing (Tech. Rep. No. ONR 82-1). Portland, OR: Portland State University.

Paulson, J. A. (1985). Latent class representation of systematic patterns in test responses (Tech. Rep. No. ONR 85-1). Portland, OR: Portland State University.

Popham, W. J. (1993). Educational testing in America: What's right, what's wrong. Educational Measurement: Issues and Practices. , 12(1). 11-14.

Rindskopf, D. (1983). A general framework for using latent class analysis to test hierarchical and nonhierarchical learning models. Psychometrika, 48(1), 85-97.

Rindskopf, D.(1987). Using latent class analysis to test developmental models. Developmental Review, 7, 66-85.

Snow, R. E. (1989). Toward assessment of cognitive and conative structures in learning. Educational Researcher, 18(9), 8-14.

Snow, R. E. & Mandinach, E. B. (1991). Integrating assessment and instruction: A research and development agenda. ETS Research Report No. RR-91-8. Princeton, N.J.: Educational Testing Service.

Tatsuoka, K. K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. Journal of Educational Measurement, 20, 345-354.

Tatsuoka, K. K. (1984). Caution indices based on item response theory. Psychometrika, 49, 95-110.

Tatsuoka, K. K., Birenbaum, M., & Arnold, J. (1989). On the stability of students' rules of operation for solving arithmetic problems. Journal of Educational Measurement, 26(4), 351-361.

Tatsuoka, K. K. & Tatsuoka, M. M. (1987). Bug distribution and statistical pattern classification. Psychometrika, 52, 193-206.

University of the State of New York Bureau of General Education Curriculum Development. (1978). General High School Mathematics. Albany, New York: Author.

Wilson, M. R. (1989). Saltus: A psychometric model of discontinuity in cognitive development. Psychological Bulletin, 105, 276-289.

Yamamoto, K. (1987). A model that combines IRT and latent class models. Unpublished doctoral dissertation, University of Illinois.

Yamamoto, K. (1989). Hybrid model of IRT and latent class models. ETS Research Report No. RR-89-41. Princeton, N.J.: Educational Testing Service.