DOCUMENT RESUME

ED 366 651 TM 021 075

AUTHOR Tatsuoka, Kikumi K.; And Others
TITLE Proficiency Scaling Based on Conditional Probability Functions for Attributes.
INSTITUTION Educational Testing Service, Princeton, N.J.
SPONS AGENCY Office of Naval Research, Arlington, VA. Cognitive and Neural Sciences Div.
REPORT NO RR-93-50-ONR
PUB DATE Oct 93
CONTRACT N00014-90-J-1307
NOTE 77p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS Cognitive Processes; Cognitive Tests; Functions (Mathematics); High Schools; High School Students; Item Response Theory; Mathematics Tests; *Probability; Rating Scales; *Scaling; Scores; Tables (Data); *Test Construction; Test Results
IDENTIFIERS Attribute Characteristic Curves; Conditional Probability Functions; *Proficiency Scales; Rule Space Model; *Scholastic Aptitude Test

ABSTRACT

This study introduces procedures for constructing a proficiency scale for a large-scale test by applying Tatsuoka's Rule Space Model. The SAT Mathematics (SAT M) Section 2 is used for illustrating the process and the results. A task analysis is summarized in a mapping sentence, and then 14 processes and content attributes are identified for explaining the underlying cognitive aspects of the examinees' performance on the SAT M. Analysis results show that almost 98 percent of 2,334 examinees are successfully classified into one of 468 cognitive states. The cognitive states are characterized by mastery or non-mastery of the 14 attributes. Attribute Characteristic Curves, which are conditional probability functions defined on the SAT Scale, are introduced and used for interpreting an examinee's proficiency. Prototypes of a student's performance report and a group performance report are given as examples of possible ways for summarizing the analysis results. The study contains 19 tables and 1 figure. One appendix (two tables) introduces the Rule Space Model. (Contains 38 references.) (Author)

RR-93-50-ONR

PROFICIENCY SCALING BASED ON CONDITIONAL PROBABILITY FUNCTIONS FOR ATTRIBUTES

Kikumi K. Tatsuoka
Menucha Birenbaum
Charles Lewis
Kathleen M. Sheehan

This research was sponsored in part by the
Cognitive Science Program
Cognitive and Neural Sciences Division
Office of Naval Research, under
Contract No. N00014-90-J-1307
R&T 4421559

Kikumi K. Tatsuoka, Principal Investigator

Educational Testing Service

Princeton, NJ

October 1993

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED Final, April 1989 - August 1993

4. TITLE AND SUBTITLE	5. FUNDING NUMBERS
Proficiency Scaling Based on Conditional Probability Functions for Attributes	C-N00014-90-J-1307 61153 N RR 04204-01 R & T 4421559
6. AUTHOR(S) Kikumi Tatsuoka, Menucha Birenbaum, Charles Lewis, and Kathleen Sheehan	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Educational Testing Service Rosedale Road Princeton, NJ 08541	8. PERFORMING ORGANIZATION REPORT NUMBER RR-93-50-ONR
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Cognitive Sciences (1142CS) Office of Naval Research 800 N. Quincy Street Arlington, VA 22217-5000	10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT	12b. DISTRIBUTION CODE
Approved for public release. Distribution unlimited.	

13. ABSTRACT (Maximum 200 words)

This study introduces procedures for constructing a proficiency scale for a large-scale test by applying Tatsuoka's Rule Space Model. The SAT Mathematics (SAT M), Section II, is used for illustrating the process and the results. A task analysis is summarized in a mapping sentence, and then 14 processess and content attributes are identified for explaining the underlying cognitive aspects of the examinees' performance on the SAT M. Analysis results show that almost 98% of 2,334 examinees are successfully classified into one of 468 cognitive states. The cognitive states are characterized by mastery or non-mastery of the 14 attributes. Attribute Characteristic Curves, which are conditional probability functions defined on the SAT Scale, are introduced and used for interpreting an examinee's proficiency. Prototypes of a student's performance report and a group performance report are given as examples of possible ways for summarizing the analysis results.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Conditional probability functions, cognitive diagnosis, classification, IRT, SAT Mathematics, proficency scaling.	65
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev 2-89)
Prescribed by ANSI Std Z39-18
298-102

PROFICIENCY SCALING BASED ON CONDITIONAL PROBABILITY FUNCTIONS FOR ATTRIBUTES

Kikumi K. Tatsuoka

Menucha Birenbaum

Charles Lewis

Kathleen M. Sheehan

EDUCATIONAL TESTING SERVICE,

Princeton, NJ

October 13, 1993

Running: proficiency scaling

This research was sponsored in part by the Cognitive Science Program, Cognitive and Neural Sciences Division, Office of Naval Research, under Contract No. N00014-90-J-1307, R & T 4421559, and was sponsored in part by the College Board and Program Research Planning Council.

ABSTRACT

This study introduces procedures for constructing a proficiency scale for a large-scale test by applying Tatsuoka's Rule Space Model. The SAT Mathematics (SAT M), Section 2, is used for illustrating the process and the results. A task analysis is summarized in a mapping sentence, and then 14 processes and content attributes are identified for explaining the underlying cognitive aspects of the examinees' performance on the SAT M. Analysis results show that almost 98% of 2334 examinees are successfully classified into one of 468 cognitive states. The cognitive states are characterized by mastery or non-mastery of the 14 attributes. Attribute Characteristic Curves, which are conditional probability functions defined on the SAT Scale, are introduced and used for interpreting an examinees' proficiency. Prototypes of a student's performance report and a group performance report are given as examples of possible ways for summarizing the analysis results.

INTRODUCTION

Recent developments in cognitive theories have shown that learning is the reorganization and integration of complex tasks. However, learning models considered by educational measurement are primarily linear, and hence measurement models that have been developed support the unidimensionality view of ability levels. The purpose and goal of these models are focused on making inferences about amount of ability or amount of knowledge that an individual possesses, which can be located on the continuum.

A new view of achievement that emerges from cognitive and domain studies emphasizes the importance of how knowledge is organized, what processes are used to solve problems, the degree to which certain procedures and processes are automated, and the ability to represent knowledge in a variety of ways. New measurement models should be able to measure such abilities, as well as traditional ability levels. The movement for searching for an instructionally useful way of assessing students's performance has indicated the need for new measurement theories and models. The movement for enhancing the interpretability of test scores also urges one to develop a new methodology by which test users with different interests in using performance results would be satisfied.

Beaton (1988) introduced a method, called empirical anchoring and applied it to the NAEP tests. Rock & Johnson (1989) applied this method to the SAT. The method starts out by empirically selecting items that discriminate between various levels on the total score distribution. These items are called "anchoring" items. Then experts review the anchor items that describe the skills necessary to achieve that particular score level. The method provides empirical probabilities of success on each of the items for students whose scores were near the anchoring points of the scale. Although this method has attracted a substantial amount of attention from educators, it also has invited criticism from researchers in educational measurement and psychometrics (Forsyth, 1991).

Marco, Crone, Braswell, Curley and Wright (1990) investigated the relationship between SAT content variables and their predictive validity and found that some cognitive tasks are important for predicting students' success in their future performance.

However, test item development has been atheoretical in terms of cognitive theory (Gitomer, 1988). It is important to understand the nature of cognitive processing involved in SAT Mathematics. Gitomer (1988) pointed out that students' errors are often linked to an inability to conceptualize a problem, to a failure to employ efficient problem-solving heuristic, and to a lack of willingness to pursue difficult problems that cannot be solved quickly.

Schoenfeld (1985) argued that some students have a view of mathematics that it is simply equivalent to the learning of algorithms. However, Gitomer (1988) developed a diagnostic test that was designed to measure knowledge, execution referred to the procedural evaluation of a problem (such as multiplying two polynomials), application involved in recognizing a procedure to execute for a given problem, decomposition processes that require decomposing a problem with multiple subgoals, and translation (that is, the process of transferring a word problem into a representation that can lead to a solution) had a strong relationship with mathematics grades.

Enright (1991) emphasized that understanding problem solving requires a description of the problem as well as a description of problem solving approaches and outcomes. Gallagher (1991) investigated sex differences on cognitive tasks for SAT Mathematics and found that female students tend to use algorithmic strategies as test-taking skills while male students tend to use a systematic trial-and-error approach regulated by some unknown reasoning. These task variables are useful for guiding an analysis of the underlying cognitive processes.

A task analysis of the SAT Mathematics, Form 8A, was performed by taking the research results mentioned above into account. This report summarizes the results of a task analysis and discusses an application of a measurement model called Rule Space (Tatsuoka, 1983) to construct a descriptive scale for SAT Mathematics. The approach is an outcome of a long-term research project supported by the Office of Naval Research, and the model actually performs individual diagnostic analyses of examinees' response patterns. The results can be used for enhancing learning, improving instruction, and remediation of examinees' weaknesses.

The model projects (or converts) examinees' item response patterns into their performance patterns on underlying cognitive tasks, which are identified by a task analysis. A set of newly converted mastery patterns of cognitive tasks (called attributes) enables one to estimate conditional probability functions for attributes (PFAs) on the SAT Scale, or IRT ability scale θ.

The report gives some tailored prototypes of performance reports suitable to various interest groups of test users. The last section discusses the generalizability of attributes across two forms of the SAT M, Section 2.

METHOD AND PROCEDURES

1. A Task Analysis of SAT Mathematics

A description of the process that led to the specification of the attributes employed in the rule space analysis is described in this Section.

1.1. A mapping sentence The cognitive requirements for solving the mathematics items of Sections 2 and 5 of SAT (form 8B administered on May 7, 1988) were specified using data from two protocols.

In order to summarize the content and process categories identified in the protocol analysis, a mapping sentence (Guttman, 1991; Tziner, 1987) was designed. The mapping sentence included 13 facets with a varying number of elements in each. Before presenting the mapping sentence, a word of caution is in order. The mapping sentence presented in Table 1.1 is a preliminary one. By no means do we contend that it is complete or exhaustive. More insight into

Insert Table 1.1 about here

the cognitive requirements underlying the SAT-M items needs to be gained by a comprehensive protocol analysis on several forms of the SAT before a complete cognitive model can be constructed.

Every item in the test can be expressed as a combination of elements from the facets of the mapping sentence. For example: Item No. 1, "If $2x - 6 = 10$, then $3x - 6 =$ ____ , (A) 0, (B) 8, (C) 11, (D) 18, (E) 24 " can be expressed in terms of the above mapping sentence as the following combination of facet elements: A3.1.1, B1, C2, D1, E2, F1, G1, H1, I2.1, J2, K3, L2, M1.

1.2. Making an incidence matrix Twenty-seven elements from the mapping sentence were selected and expressed as attributes to be used in the initial rule space analysis. Table 1.2.1 lists these attributes.

Insert Table 1.2.1 about here

An incidence matrix Q (60 items by 27 attributes) was constructed for SAT Sections 2 and 5 using the above mentioned attributes. Table 1.2.2 presents the Q matrix along with the percent correct responses for each item and values of the IRT item difficulty parameter b.

Insert Table 1.2.2 about here

For ease of referencing, Table 1.2.3 lists the items requiring each of the 27 attributes.

Insert Table 1.2.3 about here

1.3. A multiple regression analysis A multiple regression analysis was performed to predict percent correct (of 60 items) from the 27 attribute vectors. Table 1.3.1 presents the results of this analysis.

Insert Table 1.3.1 about here

As can be seen in Table 1.3.1, 83% of the variance in item difficulty (percent correct) was explained by the 27 attributes. Attributes 8, 19, 6, 3, 2, 25, 27, 11, 21, 7, 4 had the highest regression weights. The negative signs of these weights indicate that the presence of these attributes contributes to the items being more difficult. Attribute 15 had a relatively high positive weight, indicating that its presence is associated with easier items.

Based on the regression results, the initial attribute set was reduced by collapsing 10 of the content attributes into three categories and omitting five weak attributes. The reduced set of 15 attributes is presented in Table 1.3.2. Table 1.3.3 lists the 25 items of Section 2 by the reduced set of 14 attributes. (Attribute 16 is relevant to Section 5 only.)

Insert Table 1.3.2 about here

Insert Table 1.3.3 about here

1.4. Analysis of SAT M, Section 2 The incidence matrix Q for items 1-25 of Section 2 by 14 attributes (see Table 1.4.1) was subjected to multiple regression analyses for predicting item difficulties (percent correct and IRT b-values). The results of the two regression analyses are presented in Table 1.4.2.

Insert Table 1.4.1 about here

Insert Table 1.4.2 about here

As can be seen in Table 1.4.2, 83% and 91% of the variance in item difficulty (percent correct and IRT b-values, respectively) were explained by the 14 attributes. In both analyses the strongest attributes were Nos. 21, 19, 17 and 25 (analytic thinking; comprehension + application; understanding of concepts; and multiple steps toward the solution).

The Rule Space Model has recently been introduced in various ETS technical reports (Tatsuoka & Tatsuoka, 1992; Sheehan, Tatsuoka & Lewis; Birenbaum, Kelly & Tatsuoka, 1992). So a brief discussion will be given in the next section and Appendix will provide a more detailed sketch.

2. A Brief Discussion of the Rule Space Model

An alternative approach to cognitive diagnosis — in contrast to the traditional bug analyses — is the rule space model which is a probabilistic approach whose purpose is to identify the examinees' state of knowledge or cognitive states, based on an analysis of the task's cognitive requirements.

Having specified the task's cognitive requirements (also called attributes), an incidence matrix Q (K x n) (the number of attributes x the number of items) is constructed, which describes item characteristics in terms of the underlying cognitive processes involved in each item. Cognitive patterns represented by K binary elements of unobservable attributes that can be derived from the incidence matrix Q are called cognitive states (or attribute patterns). Boolean Descriptive Functions (BDFs) are used to systematically determine these cognitive states and map them into observable item score patterns (called ideal item score patterns) (Tatsuoka, 1991; Varadi & Tatsuoka, 1989). It is assumed that an item can be answered correctly if and only if all the attributes involved in the item have been mastered. Unobservable performances on the attributes can be viewed analogously to an unobservable electric current running through various switches if they are closed. A closed switch corresponds to an attribute that has been mastered. All switches in a circuit must be closed in order for the current to go through. The cognitive states are represented by a list of mastered/not mastered (or "can/cannot") attributes. The increase of the number of states is combinatorial, but Boolean algebra is a useful tool for dealing with the problem of combinatorial explosion. Boolean algebra, which has been widely used for explaining various properties of electricity and combinatorial circuits have been utilized within the rule space framework for explaining the cognitive requirements underlying test performances.

Once the cognitive states (ideal-item-score patterns) are determined, the actual data are considered. The task now is to map the actual item response patterns of the examinees onto the cognitive states, i.e., to find the ideal-item-score pattern closest to the student's actual response pattern. Since the performance on test items usually includes slips or random errors, the observed item-response patterns are likely to deviate to some extent from the ideal-item-score patterns represented by the various cognitive states. Thus one is faced with a pattern classification problem which is handled by the rule space model (Tatsuoka & Tatsuoka, 1989). The model formulates the classification space and procedures. Item Response Theory (IRT) is utilized for formulating the classification space, which is a Cartesian product space of IRT ability θ and a variable ζ which measures the unusualness of item score patterns (Tatsuoka, 1984, Tatsuoka & Linn, 1983). The cognitive states as well as the students' item response patterns are mapped as points in the classification space by computing their θ and ζ values. Tatsuoka (1990) has shown that the swarm of mapped "fuzzy" points of students' item-response patterns follows approximately a multivariate normal distribution with the centroid being a given cognitive state. Bayes' decision rules are applied for the final classification and for computation of misclassification probabilities.

Once this classification has been carried out, one can indicate with a specified probability level which attributes a given examinee is likely to have mastered or failed to master. If classification rates are as high as 80 % or above, then the attribute mastery patterns can be used for statistical analyses. For example, a factor analysis can be applied to examine the dimensionality of attributes, or a discriminant analysis can be used for investigating subgroup differences if the demographic information is available. Similar to the estimation of Item Response Curves from the item response patterns, it is possible to investigate the conditional probability functions of the attributes defined on the SAT scale or IRT θ.

3. The Classification Results of SAT M, Section 2

A computer program, BUGLIB, classified 2335 examinees who took the SAT M, Form 8A, into one of 600 cognitive states. Since the squared Mahalanobis distance in this case follows a Chi-square distribution with 7 degrees of freedom, $\chi^2 = 2.76$ ($p=.01$) is set as the first criterion for whether or not X can be classified into a cognitive state. It turned out that 98 % of the 2335 examinees qualified according to the first criterion, and were thus classified into one of 600 cognitive states. The examinees who were not classified are mostly very high scoring students and their θ values are larger than 2.5. After Bayes' rule was applied for the final classification, 468 cognitive states become non-empty, with 136 states having one examinee classified, 64 states having 2 classified, 32 states having 3 classified, 26 states having 4, 14 states having 5, 13 states having 6, 13 having 7, 8 having 8, and 5 having 9. The states to which at least 11 examinees were classified are listed in Table 3.1. One hundred thirty two examinees are

Insert Table 3.1 about here

classified into State 472, which is characterized by the deficiency of attributes 2,19,21, and 25. State 2, which is characterized by the lack of skill 21, has 180 examinees classified.

The θ-values and ζ-values for the cognitive states which are listed in Table 3.1 are given in Table 3.2.

Insert Table 3.2 about here

Table 3.2 indicates some interesting trends for the lack of skills across various levels of θ. For example, the low-ability examinees missed Attributes 1, 3 and 21 (Arithmetic, advanced algebra and analytical thinking skill) while high-ability examinees missed Attribute 21 and could do most content areas except for advanced algebra. Probability Functions for the attributes (PFAs) will provide us trends of the 14 attributes across θ.

However, before discussing PFAs, simple descriptive statistics of the 14 attributes are summarized. Table 3.3 shows the summary statistics of

Insert Table 3.3 about here

the 14 attributes and θ, ζ and five generalized ζs. Attributes 21, 19, and 3 are difficult attributes while Attributes 18, 6, 15, and 23 are easy ones. The means of θ, ζ and five generalized ζs are closer to zero and the standard deviations are almost 1 as their theoretical means and standard deviations indicate. The correlations of θ with the 14 attributes range from .05 (Attribute 23) to .30 (Attribute 3). The correlations of ζ with the 14 attributes are between .14 (Attribute 19) and -.34 (Attribute 24), except for that of Attribute 21 which is .58. The value .58 indicates that the behavior of Attribute 21 is unusual, and examinees with unusual response patterns tend to have the mastery score of one for this attributes. The dimensionality of the 14 attributes is tested by computing the eigenvalues of the correlation matrix of 14 attributes. The results of Principal Component analysis indicated that the 14 attributes are not unidimensional. Of course we could have examined the dimensionality with better statistics such as Stout's method (Stout, 1987), but we will leave it for a future work.

4. Probabilities for Attributes

When examinees' item response patterns are classified into particular states, their corresponding attribute mastery patterns are then known. We use the attribute mastery patterns to estimate probability functions for the attributes (PFAs). PFAs are the conditional probability functions defined on θ, and they describe the basic characteristics of the behavior for the attribute variables. By looking at the graphs of PFAs, one can see the relationships between the performances on the attributes and the IRT θ-scale or SAT scale. Each attribute should have its unique curve, different from those of the others. By comparing two curves, one can see which attribute is harder. They may intersect at some point, with abscissa θ_0. In that case there is an interaction between item difficulty and ability level exists. Unlike Item Response Theory, we do not restrict the possible forms of the conditional probability functions by assuming that they belong to a prespecified family of parametric functions such as logistic or normal ogive. Since our intention is to "let the data speak for themselves," a nonparametric estimation approach is adopted in this report.

4.1 Non-parametric regression estimates as probability functions for attributes Non-parametric estimation of the unknown density function f from a plot of frequencies, the histogram, has been well investigated by many statisticians (Hardle, 1991; Scott, 1985). Several psychometricians have applied these techniques to estimating Item Response Curves, which are not density functions (Ramsay, 1991; Mokken & Lewis, 1982, Lewis, 1990).

Instead of plotting a histogram of observed frequencies an PFAs for a particular attributes constructed by first classifying examinees into bins b_j based on their estimated θ values and then computing the proportion of examinees in each bin who have been classified as having mastered the attribute. These proportions are then plotted against θ and smoothed. Alternatively, examinees may be classified into bins based on their SAT Scale scores. The PFA would then be plotted as a function of the SAT Scale score.

4.2 Results Using SPLUS on a SUN SPARC station, a computer program for estimating Attribute was written. In the program, examinees were classified into one of 12 bins based on their SAT Scale score. Figure 4.1 contains the resulting PFAs for each of the 14 attributes.

Insert Figure 4.1 about here

The curves in Figure 4.1 are not well smoothed yet, but they should suffice for the purpose of introducing the concept of PFAs for an attribute variable to the reader of this report. Improved methods for estimation of PFAs and estimation of confidence intervals will be given in a future report.

4.3 Interpretation Once examinees' SAT Scale scores are known, their probabilities of mastering each of the attributes can be read off the curves given in Figure 4.1. As an example, Table 4.3.1 provides attribute mastery probabilities for the first eight examinees in the data set.

Insert Table 4.3.1 about here

Examinee 5 has a very high SAT Scale score, and he/she is doing very well on most attributes except for 3,17,19, and 21. His SAT Scale score is almost as high as Examinee 7, but his attribute sores are much lower for 17, 19, and 21. By looking at the profile of each student, one can get useful information for remediation planning. Alternatively, by looking at the unit of classrooms or schools, one can make useful curriculum design, or evaluation of the past instruction or planning.

4.4 Percentile scores Mokken & Lewis (1982) developed a non-parametric, Bayesian IRT model which is based on the Mokken-scale, and Lewis (1990) developed an algorithm for estimating the x% threshold for a monotone regression function. His program MonoReg2 (1990) computes the posterior mean estimate of a percent point of interest. For example, Attribute 19 has 546 for the 50% point, 277 for 25% point and 760 for the 75% point.

Insert Figure 4.4.1 about here

Figure 4.4.1 shows the empirical curve for Attribute 19 and posterior median estimates of selected values of the corresponding theoretical function (connected by straight lines). With this method, a desired percent point and its corresponding SAT scale score can be obtained. A summary table could then be prepared, describing the location of the attribute on the SAT scale.

4.5. Enhancing score reports Enhancing a score report can be done by utilizing the probability of successful performance on each attribute, together with the information obtainable from item-level analyses such as computing IRT conditional probabilities on θ. The incidence matrix Q can be used to retrieve a meaningful subset of items that involves, say, "test taking skills" or "higher level thinking skills". Therefore, the results from the rule-space model can be used for preparing a variety of reports that are tailored to different groups of test users. The purposes for using the test reports may vary among different groups of test users.

The optimal use of test results should be recommended. If the audience is higher educational institutes, test results are used for selection or placement of applicants. Individual examinees in high schools may use test results for guiding themselves for further study or remediation, and teachers for evaluating their instructions, for designing of curricula and future instruction planing. The test results can also be used for preparing reports for group performance. Summary statistics of attribute-level performance as well as item-level performance can be useful for schools, for various districts and state offices of education. The following figure gives an example of what we can offer to the test users.

Insert Figure 4.5.1 about here

The data banks available for enhancing scoring reports consist of four parts: 1) The score matrix, each row of which contains a student ID, an item response pattern, a θ–value, a ζ–value (an index for measuring atypicality of a response pattern) and an attribute–mastery pattern; 2) the incidence matrix; 3) the probability matrix for indicating each item's success rate at various levels of θ and SAT scale; and 4) the probability matrix for indicating each attribute's mastery rate at various levels of θ and SAT scale. The information mentioned above, together with demographic information can provide test users with a variety of reports tailored to different groups based on their needs and interests. The following figures show prototypes of reports that can be assembled from the database (see Appendix).

Figures 4.5.2, and 4.5.3

Figures 4.5.2 and 4.5.3 are prepared for examinees who are interested in understanding their weaknesses and strengths, while Figure 4.5.4 is for a class room teacher who is interested summary statistic and class evaluation. Rearranging the probability matrix by the order of total scores and item difficulties enables teachers and administrators to identify possible problem areas (Birenbaum, 1992).

5. Are the 14 Attributes Invariant Across Different Forms of SAT Mathematics ?

A replication study was carried out by applying the 14 attributes to a different SAT form (0A March, 1990). Table 5.1 presents the incidence matrix

Insert Table 5.1 about here

for the 25 items of Section 2 of that form by the 14 attributes, along with the item difficulties (percent correct).

Insert Table 5.2 about here

Table 5.2 presents the regression results for predicting item difficulties of the 25 items of Form 0A (section 1) from the 14 attributes.

As can be seen in the table, 91% of the variance in item difficulty (percent correct) was explained by the 14 attributes. The strongest attributes were Nos. 3, 21 20 and 25 (advanced algebra; analytic thinking; reasoning; and multiple steps toward the solution, respectively).

Upon reviewing the items of Form 0A an additional attribute was introduced to the original set, namely, Attribute 26 "changing the unit of measurement". That attribute appeared in items 10 and 18 of Form 0A, Section 2. The incidence matrix Q for the 15 attributes appears in Table 5.3.

Insert Table 5.3 about here

For ease of referencing, Table 5.4 lists the attributes involved in each of the 25 items (Form 0A, Section 2) and Table 5.5 lists the items that involve each of the 15 attributes.

Insert Table 5.4 and 5.5 about here

A regression analysis of the incidence matrix with the additional attribute (No. 26) is presented in Table 5.6.

Insert Table 5.6 about here

As can be seen in the table, 94% of the variance in item difficulties is explained by the 15 attributes. The strongest attributes in this analysis are:

26, 3, 20 ,6, and 25 (changing the unit of measurement; advanced algebra; reasoning; elementary geometry; and multiple steps toward the solution).

This routine multiple regression analysis suggests that the attributes valid for one form may be valid for another form. However, it does not give any direct information for assurance that an estimated PFA for Attribute A_k involved in one form will be very close to the estimated PFA from a different form. If the construction of parallel test forms were to be based on the matching of attributes across different forms, then our concern for invariance of PFAs across the forms may not be so important. However, the current practice of test construction procedures do not consider the underlying cognitive attributes of test performance. The procedures emphasize matching of content domains although SAT Mathematics tests is designed for measuring reasoning rather than for measuring the competency in content domains.

Discussion

The influence of SAT Verbal and Mathematics tests on American education is so noticeable that maximizing the amount of information obtainable from the test scores, and searching for ways to utilize such information optimally are very important. This study introduced a new way to construct a proficiency scale by applying the rule space model.

The rule space model is a symbolic parametric model in which the performances on unobservable cognitive tasks are inferred from observable item scores. The inferred attribute-mastery patterns are used for estimating Attribute Characteristic Curves defined on the θ or SAT scale. The proficiency scale in this paper is derived from these PFAs.

Statistical matters such as construction of confidence intervals for PFAs and further improvement of non-parametric estimation methods are not discussed in this paper. The technical aspect of obtaining percentile scores from PFA should also be sought in a future paper. A multidimensional rule space has been introduced for the first time in this paper, but technical details of the multidimensional space will be discussed elsewhere in the near future.

A list of the 14 attributes should be examined more carefully before the proficiency scale for SAT M is to be used in practice. The regression analysis and the rule-space classification don't necessarily provide the best unique set of attributes. Instead, they can indicate whether or not these attributes provide a useful representation of the underlying cognitive processes of the test. There may exist other sets of attributes that are as good as the original 14 attributes. Further investigation on the determination of the optimal set of attributes is needed.

References

Beaton, A. E. (1988). Expanding the new design: The NAEP 1985-1986 technical report. (No.17-TR-20) Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Birenbaum, M. (1992). Feedback from tests- A practical handbook for analyzing test results. (include a computer software for analyzing test results, using IBM PCs) Tel-Aviv: Koltar. (in Hebrew).

Birenbaum, M., Kelly, A. E. & Tatsuoka, K. K. (in press). Applying the rule space model to diagnose students, knowledge states in exponents. Journal of Teaching Mathematics.

Birkoff, B. (1970). Lattice theory. New York, NY: MacMillan

Dibello, L. V., & Baillie, R. J. (1991). Separating points in Rule space: A contribution to diagnostic testing theory. Technical Report, Urbana, IL: University of Illinois, CERL.

Enright, M. (1991). A Cognitive analysis of GRE Algebra word problems. Unpublished paper.

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

Forsyth, R. A. (1991). Do NAEP scales yield valid criterion-referenced interpretation? Educational Measurement: Issues and Practice, 10(3), 3-9.

Gallagher, A. (1991). Strategy use on multiple-choice and free-response items: An analysis of sex differences among high scoring examinees on the SAT-M. Technical Report RR-Educational Testing Service, Princeton; NJ.

Gitomer, D. H. (1988). Sources of difficulty in calculus performance. A paper presented at AERA in New Orleans.

Guttman, L. (1991). Louis Guttman on facet theory: excerpts from unfinished writings. Jerusalem: The Israel Academy of Sciences and Humanities and the Hebrew University of Jerusalem.

Hardle, W. (1991). Smoothing Techniques with implementation in S. Springer-Verlag, New York; NY.

Kim, S. Hoon. (1990). Classification of item-response patterns into misconception groups. Unpublished doctoral dissertation. University of Illinois, Champaign, IL.

Kim, S. Ho. (1992). Issues in building influence diagrams of cognitive attributes and item responses. Technical Report, Educational Testing Service, Princeton.

Lachenbruch, p. A. (1975). Discriminant Analysis. New York: Hafner Press.

Lewis, C. (1990). Notes on computing algorithm used in monotone regression program. Unpublished manuscript.

Lewis, C. (1992). MONOREG2. Unpublished computer program.

Marco. G. L., Crone, C. R., Braswell, J. S., Curley, W. E., & Wright, N. K. (1990). Trends in SAT content and statistical characteristics and their relationship to SAT predictive validity. Research Report, Princeton, NJ: Educational Testing Service.

Mislevy, R. J., Sheehan, K. M., and Wingersky, M. (1992). How to equate with little or no data. Research report, Princeton, NJ: Educational Testing Service.

Mokken, R. J., & Lewis, C. (1982). A nonparametric approach to the analysis of dichotomous item responses. Applied Psychological measurement, 6(4), 417-430.

Ramsay, J. O. (1991). Kernel smoothing approaches to nonparametric item characteristic curve estimation. Psychometrika, 56(4), 611-630.

Rock, D. A., & Johnson, G. (1989). Empirical Anchoring of the SAT-M. Unpublished manuscript.

Scott, D. W. (1985). Averagedshifted histograms: effective non-parametric density estimators in several dimensions. Annals of Statistics, 13, 1024-1040.

Sheehan, K., Tatsuoka, K. K., & Lewis, C. (1993). A diagnostic classification model for document processing skills. Research Report, Princeton, NJ: Educational Testing Service.

Schoenfeld, A. H. (1985). Mathematical problem solving. Orlando: Academic Press.

Stout, W. F. (1987). A nonparametric approach for assessing latent trait dimensionality. Psychometrika, 52, 589-617.

Tatsuoka, K. K., & Linn, R. L. (1983). Indices for detecting unusual patterns: Links between two general approaches and potential applications. Applied Psychological Measurement, 7(1), 81-96.

Tatsuoka, K. K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. Journal of Educational Measurement, 20(4).

Tatsuoka, K. K. (1984). Caution indices based on item response theory. Psychometrika, 49(1).

Tatsuoka, K. K. (1985). A probabilistic model for diagnosing misconceptions in the pattern classification approach. Journal of Educational Statistics, 12(1).

Tatsuoka, K. K. (1990). Toward an integration of item-response theory and cognitive error diagnoses. In N. Frederiksen, R. L. Glaser, A. M. Lesgold, & M. G. Shafto (Eds.), Diagnostic monitoring of skill and knowledge acquisition. Hillsdale, NJ: Erlbaum.

Tatsuoka, K. K. (1991). Boolean algebra applied to determination of universal set of knowledge states. Research report RR-91-44-ONR, Princeton, NJ: Educational Testing Service.

Tatsuoka, K. K. (1992). Item construction and psychometric models appropriate for constructed responses. In R. E. Bennett & W. C. Ward (Eds.), Construction Versus Choice in Cognitive Measurement. Hillsdale, NJ: Erlbaum.

Tatsuoka, K. K., & Tatsuoka, M. M. (1987). Bug distribution and pattern classification. Psychometrika, 52(2), 193-206.

Tatsuoka, K. K., & Tatsuoka, M. M. (1992). A psychometrically sound cognitive diagnostic model: Effect of remediation as empirical validity. Research report, Princeton, NJ: Educational Testing Service.

Tatsuoka, M. M., & Tatsuoka, K. K. (1989). Rule space. In Kotz and Johnson (Eds.), Encyclopedia of statistical sciences. New York: Wiley.

Tziner, A. E. (1987). The facet analytic approach to research and data processing. New York: Peter Lang.

Varadi, F. & Tatsuoka, K. K. (1989). BUGLIB, unpublished software.

Table 1.1

A Mapping Sentence For SAT-M

In order to solve item x, which represents a task with the following characteristics:

A
content

1. Arithmetic
   - 1) basic operations with whole numbers
   - 2) signed numbers operations
   - 3) fractions, decimals
   - 4) square root, exponents
2. Mathematics
   - 1) properties of numbers, combinatorial, inequality unit of measurement
3. Algebra
   - 1) basic
     - (1) linear equations
     - (2) simultaneous linear eq
   - 2) advanced
     - (1) quadratic eq.
     - (2) functions
4. Geometry
   - 1) elementary
     - (1) lines, rectangles
     - (2) triangles
     - (3) circles
   - 2) analytic
5. Statistics
   - 1) Probability

B context
1. regular math
2. quantitative comparisons

C setting
1. concrete
2. abstract

D question type
1. routine
2. non routine

E language of presentation
1. verbal (word problems)
   - 1) realistic context
   - 2) imaginary context
2. numerical
3. spatial

F Q. Structure Logic (if...then)
1. yes
2. no

G answer type
1. exact number
2. approximation
3. variable

H response format
1. multiple choice
2. constructed (grid)
3. stem includes options

I susceptibility to "test wiseness"
1. low
2. high
   - 1) options can be used to get the answer
   - 2) can be solved intuitively / by example
   - 3) visual solution possible

and which the solution process involves:

J no. of steps
1. one
2. two
3. three

K requiring to read
1. charts
2. figures
3. math notations

L calculator
1. not needed
2. can be helpful
3. needed

the examinee has to demonstrate the following:

M
Processes

1. Application of simple rules/algorithms (perform computations)
2. Comprehension + application of rules/theorems/definitions/principals/laws
3. Translation from one mode to another
4. Creation of an equation with
   - 1) one unknown
   - 2) more than one unknown
5. Analytic thinking
   - 1) decomposition of a simple problem
   - 2) decomposition of a complex problem

   and restructuring
6. Reading comprehension
   - 1) general
   - 2) specific terminology

Table 1.2.1

SAT-M 27 Attributes

Attribute No.	Attribute's Description
	A. Content related attributes
1.	Arithmetics (+ - X : ; signed #s; # line; ( ); factoring, properties of #s; combinatorial).
4.	Arithmetics - fractions (+ ratio; decimals; probability; %)
5.	Arithmetics - exponents (+ sq. root).
22.	Arithmetic - inequality.
2.	Algebra - linear equations (+ simultaneous linear).
3.	Algebra - quadratic equations.
27.	Algebra - Functions (+ relationships between number and symbols).
6.	Geometry - lines; rectangles.
7.	Geometry - triangles.
8.	Geometry - Circles.
26.	Analytic geometry/reading charts.
9.	Measurement related concepts.
10.	Nonroutine problems (nonconventional).
11.	Language of presentation: Verbal (Word problem).
12.	Language of presentation: Numerical (math notations)
13.	Language of presentation: V + Spatial (figure given).
14.	Language of presentation: V + Spatial (figure to be drawn).
15.	Logic (if...then).
16.	Quantitative comparisons.
	B. Process Related Attributes
17.	Understanding of the meaning of concepts.
18.	Application of simple rules/algorithms (SOLVE: perform computations).
19.	Comprehension + application of rules/theorems (chooses and applies correctly).
20.	Reasoning (creates an equation).
21.	Analytic thinking, cognitive restructuring (higher mental processes).
23.	Reading comprehension (+ follow instructions; math/geometry terminology).
24.	Test-wiseness (solves intuitively; by example; goes backwards from the given answers).
25.	Number of steps in the solution > 1

Table 1.2.2

Incidence Matrix Q for 27 Attributes and 60 SAT-M Items

Attributes

Item No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	% Correct	b Value
01	0	1	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0	1	0	0	81	-1.513
02	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	89	-2.037
03	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	1	1	0	88	-1.554
04	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	1	1	0	76	-1.214
05	0	0	0	1	0	0	0	0	0	0	0	1	0	0	1	0	0	1	0	0	0	0	0	1	0	0	0	80	-1.021
06	0	0	0	0	1	0	0	0	0	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	85	-1.371
07	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	1	1	0	0	0	72	-2.916
08	1	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0	1	0	0	1	1	0	0	0	59	-0.538
09	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	76	-1.056
10	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	1	0	0	53	.157
11	1	1	0	1	0	0	0	0	0	1	1	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0	0	60	-.433
12	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	64	-.218
13	0	0	1	1	1	0	0	0	0	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	58	-.449
14	0	0	0	0	0	1	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0	1	0	0	0	0	72	-1.190
15	0	0	0	1	0	0	0	0	0	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0	1	0	0	48	.777
16	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	55	.328
17	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	1	0	1	0	0	38	.619
18	0	0	0	0	0	0	0	1	0	0	0	0	1	0	1	0	0	0	1	1	0	0	0	0	1	0	0	35	.881
19	0	1	0	1	0	0	0	0	0	1	1	0	0	0	1	0	0	1	0	1	0	0	0	0	1	0	0	33	.842
20	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0	1	1	1	0	0	32	.805
21	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0	1	0	1	0	1	27	.840
22	0	0	0	0	0	0	0	1	0	0	1	0	0	0	1	0	0	1	0	1	0	1	1	0	1	0	0	23	.899
23	0	1	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1	0	1	1	0	1	0	1	0	0	19	1.534
24	0	0	0	0	0	1	1	0	0	0	0	0	1	0	1	0	0	1	1	1	1	0	0	0	1	0	0	14	1.788
25	0	1	0	1	0	0	0	0	1	1	0	0	1	0	0	0	0	1	0	1	1	0	0	0	1	0	0	17	1.351
26	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	92	-2.977
27	1	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0	0	92	-1.888
28	0	0	0	0	0	0	0	0	1	1	1	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0	0	80	-2.466
29	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	79	-1.265
30	0	0	0	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	68	-.662
31	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1	0	1	0	0	1	0	0	1	0	0	0	1	75	-1.182
32	0	0	0	0	0	1	0	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	64	-.765
33	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	74	-.217
34	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	76	-.920
35	0	0	0	1	0	0	0	0	1	0	0	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	67	-.511
36	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	72	-.372
37	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	65	-.215
38	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	1	0	83	-1.437
39	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1	1	0	73	-1.246
40	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	79	-1.081
41	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	1	1	0	0	0	0	1	0	1	0	0	49	.175
42	1	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	1	1	0	0	0	0	1	0	1	0	0	45	.254
43	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	67	-.741
44	1	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	61	-.509
45	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0	1	0	0	0	0	0	56	-.104
46	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	61	-.094
47	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	1	0	0	0	0	49	.543
48	0	0	0	1	0	0	0	0	1	0	0	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	35	.985
49	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	1	1	0	0	0	1	0	1	0	1	0	0	26	2.042
50	0	1	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0	1	0	0	1	0	0	0	1	0	0	28	1.303
51	0	1	1	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	30	1.378
52	0	0	0	0	0	1	1	0	0	0	0	0	0	1	0	1	0	0	1	0	0	0	0	0	1	0	0	09	1.939
53	0	0	0	0	0	1	0	0	1	0	1	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0	0	57	.577
54	1	1	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	1	1	1	0	0	0	54	-.147
55	0	0	0	0	0	1	0	0	0	1	0	0	0	1	0	1	0	0	0	0	1	0	0	0	1	0	0	41	1.150
56	1	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	1	1	0	1	1	1	1	0	22	1.286
57	0	1	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	15	1.597
58	1	1	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	1	0	0	1	1	0	0	26	1.362
59	0	1	0	1	0	0	0	0	0	1	1	0	0	0	0	0	1	1	0	0	0	0	0	0	1	0	0	09	1.708
60	0	0	0	1	0	0	0	1	0	0	0	0	1	0	1	0	0	1	1	0	1	0	1	0	1	0	0	07	1.805

Note:
Items 1-25 are from section 2 and items 26-60 are from section 5

Table 1.2.3

Items Required in Each of the 27 Attributes

Attribute	Items (1-60)
01	3, 4, 8, 9, 11, 16, 20, 26, 27, 23, 38, 39, 40, 41, 42, 44, 54, 56, 58
02	1, 11, 19, 23, 25, 43, 50, 51, 54, 57, 58, 59
03	13, 51
04	2, 5, 10, 11, 13, 15, 19, 25, 33, 35, 36, 45, 47, 48, 59, 60
05	2, 6, 13, 21, 29
06	14, 17, 24, 30, 32, 37, 49, 52, 53, 55
07	7, 10, 24, 34, 46, 52, 57
08	18, 22, 60
09	12, 25, 28, 35, 48, 53
10	8, 11, 12, 16, 19, 20, 23, 25, 28, 30, 31, 32, 41, 42, 44, 50, 54, 55, 58, 59
11	8, 11, 12, 16, 19, 20, 22, 23, 27, 28, 34, 41, 42, 47, 53, 54, 56, 58, 59
12	1, 2, 5, 6, 9, 13, 15, 21, 26, 29, 31, 33, 35, 36, 40, 43, 44, 45, 48, 50, 51
13	3, 4, 7, 10, 18, 24, 25, 30, 32, 37, 38, 39, 46, 57, 60
14	14, 17, 49, 52, 55
15	1, 5, 6, 8, 11, 13, 14, 15, 18, 19, 22, 24, 27, 28, 31, 53, 54, 56, 60
16	33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52
17	14, 21, 31, 32, 34, 35, 41, 42, 45, 47, 48, 49, 55, 59
18	1, 2, 3, 4, 5, 6, 9, 10, 11, 12, 13, 15, 19, 21, 22, 23, 24, 25, 26, 27, 29, 36, 37, 40, 41, 42, 43, 46, 50, 53, 57, 59, 60
19	10, 18, 24, 52, 57, 60
20	7, 8, 11, 16, 18, 19, 20, 22, 23, 24, 25, 27, 28, 31, 51, 53, 56
21	17, 23, 24, 25, 49, 50, 55, 56, 58, 60
22	22, 33, 45, 54
23	3, 7, 8, 14, 17, 20, 21, 22, 23, 30, 31, 32, 41, 42, 44, 47, 49, 54, 56, 60
24	5, 7, 8, 20, 28, 29, 51, 54, 56, 58
25	1, 3, 4, 10, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 39, 41, 42, 49, 50, 52, 55, 56, 58, 59, 60
26	3, 4, 38, 39, 56
27	21, 31

Table 1.3.1

Multiple Regression Results: Predicting Item Difficulties from 27 Attributes.

Attribute	b	SEb	t
A1	.02	6.61	.01
A2	-19.54	6.41	-3.05***
A3	-21.67	11.36	-1.91*
A4	-11.51	5.47	-2.10**
A5	-4.44	8.84	-.50
A6	-21.80	10.10	-2.16**
A7	-12.35	11.25	-1.10
A8	-29.39	16.56	-1.78*
A9	-10.46	7.33	-1.43
A10	-2.78	6.27	-.44
A11	-14.92	12.10	-1.23
A12	-8.11	12.67	-.64
A13	3.03	10.55	.29
A14 (-)	0.00	0.00	0.00
A15	10.57	5.51	1.92*
A16	-7.90	5.38	-1.47
A17	-4.18	6.38	-.66
A18	-1.92	5.51	-.35
A19	-26.93	10.84	-2.48**
A20	-4.32	6.01	-.72
A21	-13.44	6.28	-2.14**
A22	-3.76	8.51	-.44
A23	-8.57	5.10	-1.68
A24	-8.45	6.59	-1.28
A25	-16.71	5.44	-3.08***
A26	-4.75	11.72	-.41
A27	-15.87	13.86	-1.15

$R^2 = 0.83$

Note:
A1 to A27 : Initial set of attributes (see Table 1.2.1).
Y : Percent of correct responses (as reported in "Taking the SAT 1990-91).
Number of items: 60 (1-25 from Section 2; 26-60 from Section 5).
(-) Parameter not estimated (A14 is a linear combination of A11, A12, A13)
* $p<.10$. ** $p<.05$. *** $p<.01$.

Table 1.3.2

The Reduced set of 15 Attributes

Attribute No.	Attribute's Description
	A. Content related attributes
1.	Arithmetics (including content of attributes: 1, 4, 5, 22).
2.	Elementary Algebra (including content of attributes: 2, 27).
3.	Advanced Algebra.
6.	Elementary Geometry (including content of attributes: 6, 7, 8, 26).
11.	Word problems.
15.	Logic (if...then).
16.	Quantitative comparisons*.
	B. Process Related Attributes
17.	Understanding of the meaning of concepts.
18.	Application of simple rules/algorithms (SOLVE: perform computations).
19.	Comprehension + application of rules/theorems (chooses and applies correctly).
20.	Reasoning (creates an equation).
21.	Analytic thinking, cognitive restructuring (higher mental processes).
23.	Reading comprehension (+ follow instructions; math/geometry terminology).
24.	Test-wiseness (solves intuitively; by example; goes backwards from the given answers).
25.	Number of steps in the solution > 1

* Applies to section 5 only.

Table 1.3.3

The 25 items of section 2 Listed by the Reduced set of 14 Attributes

Item	Attribute
01	2, 15, 18, 25
02	1, 18
03	1, 6, 18, 23, 25
04	1, 6, 18, 25
05	1, 15, 18, 24
06	1, 15, 18
07	6, 20, 23, 24
08	1, 11, 15, 20, 23,24
09	1, 18
10	1, 6, 18, 19, 25
11	1, 2, 11, 15, 18, 20
12	11, 18
13	1, 3, 15, 18
14	6, 15, 17, 23
15	1, 15, 18, 25
16	1, 11, 20, 25
17	6, 21, 23, 25
18	6, 15, 19, 20, 25
19	1, 2, 11, 15, 18, 20, 25
20	1, 11, 20, 23, 24, 25
21	1, 2, 17, 18, 23, 25
22	1, 6, 11, 15, 18, 20, 23, 25
23	2, 11, 18, 20, 21, 23, 25
24	6, 15, 18, 19, 20, 21, 25
25	1, 2, 18, 20, 21, 25

Table 1.4.1

Incidence Matrix Q for 14 Attributes by 25 Items and the Item Parameters

(IRT: a's, b's) and Percent Correct

Item No.	1	2	3	6	11	15	17	18	19	20	21	23	24	25	IRT a's	IRT b's	% Correct
1	0	1	0	0	0	1	0	1	0	0	0	0	0	1	.685	-1.518	81
2	1	0	0	0	0	0	0	1	0	0	0	0	0	0	.833	-1.938	89
3	1	0	0	1	0	0	0	1	0	0	0	1	0	1	1.089	-1.499	88
4	1	0	0	1	0	0	0	1	0	0	0	0	0	1	.593	-1.285	76
5	1	0	0	0	0	1	0	1	0	0	0	0	1	0	.855	-1.298	80
6	1	0	0	0	0	1	0	1	0	0	0	0	0	0	1.274	-1.309	85
7	0	0	0	1	0	0	0	0	0	1	0	1	1	0	.263	-2.180	72
8	1	0	0	0	1	1	0	0	0	1	0	1	1	0	.491	-.580	59
9	1	0	0	0	0	0	0	1	0	0	0	0	0	0	.902	-1.036	76
10	1	0	0	1	0	0	0	1	1	0	0	0	0	1	.869	-.066	53
11	1	1	0	0	1	1	0	1	0	1	0	0	0	0	.855	-.480	60
12	0	0	0	0	1	0	0	1	0	0	0	0	0	0	.692	-.566	64
13	1	0	1	0	0	1	0	1	0	0	0	0	0	0	.847	-.452	58
14	0	0	0	1	0	1	1	0	0	0	0	1	0	0	.677	-1.134	72
15	1	0	0	0	0	1	0	1	0	0	0	0	0	1	.510	.039	48
16	1	0	0	0	1	0	0	0	0	1	0	0	0	1	.593	-.245	55
17	0	0	0	1	0	0	0	0	0	0	1	1	0	1	.725	.422	38
18	0	0	0	1	0	1	0	0	1	1	0	0	0	1	.614	.645	35
19	1	1	0	0	1	1	0	1	0	1	0	0	0	1	.664	.694	33
20	1	0	0	0	1	0	0	0	0	1	0	1	1	1	.574	.772	32
21	1	1	0	0	0	0	1	1	0	0	0	1	0	1	.858	.860	27
22	1	0	0	1	1	1	0	1	0	1	0	1	0	1	.993	.987	23
23	0	1	0	0	1	0	0	1	0	1	1	1	0	1	.464	2.035	19
24	0	0	0	1	0	1	0	1	1	1	1	0	0	1	.501	2.333	14
25	1	1	0	0	0	0	0	1	0	1	1	0	0	1	.536	1.944	17

Table 1.4.2

Multiple Regression Results: Predicting Item Difficulties for Items 1-25 from 14 Attributes.

	Proportion Correct				IRT b-values			
Attribute	b	SEb	β	t	b	SEb	β	t
25	-16.06	8.53	-.33	-1.88	.85	.31	.35	2.72*
23	-4.38	12.85	-.09	-.43	.21	.47	.08	.45
21	-36.60	12. 52	-.57	-2.92*	2.35	.46	.70	5.10**
11	-13.84	11.94	-.27	-1.16	.84	.44	.32	1.92
03	-21.46	17.77	-.18	-1.21	.75	.65	.12	1.14
15	-2.81	7.24	-.06	-.39	.26	.27	.11	.98
02	3.69	12.89	.07	.29	-.42	.47	-.15	-.89
01	-5.29	9.45	-.10	-.56	.36	.35	.14	1.05
24	-.29	13.59	-.00	-.02	-.16	.50	-.05	-.32
19	-19.39	15.42	-.27	-1.26	1.26	.57	.33	2.22
18	-1.44	11.56	-.03	-.13	.23	.43	.09	.55
17	-26.97	17.40	-.31	-1.55	1.35	.64	.30	2.10
06	6.61	12.48	.11	.45	-.57	.46	-.22	-1.24
20	-12.54	13.43	-.26	-.93	.37	.49	.15	.76
a	89.00	12.59			-2.06	.46		
R^2	.83				.91			
R^2adj.	.59				79			

* p<.05 ; ** p< .001

Table 3.1 A list of Cognitive States in which at least Five Percent of Examinees are Classified (N = 2334)

Cognitive States	Frequency	Attribute Mastery Pattern	Attributes not mastered
		1111122222	
		12361578901345	
1	19	11111111111111	
2	180	11111111110111	21
4	32	11111101111111	17
5	18	11011111111111	3
6	94	11011111110111	3, 21
8	11	11011101110111	3,17, 21
9	37	11111111011111	19
10	46	11111111010111	19, 21
12	28	11111101010111	17,19,21
14	30	11011111010111	3,19,21
22	18	11010111110111	3,11,21
28	12	11110101010111	11,17,19,21
30	38	11010111010111	3,11,19,21
34	17	11111111100111	20,21
66	87	11111111110101	21,24
70	11	11011111110101	21,24
126	25	11111101110011	21,23
128	14	11011101110011	3,17,21,23
138	40	11111101010011	17,19,21,23
140	16	11011101010011	3,17,19,21,23
215	12	01011111111111	1,3
217	30	01011111110111	1,3,21
218	14	01011111010111	1,3,19,21
220	13	01011101011111	1,3,17,19
221	24	01011101110111	1,3,17,21
222	22	01011101010111	1,3,17,19,21
253	43	10111111110111	2,21
257	15	10011111110111	2,3,21
261	30	10111111010111	2,19,21
268	11	10110111111111	2,11
269	31	10110111110111	2,11,21
273	12	10010111110111	2,3,11,21
277	13	10110111010111	2,11,19,21
468	67	11111111010110	19,21,25
469	18	11111101010110	17,19,21,25
472	132	10111111010110	2,19,21,25
473	25	10111101010110	2,17,19,21,25
474	32	10011111010110	2,3,19,21,25
475	15	10011101010110	2,3,17,19,21,25
476	39	10110111010110	2,11,19,21,25
477	16	10110101010110	2,11,17,19,21,25
478	36	10010111010110	2,3,11,19,21,25
488	11	10111111000110	2,19,20,21,25
502	16	11111111010100	19,21,24,25
520	33	10011001010110	2,3,15,17,19,21,25
547	11	00011111111111	1,2,3

Table 3.2 Ability Levels and Atypicality of Cognitive States (sorted by θ values)

Cognitive States	Frequency	θ	ζ	Attributes not mastered
1	19	5.00	0.52	
5	18	3.06	1.43	3
9	37	1.98	0.67	19
2	180	1.83	-1.37	21
6	94	1.40	-0.59	3,21
10	46	1.14	-0.69	19,21
4	32	1.12	-0.55	17
66	87	0.88	0.16	21,24
14	30	0.83	-0.01	3,19,21
8	11	0.82	0.14	3,17,21
12	28	0.61	-0.22	17,19,21
70	11	0.60	0.94	21,24
253	43	0.59	0.06	2,21
257	15	0.34	0.55	2,3,21
261	30	0.15	-0.06	2,19,21
126	25	0.04	0.52	21,23
34	17	0.01	-0.74	20,21
268	11	-0.02	0.79	2,11
22	18	-0.17	-0.26	3,11,21
128	14	-0.19	0.88	3,17,21,23
269	31	-0.35	-0.51	2,11,21
138	40	-0.35	0.19	17,19,21,23
30	38	-0.54	-1.18	3,11,19,21
273	12	-0.56	-0.54	2,3,11,21
140	16	-0.57	0.26	3,17,19,21,23
28	12	-0.71	-1.58	11,17,19,21
277	13	-0.71	-1.37	2,11,19,21
468	67	-0.75	-0.45	19,21,25
469	18	-0.91	-0.10	17,19,21,25
472	132	-1.11	-0.33	2,19,21,25
473	25	-1.12	-0.32	2,17,19,21,25
488	11	-1.13	0.07	2,19,20,21,25
502	16	-1.13	0.70	9,21,24,25
474	32	-1.16	-0.87	2,3,19,21,25
476	39	-1.24	-0.64	2,11,19,21,25
257	15	-1.32	-0.65	2,3,17,19,21,25
477	16	-1.40	-0.40	2,11,17,19,21,25
478	36	-1.45	-1.00	2,3,11,19,21,25
215	12	-1.49	2.32	1,3
547	11	-1.81	1.82	1,2,3
217	30	-1.99	-0.11	1,3,21
220	13	-2.02	0.66	1,3,17,19
520	33	-2.07	-0.69	2,3,15,17,19,21,25
218	14	-2.22	-0.82	1,3,19,21
221	24	-2.25	-0.18	1,3,17,21
222	22	-2.55	-0.90	1,3,17,19,21

Table 3.3 Descriptive Statistics of the 14 Attributes and θ, ζ and Generalized ζs (N = 2334)

Attributes	mean	S.D.	Corr. with θ	Corr. with ζ
1	.896	.305	.25	-.11
2	.631	.483	.21	-.03
3	.542	.498	.30	-.17
6	.958	.201	.16	-.25
11	.764	.425	.21	-.01
15	.939	.240	.19	-.12
17	.668	.471	.25	-.18
18	.978	.152	.15	-.14
19	.461	.499	.22	.14
20	.879	.326	.19	-.01
21	.213	.409	.11	.58
23	.901	.298	.05	-.11
24	.807	.395	.17	-.34
25	.790	.408	.15	.27

Dimension	mean	S.D.
θ	.060	1.200
ζ	-.147	1.067
ζ_1	-.089	1.002
ζ_2	-.050	.992
ζ_3	-.055	1.028
ζ_4	-.076	1.010
ζ_5	-.027	1.008

Table 4.3.1 Examples of Probability Vectors for the First Ten Examinees

		14 Attributes													
ID	SAT	1	2	3	6	11	15	17	18	19	20	21	23	24	25
1	500	93	65	59	98	82	96	74	99	47	89	15	92	82	77
2	640	98	75	73	97	84	99	80	100	58	96	23	90	86	86
3	420	89	59	49	97	75	94	71	97	41	85	13	92	82	76
4	510	94	65	60	98	82	96	75	99	47	89	14	91	82	77
5	730	100	82	79	100	87	100	75	100	66	100	37	91	89	95
6	340	80	53	36	93	64	89	49	94	36	81	26	88	74	75
7	790	100	86	83	100	99	100	100	100	72	100	79	94	99	100
8	230	53	44	17	71	46	66	21	90	28	75	32	78	52	68

Table 5.1

Form 0A: Incidence Matrix Q for 14 Attributes and 25 Items

Item No.	1	2	3	6	11	15	17	18	19	20	21	23	24	25	% Correct
01	0	1	0	0	0	1	0	1	0	0	0	0	0	0	89.2
02	1	0	0	0	1	1	1	0	0	0	0	0	0	0	89.0
03	1	0	0	0	1	1	0	1	0	0	0	0	1	0	78.6
04	1	1	0	0	0	1	0	1	1	0	0	0	1	0	70.6
05	0	1	0	0	1	0	1	1	0	1	0	0	0	1	73.9
06	1	0	0	0	1	1	1	1	0	1	0	0	0	0	54.9
07	0	0	0	1	1	0	1	0	0	0	0	1	0	0	73.7
08	1	0	0	0	0	1	0	1	0	0	0	0	0	0	79.5
09	0	0	0	1	1	0	1	0	0	0	0	1	0	0	73.6
10	1	1	0	0	1	0	0	1	0	0	0	0	0	1	63.1
11	0	0	0	1	1	0	0	0	1	0	0	1	0	0	68.1
12	1	0	0	0	1	0	0	0	0	0	0	1	1	1	83.4
13	0	1	1	0	1	1	0	1	0	1	0	0	0	0	51.8
14	0	0	1	1	0	0	1	1	1	0	1	1	0	1	25.7
15	0	0	0	1	1	0	1	0	0	1	1	1	1	0	46.1
16	0	0	0	1	0	1	1	0	0	0	1	0	0	0	59.6
17	1	1	1	0	0	1	0	1	0	0	0	0	0	1	34.5
18	1	0	0	0	1	0	1	0	0	0	1	0	0	1	29.5
19	0	1	1	1	1	0	0	1	1	0	1	0	0	1	21.8
20	0	0	1	0	0	1	1	0	0	0	1	1	0	0	51.0
21	0	0	1	0	0	1	1	1	0	0	1	0	0	1	22.4
22	0	0	0	1	0	0	1	1	0	1	1	0	0	1	26.2
23	1	0	0	0	1	1	0	0	1	1	1	1	1	1	20.3
24	0	0	1	0	0	1	0	1	1	0	0	0	0	1	25.1
25	1	0	0	0	1	0	1	1	1	1	1	1	0	1	19.8

Table 5.2

Multiple Regression Results: Predicting Item Difficulties of 25 Items (Form 0A Section 2) From 14 Attributes.

Attribute	b	SEb	β	t
25	-14.60	9.00	-.31	-1.62
11	.85	8.36	.02	.10
17	3.10	7.91	.07	.39
06	-9.80	9.78	-.19	-1.00
24	6.43	7.86	.11	.82
20	-17.38	7.97	-.33	-2.18
19	-10.69	7.05	-.20	-1.52
03	-23.12	9.41	-.44	-2.46*
02	8.42	7.69	.16	1.10
23	6.57	9.52	.13	.69
21	-19.19	7.68	-.40	-2.50*
01	-10.27	8.05	-.21	-1.28
18	1.65	9.02	.03	.18
15	-1.29	12.67	-.03	-.10
a	81.47	19.97		
R^2	.91			
R^2adj.	.78			

* p<.05 ; ** p< .001

Table 5.3

Incidence Matrix Q for 15 Attributes and 25 Items (Form 0A section 2) and Percent Correct

Item No.	1	2	3	6	11	15	17	18	19	20	21	23	24	25	26	% Correct
01	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0	89.2
02	1	0	0	0	1	1	1	0	0	0	0	0	0	0	0	89.0
03	1	0	0	0	1	1	0	1	0	0	0	0	1	0	0	78.6
04	1	1	0	0	0	1	0	1	1	0	0	0	1	0	0	70.6
05	0	1	0	0	1	0	1	1	0	1	0	0	0	1	0	73.9
06	1	0	0	0	1	1	1	1	0	1	0	0	0	0	0	54.9
07	0	0	0	1	1	0	1	0	0	0	0	1	0	0	0	73.7
08	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	79.5
09	0	0	0	1	1	0	1	0	0	0	0	1	0	0	0	73.6
10	1	1	0	0	1	0	0	1	0	0	0	0	0	1	1	63.1
11	0	0	0	1	1	0	0	0	1	0	0	1	0	0	0	68.1
12	1	0	0	0	1	0	0	0	0	0	0	1	1	1	0	83.4
13	0	1	1	0	1	1	0	1	0	1	0	0	0	0	0	51.8
14	0	0	1	1	0	0	1	1	1	0	1	1	0	1	0	25.7
15	0	0	0	1	1	0	1	0	0	1	1	1	1	0	0	46.1
16	0	0	0	1	0	1	1	0	0	0	1	0	0	0	0	59.6
17	1	1	1	0	0	1	0	1	0	0	0	0	0	1	0	34.5
18	1	0	0	0	1	0	1	0	0	0	1	0	0	1	1	29.5
19	0	1	1	1	1	0	0	1	1	0	1	0	0	1	0	21.8
20	0	0	1	0	0	1	1	0	0	0	1	1	0	0	0	51.0
21	0	0	1	0	0	1	1	1	0	0	1	0	0	1	0	22.4
22	0	0	0	1	0	0	1	1	0	1	1	0	0	1	0	26.2
23	1	0	0	0	1	1	0	0	1	1	1	1	1	1	0	20.3
24	0	0	1	0	0	1	0	1	1	0	0	0	0	1	0	25.1
25	1	0	0	0	1	0	1	1	1	1	1	1	0	1	0	19.8

Table 5.4

The 25 Items (Form 0A Section 2) Listed by the 15 Attributes

Item	Attribute	Item	Attribute
01	2, 15, 18		
02	1, 11, 15, 17		
03	1, 11, 15, 18, 24		
04	1, 2, 15, 18, 19, 24		
05	2, 11, 17, 18, 20, 25		
06	1, 11, 15, 17, 18, 20, 07, 6, 11, 17, 23		
08	1, 15, 18		
09	6, 11, 17, 23		
10	1, 2, 11, 18, 25, 26*		
11	6, 11, 19, 23		
12	1, 11, 23, 24, 25		
13	2, 3, 11, 15, 18, 20		
14	3, 6, 17, 18, 19, 21, 23, 25		
15	6, 11, 17, 20, 21, 23, 24		
16	6, 15, 17, 21		
17	1, 2, 3, 15, 18, 25		
18	1, 11, 17, 21, 25, 26		
19	2, 3, 6, 11, 18, 19, 21, 25		
20	3, 15, 17, 21, 23		
21	3, 15, 17, 18, 21, 25		
22	6, 17, 18, 20, 21, 25		
23	1, 11, 15, 19, 20, 21, 23, 24, 25		
24	3, 15, 18, 19, 25		
25	1, 11, 17, 18, 19, 20, 21, 23, 25		

Table 5.5

The 15 Attributes Listed by the Items in which They Are Required

Attribute	Items (1-25 form 0A section 2)
1	2, 3, 4, 6, 8, 10, 12, 17, 18, 23, 25
2	1, 4, 5, 10, 13, 17, 19
3	17, 19, 20, 21, 24
6	7, 9, 11, 14, 15, 16, 19, 22
11	2, 3, 5, 6, 7, 9, 10, 11, 12, 13, 15, 18, 19, 23, 25
15	1, 2, 3, 4, 6, 8, 13, 16, 17, 20, 21, 23, 24
17	2, 3, 5, 6, 7, 14, 15, 16, 18, 20, 21, 22, 25
18	1, 3, 4, 5, 8, 10, 13, 17, 19, 22, 24, 25
19	12, 14, 19, 25
20	10, 13, 15, 22, 23, 25
21	12, 16, 18, 19, 20, 21, 23
23	3, 7, 9, 11, 12, 14, 15, 23, 25
24	3 4, 12, 15, 23
25	5, 10, 12, 14, 17, 18, 19, 21, 22, 24, 25
26	1, 18

Table 5.6

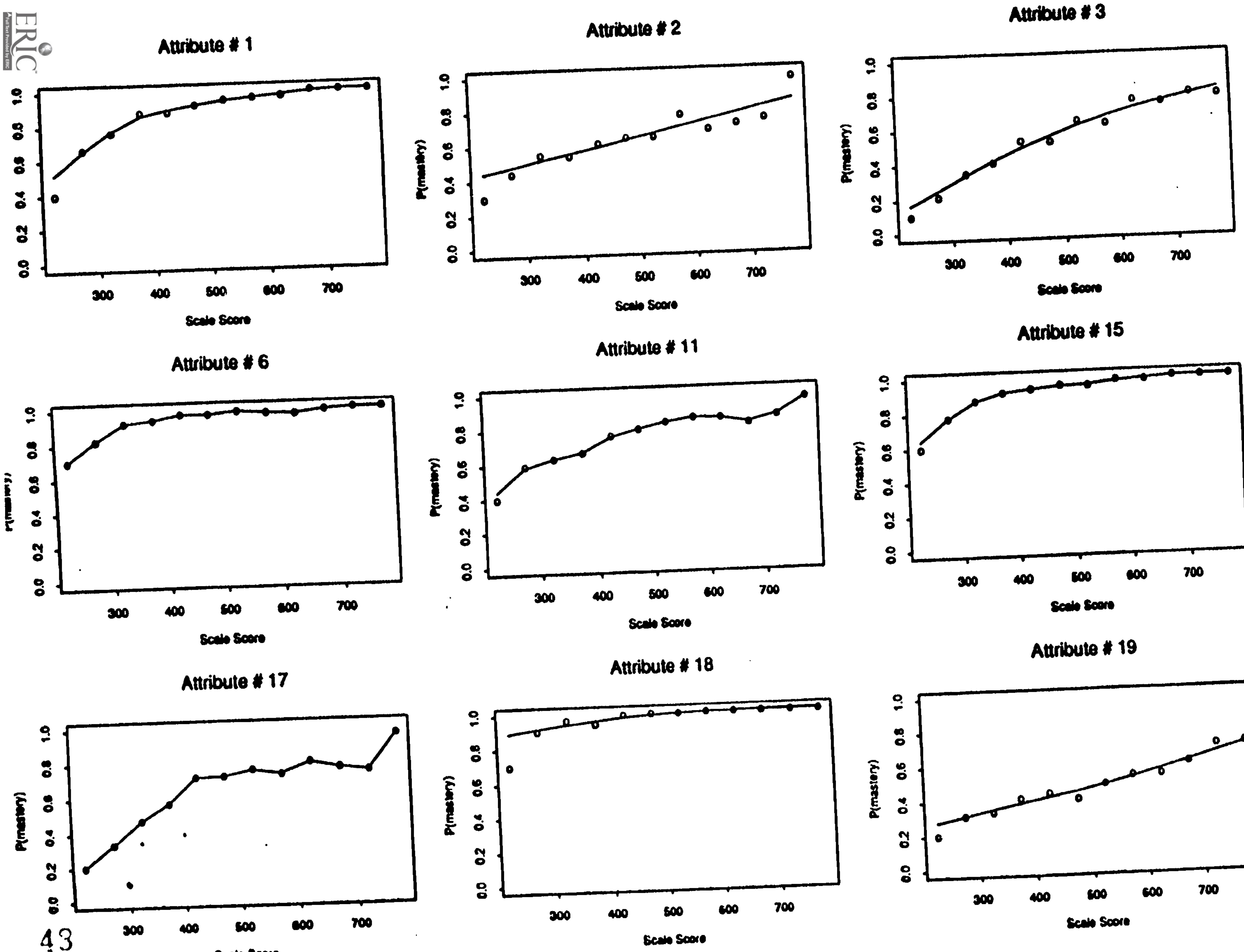
Multiple Regression Results: Predicting Item Difficulties (percent correct) for Items 1-25

(Form 0A Section 2) From 15 Attributes.

Attribute	b	SEb	β	t
26	-26.35	12.57	-.30	-2.10
17	-.49	7.05	-.01	-.07
03	-23.53	8.13	-.44	-2.89*
23	-.53	8.89	-.01	-.06
20	-20.72	7.06	-.39	-2.93*
06	-17.82	9.28	-.35	-1.92
19	-12.38	6.15	-.23	-2.01
24	2.04	7.10	.03	.29
11	1.22	7.23	.03	.17
25	-16.94	7.86	-.36	-2.16
02	9.23	6.65	.17	1.39
18	-2.97	8.10	-.06	-.37
01	-6.61	7.17	-.14	-.92
21	-12.63	7.33	-.26	-1.72
15	-13.27	12.35	-.28	-1.07
a	98.41	19.06		
R^2	.94			
$R^2_{adj.}$	.83			

* $p<.05$

List of Figures

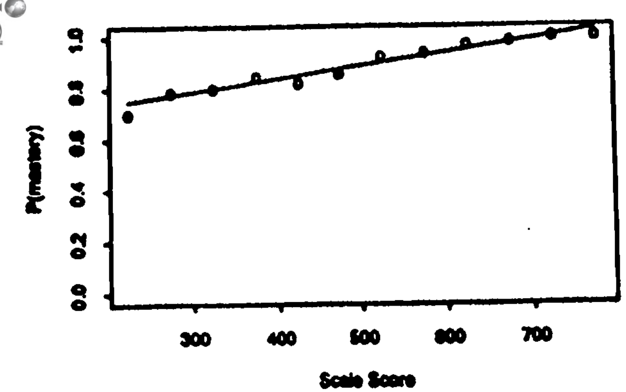


SAT-Q, Section 1

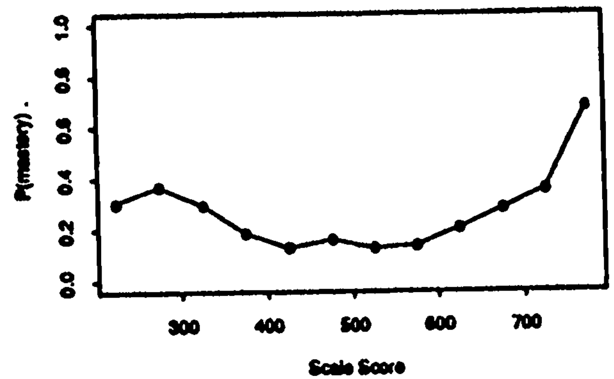
43

49

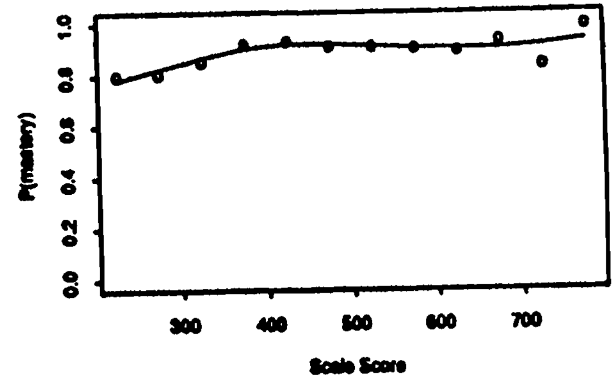
Attribute # 20



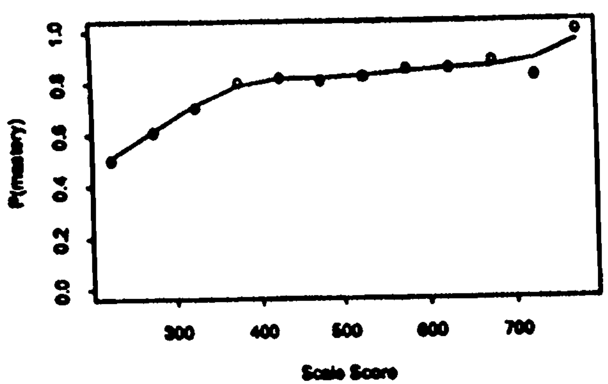
Attribute # 21



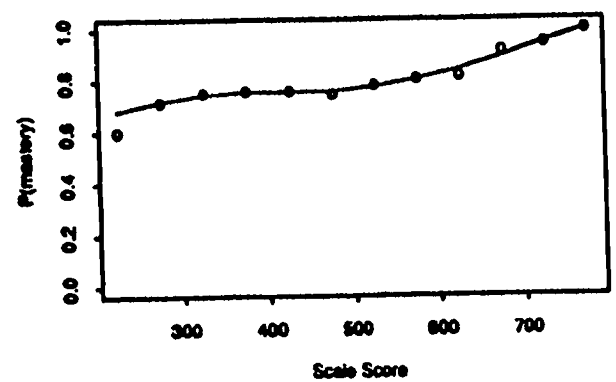
Attribute # 23

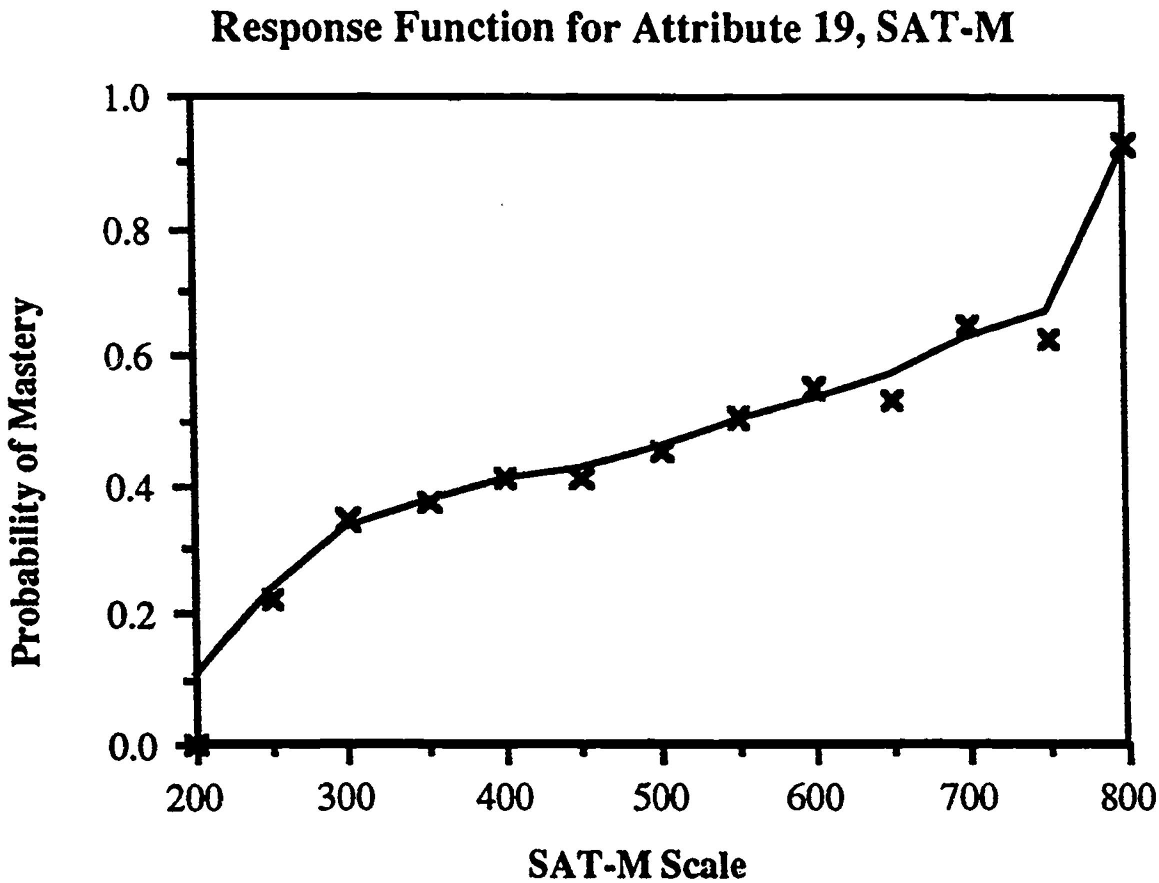


Attribute # 24



Attribute # 25





Attribute 19 represents the ability to comprehend and apply rules and theorems correctly.

The figure shows an empirical attribute response function (points denoted by x) and posterior median estimates of selected values of the corresponding theoretical function (connected by straight lines).

The posterior mean estimate of the 25% point for this function is 277.
The posterior mean estimate of the 50% point for this function is 546.
The posterior mean estimate of the 75% point for this function is 760.

1) A student report, Kumi Tatsuoka

SAT percentile score based on item-level: 60

Your percentile scores on <u>the content area</u> are:

Arithmetic.........A
Algebra............C
Geometry...........C
Miscellaneous......D

Performance underlying cognitive processes:

Understanding the meaning of concepts..... C

Application of simple rules/algorithms (solving equations, computation, derivation of simple algebraic expressions).......... A

Comprehension and application of rules/theorems, principles correctly...... C

Reading comprehension (+follow instructions;math/geometry terminology)... B

Reasoning (create an equation, identifying components and follow procedures)......... C

Analytic thinking, cognitive restructureing (higher mental processes)................. D

Strategies (trial-and-errors by plug in numbers, make an inference of the correct answer from options with unknown systematic methods................................. C

The complex problems with steps > 1....... C

A: top 10 percent
B: 70 - 89 percentile
C: Average
D: 30 - 49 percentile
E: 10 - 29 percentile
F: bottom 10 percent

2) A student report for Jane Smith

SAT-scaled score based on item performance: 600

Probability of success on attribute(s) associated with:

	Mean at 600-level	Your score
Arithmetic	97 %	ok
Elementary Algebra	72 %	no
Advanced algebra	69 %	no
Geometry	97 %	ok
Understanding the meaning of concepts	76 %	ok
Application of simple rules/algorithms (solving equations, computation, derivation of simple algebraic expressions)	100 %	ok
Comprehension and application of rules/theorems, principles correctly	54 %	no
Reading comprehension (+follow instructions;math/geometry terminology)	90 %	no
Reasoning (create an equation, identifying components and follow procedures)	94 %	no
Analytic thinking, cognitive restructureing (higher mental processes)	17 %	ok
Strategies (trial-and-errors by plug in numbers, make an inference of the correct answer from options with unknown systematic methods	85 %	ok
Mastery of complex problems with steps > 1	82 %	ok

Additional Comments:

Your performance pattern is rather unusual, so we provide you with your diagnosed cognitive state on the right most side of the above table.

we recommend that you practice word problems and pay more attention to the meaning of principles, theorems and properties.

You should also follow the instructions more carefully.

II. A report for a class room teacher

Class size 10, five girls and five boys
junior year, Teacher is Mrs Smith

The mean of SAT-scale score: 450
The standard deviation : 30

names	SAT scale	percentile rank	A1	A2	A3	A4	A5	A6	A7	A8	A9
			attributes								
1. Donald Duck	750	95%	90	85	50	85	60	77	81	92	65
2. Wylie Cayote	540	61%	81	64	55	42	89	45	32	75	18
3. Mickey Mouse	605	80%	82	71	62	40	80	55	54	67	32
4. Olive Oyl	680	90%	88	67	32	97	65	46	98	63	88
5. Bo Peep	442	67%	43	53	65	24	35	36	56	46	67
......											
......											
......											
10. Charlie Brown	590	69%	75	60	50	40	85	46	42	77	29
Average	620	74%	76	72	65	54	67	51	46	43	25
S.D.	42		5	7	8	10	11	9	15	12	17

55

APPENDIX

Appendix I

The rule-space-model has recently been introduced in various ETS technical reports (Tatsuoka & Tatsuoka, 1992; Sheehan, Tatsuoka & Lewis, 1993; Birenbaum & Tatsuoka, 1993; Birenbaum, Kelly & Tatsuoka, 1993). This paper emphasizes the introduction of the procedures that lead to probability functions for attributes (PFAs), which are applied to SAT Mathematics tests. An PFA is the conditional probability function for successful performance on each attribute at given IRT ability level θ,

$$P_{Ak}(\theta) = \text{Prob}(\, A_k = 1 \mid \theta), \; k = 1, 2, \ldots\ldots, K \qquad (1)$$

Since PFAs are defined on the IRT ability variable θ or equivalently, on the SAT scale that are obtained by transforming the θ-scale, each scale point is associated with a probability vector of the cognitive attributes.

1. An Incidence Matrix and All Possible Ideal-Item-Score Patterns

Tatsuoka (1990) organized the underlying cognitive tasks that are required in answering test items in an incidence matrix, Q-matrix, whose rows represent attributes (i.e., knowledge, cognitive processes and skills etc.) and columns represent items. The entries in each column indicate which attributes are involved in the solution of each item. The incidence matrix of order that relates the 25 items in Section 2 of the SAT M with the 14 attributes selected in the previous section is used for deriving all possible ideal-item-score patterns which correspond to attribute mastery patterns (Tatsuoka, 1991). The expression "ideal-item-score patterns" will be used hereafter to refer to logically determined knowledge states, as contrasted with the examinees' actual item-response patterns. The logically determined ideal-item-score patterns also represent classification groups, which correspond to the attribute mastery patterns. The ideal-item-score patterns are the images of a Boolean Descriptive Function (BDF) that is defined on the lattice of attributes. The BDF takes the value of either one or zero, for right or wrong on the items. The definition of the BDF can be stated by hypothesizing that "if Attribute A_k cannot be done correctly" or equivalently "if A_k is not mastered" then the items involving A_k cannot be answered correctly. The value of one for A_k means that "one can do A_k correctly" which is equivalent to "mastery of A_k" (Tatsuoka, 1991).

An algorithm that was developed by Varadi & Tatsuoka, 1989 produces all possible ideal-item-score patterns from an incidence matrix. An intuitive illustration is given by Tatsuoka (1993). A computer program BUGLIB (Varadi & Tatsuoka, 1989) produced more than 3000 ideal-item-score patterns for the incidence matrix of order 27 x 60 in Table 1.2.3, and 600 for that of order 14 x 25 associated with Table 1.3.3. Since the current form of BUGLIB cannot further analyze data from more than 2000 groups, the discussion in this report is restricted to the analysis results from Table 1.3.3, which relates to Section 2 of SAT M, Form 8A. Table A.1 shows a partial list of the 600 ideal-item-score patterns.

Insert Table A.1 about here

The first 25 columns after the IDs give the ideal-item-score patterns, followed by the two columns showing the values of θ estimated by the Maximum Likelihood Method and ζ (Tatsuoka, 1984, 1985; Tatsuoka & Linn, 1983), and the last 14 columns show the corresponding attribute patterns. The m-th ideal-item-score pattern is the image of the m-th attribute pattern by the BDF. The variable ζ will be described in Section 2.3.

<u>2. A set of "fuzzy" response patterns</u> There are 2^{14} possible attribute patterns for 14 attributes, but the BDF reduces the number of reliable attribute patterns to 600. These 600 attribute patterns correspond to 600 ideal-item-score patterns. Conceptually, an item-response pattern that does not correspond to one of these 600 ideal-item-score patterns is considered to be a "fuzzy item patterns" produced by slips. Slips are regarded as deviations from an ideal-item-score pattern.

Bayes' decision rules for minimum error are known to produce optimal classification and are also known to be relatively unaffected by the distribution of scores in a group. Application of Bayes' decision rules to our classification problem requires that the distribution of each cognitive state should be obtained statistically.

Tatsuoka & Tatsuoka (1987) introduced a slippage random variable and slippage probabilities for the items, and explained fuzzy response patterns as outcomes of inconsistent performance. The fuzzy response patterns around each ideal-item score pattern will cluster together. They showed that a set of fuzzy response patterns around an ideal-item-score pattern follows a compound binomial distribution with slippage probabilities for each item. Falmagne (1989) formulated a model that estimates these slippage probabilities.

However, if the number of cognitive states is as large as 600, we would need an enormously large sample for estimating the parameters of the model such as latent class models. An efficient algorithm for estimating very large numbers of state parameters has not been developed yet. The rule space model does not require the estimation of state parameters because it is an analytical approach, and the probabilities of state membership for an individual will be obtained through a classification procedure.

3. Classification space and bug distributions The rule-space model takes a statistical pattern classification approach to achieve classification of examinees into one of 600 cognitive states. An advantage of this approach is that the problem of combinatorial explosion is treated geometrically by mapping all patterns — both the examinees' response patterns and ideal-item-score patterns — into a vector space in which an appropriate distance is defined. Moreover, the dimension of the classification space usually equals the number of groups, in our context the number of cognitive states, but the model reduces the dimension of, say 600, to as few as two dimensions. If two states are similar in terms of mastery of the attributes, they are located close to each other in the rule space.

The vector space is a Cartesian Product space of θ and the image of a mapping function $f(x, \theta)$ defined by Equation 2.3.1,

$$f(X, \theta) = (P_j(\theta) - X,\ P_j(\theta) - T(\theta))$$

$$= b_1X_1 + b_2X_2 + \ldots + b_nX_n + \text{constant}. \quad (2)$$

Since this function is continuous, the fuzzy response patterns around a given ideal-item-score pattern, R, will be mapped onto points in the vicinity of the image of R, $f(\theta_R, R)$, and θ_R. These image points are denoted by $\{(\theta_R, f(\theta_R, R))$. In practice, $f(\theta_X, X)$ for any X will be standardized and denoted by ζ_X. The second coordinate, $f(\theta_R, R)$ will be replaced by ζ_R. We assume that these points (the images of fuzzy response patterns) swarm around R, and that $\{(\theta_R,$

ζ_R)) follow a bivariate normal distribution (Tatsuoka & Tatsuoka, 1987; Tatsuoka, 1990), called a "bug distribution".

The cognitive state R whose θ_R is in somewhere between -3 and +3, but for which the absolute value of ζ_R is larger than 3 may not really exist (Tatsuoka, 1984). If the values of ζ for some states are close to zero, many examinees will be classified into such states.

The mapping by f may not be one-to-one, but DiBello and Baillie (1992) proved that f is indeed almost one-to-one everywhere. The cases for the mapping not being one-to-one will never happen when the IRT parameters a_j and b_j are estimated from a real dataset. The standardized $f(x,\theta)$, ζ, will be the y-axis of the classification space, called Rule Space (Tatsuoka, 1985). However, the name "Rule Space" may be misleading because the mapped cognitive states can be misconception states, knowledge states or even be personality states. Tatsuoka (1985) showed that the expectation of $f(x,\theta)$ is zero and the variance is given by 3,

$$\mathrm{Var}[f(x, \theta)] = \sum_j P_j(\theta)Q_j(\theta)(P_j(\theta) - T(\theta))^2 \tag{3}$$

The configuration in rule space is something like what is shown in Figure 1.

Insert Figure 1 about here

In this figure, the ellipses represent equal density contours for the bug distributions. The covariance matrix of a bug distribution will be a diagonal matrix with the variances of θ and ζ as the diagonal elements since these variables are uncorrelated (Tatsuoka, 1985).

4. Classification Procedure Suppose an examinee's response patterns are mapped into the rule space. Then, the distance D^2 between the individual examinee's point, (θ_x, ζ_x) and the centroid (θ_R, ζ_R) of the bug distribution R is given by (4), since the covariance matrix Σ of the distribution is as shown in Equation (5).

$$D^2 = (\theta_x - \theta_R)^2/(1/I(\theta_R)) + (\zeta_x - \zeta_R)^2. \tag{4}$$

$$\begin{vmatrix} 1/I(\theta_R) & 0 \\ 0 & 1 \end{vmatrix} \tag{5}$$

The Mahalanobis distance (4) follows a Chi-Square distribution with two degrees of freedom (Lachenbruch, 1975). Suppose an examinee's point X is classified into one of the 600 predetermined groups (or, equivalently, knowledge states) determined from Table 1.3.3. Then, 600 Mahalanobis distances are first computed. If the criterion value of $\chi_2{}^2$ is set to 4.605 ($p=.25$), then the cognitive states whose Mahalanobis distance D^2 from X is less than 4.605 will be considered as eligible cognitive states for classification of X. If there is no cognitive state whose Mahalanobis distance from X is less than 4.605, then X will be left unclassified.

Suppose States R_1 and R_2 are the two closest ones to X, that is, which have the two smallest Mahalanobis distances from X; then Bayes' decision rule for minimum error will be applied to them to determine the final group for X, and the total classification error will be computed (Tatsuoka & Tatsuoka, 1987). If the covariance matrices of two states are almost identical, as they are in cases with which we deal, and their distributions are normal, then the Bayes' decision rule becomes equivalent to considering a linear discriminant function. That is, the negative of the logarithm of the ratio of the posterior probabilities of R_1 and R_2 for X will be a linear function under the normality and equal covariances conditions.

Kim (1990) examined the effect of violation of the normality requirement with simulated data in the rule space, and found that the linear discriminant function is robust against this violation. Kim further compared the classification results by the linear discriminant functions and K nearest neighbors method, which is a non-parametric classification approach and does not assume the normality of a bug distribution, and found that the linear discriminant functions performed better.

Suppose R_1 is the cognitive state to which X belongs, then the response pattern X and the ideal-item-score pattern for R_1 should be close to each other. Since R_1 corresponds to an attribute mastery pattern A_{R1}, the response pattern X also corresponds to A_{R1} with high probability. In other words, the response pattern X is converted to the attribute mastery pattern corresponding to R_1.

Since the bug distribution for R_1 is assumed to be bivariate normal, the posterior probability of R_1 given X^* can be computed by using the prior probability of R_1, as discussed in Tatsuoka & Tatsuoka (1987).

5. Multidimensional Rule Space and Generalized Zetas After mapping 600 ideal-item-score patterns into the Cartesian Product space of θ and ζ, the images of these 600 ideal-item-score patterns may become too close and too crowded, that is they may be too densely packed on the plane for classification purposes. If the mapped cognitive states are not well separated, then the error rates for classification become unacceptably large. In order to separate the images of ideal-item-score patterns, additional

dimensions may be needed. For the analysis of SAT M, Section 2, five dimensions are added.

Generalized ζs were first defined by Varadi & Tatsuoka (1989). Suppose Γ is a subset of items, then the generalized ζ_Γ is defined as the sum of the scalar product of two residuals, $(P_j(\theta)-X_j)'\ (P_j(\theta)-T(\theta))$, over all j in Γ, divided by the standard deviation of the sum. Selection of Γ is still an art and its further development is left as a research topic for the future. However, it is recommended to take union and intersection sets of the items which correspond to the attribute row vectors, $A_1,\ldots,A_k$ of the incidence matrix. Generalized zeta defined on the items involving A_k, ζ_{Ak} is given below with its numerator function f:

$$\begin{aligned} f(z,\theta_z) &= (P_j(\theta_z)-Z_j,\ P_j(\theta_z)-T(\theta_z)) \quad (6) \\ &= (Q_k'[P_j(\theta_x)-X_j],\ Q_k'[P_j(\theta_x)-T(\theta_x)]) \\ &= Q_k'([P_j(\theta_x)-X_j],\ [P_j(\theta_x)-T(\theta_x)]) \\ \zeta_z &= f(z,\theta_z)/SQRT\{Var[f(z,\theta_z)]\} \quad (7) \end{aligned}$$

where $z = Q_k x$, and θ_z is the Maximum Likelihood Estimate obtained from the items involving A_k.

The expectation and variance of $f(z,\theta_z)$ are given by (8) and (9).

$$E[f(z,\theta_z)] = 0 \quad (8)$$

$$Var\{f(z,\ \theta_z)\} = \sum_{j \in \{Q_k \neq 0\}} P_j(\theta_x)\,(1 - P_j(\theta_x))\ [P_j(\theta_x) - T(\theta_x)]^2 \quad (9)$$

The generalized ζs are uncorrelated with θ, which can be shown in exactly in the same manner as the proof for the uncorrelatedness of θ and ζ given in Tatsuoka (1985). Furthermore, a generalized ζ computed by using the items involving any combination of A_k -- defined as the union or intersection sets of A_k ,k=1,..L -- also has the orthogonality property with θ.

Any generalized ζ can be added to the original two-dimensional Cartesian product space as a new dimension, and a multidimensional classification space can be formulated. Both the ideal-item-score patterns and examinees' response patterns are mapped into the (m+2)- dimensional Cartesian product space $\{(\theta, \zeta, \zeta_1, \zeta_2, \ldots, \zeta_m)\}$. The larger the value of ζ_{Ak} is, the more unusual the performance on the items involving Attribute A_k is. Thus, each coordinate in the multidimensional rule space can maintain interpretability.

The set of atoms in the lattice of K attributes forms a basis (Tatsuoka, 1991, Birkhoff, 1970), but it is very difficult to give intuitive interpretations to the atoms unless the incidence matrix is diagonal — each attribute being involved in only one item and each item involving only one attribute. So, the atoms are not used in the rule-space model although they are mathematically useful entities. However, if intuitive interpretations of the coordinates are not required, then the atoms can be used for formulating a multidimensional space, after transforming item score patterns.

For SAT M, Section 2, five generalized ζs were added to the original two-dimensional space, and classification of examinees was done in the resulting seven dimensional space. The new dimensions are shown in Table A.2.

Insert Table A.2 about here

The interpretation of each new axis is similar to that of ζ which uses all the items. For example, ζ_1 is computed using the items involving the attributes 1,3 and 4 in which 14 items are considered. If the value of ζ_1 is large, then the pattern of the 14 relevant items is aberrant, while a smaller value (including a negative value) of ζ_1 indicates that the pattern conforms well to the order of difficulty for the 14 items.

The bug distributions for cognitive states — the images of the ideal-item-score patterns and their fuzzy response patterns into the m+2 dimensional classification space — are assumed to be multivariate normal distributions. Their centroids are the images of the ideal-item-score patterns. A squared Mahalanobis distance between X and the image of R that is the centroid of bug distribution R, or a cognitive state R follows a χ^2 distribution with m+2 degrees of freedom (Lachenbruch, 1975). The classification procedure and computation of error probabilities, prior and posterior probabilities are a straightforward extension of the two dimensional case.

After classifying examinees' response patterns into one of the predetermined groups or cognitive states, their item response patterns correspond to the attribute mastery patterns along with the information about D^2, error probabilities, probability of belonging to the cognitive state to which the examinees are classified, ML estimates of θ, ζ and generalized ζs (Varadi & Tatsuoka, 1989). We propose to use the attribute patterns to estimate Attribute Characteristic Curves, which is comparable to the estimation of Item Response Curves. However, we don't use parametric functions for PFAs. Non-parametric estimation of PFAs will be illustrated with the attribute mastery patterns of SAT M Section 4. In the next section, analysis results will be described.

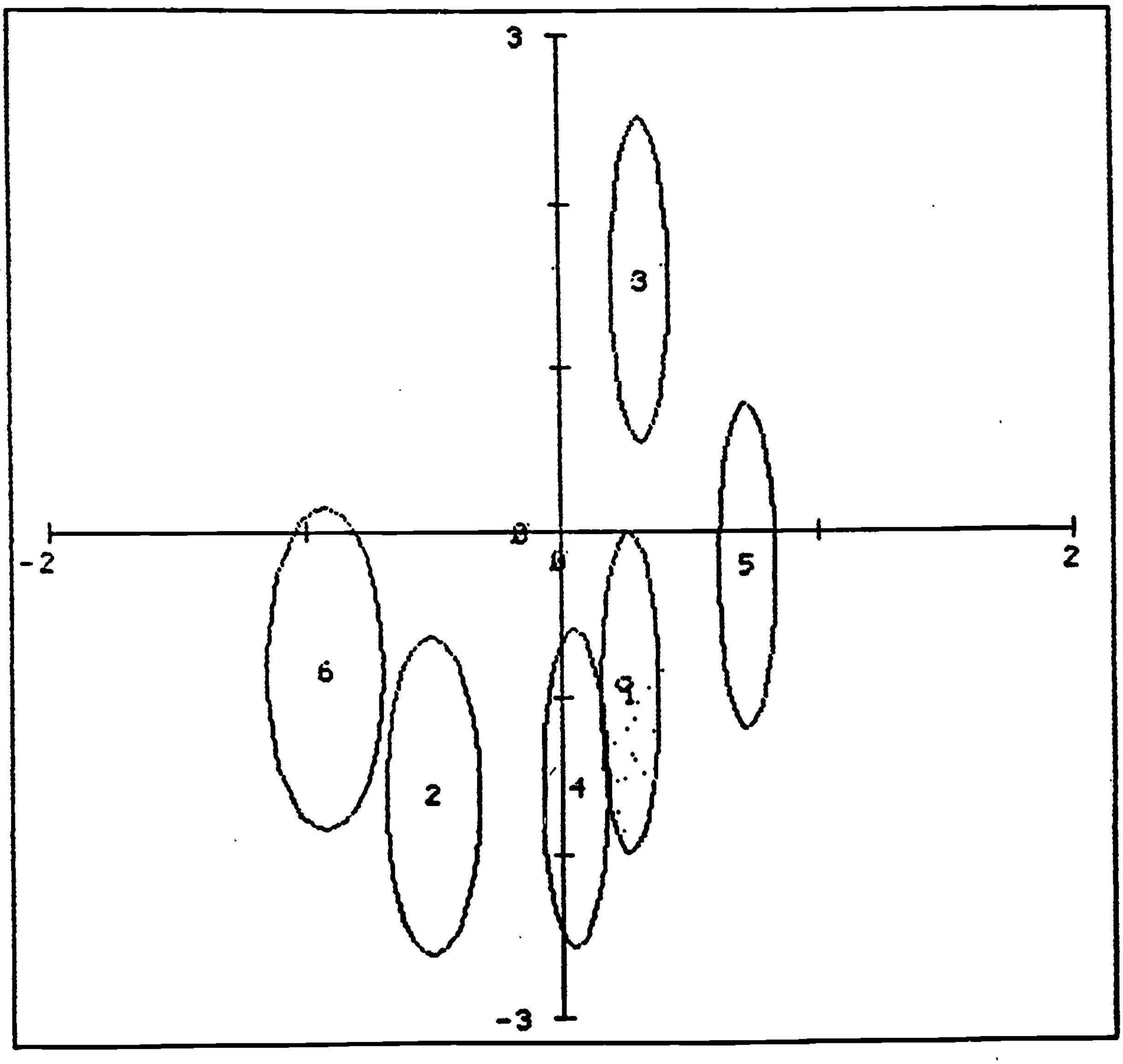
Table A.1 The first 10 out of 600 ideal-item-score patterns derived from the incidence matrix given in Table 1.

Cognitive States	Ideal-Item-Score Patterns, 25 items	θ	ζ	Attribute Patterns
1	1111111111111111111111111	5.00	.52	11111111111111
2	1111111111111111011111000	1.83	-1.37	11111111110111
3	1111111111111011111101111	2.30	1.88	11111101111111
4	1111111111111011011101000	1.12	-.55	11111101110111
5	1111111111110111111111111	3.06	1.43	11011111111111
6	1111111111110111011111000	1.41	-.59	11011111110111
7	1111111111110011111101111	1.74	2.59	11011101111111
8	1111111111110011011101000	.82	.14	11011101110111
9	1111111101111111101111101	1.98	.67	11111111011111
10	1111111101111111001111000	1.14	-.69	11111111010111
.	....	.	.	....
.	....	.	.	....
.	....	.	.	....
.	....	.	.	....
301	0111110010010011110000000	-.61	-.94	10111011000111
302	0111110010011010100000000	-.60	-.05	10111001001111
.	....	.	.	....
.	....	.	.	....
.	....	.	.	....
.	....	.	.	....
591	0000000000000001100000000	-2.76	1.52	10011000011101
592	0000000000000001010000000	-2.84	1.42	10011100110001
593	0000000000000001000100000	-2.87	1.38	10001000010111
594	0000000000000001000000000	-3.52	.86	10001000010001
595	0000000000000000110000010	-2.44	2.44	00010101011101
596	0000000000000000110000000	-2.74	1.76	00010100111101
597	0000000000000000100000000	-3.32	1.13	00010000001101
598	0000000000000000010000010	-2.91	1.65	00010101111001
599	0000000000000000010000000	-3.48	1.02	00010100110001
600	0000000000000000000000000	-5.00	.53	00000000000000

Table A.2 The Generalized ζs Added as New Dimensions and Their Attribute Sub Space

	Attributes		Corresponding items
1	ζ_1	$A_1+A_3+A_4$	2,3,4,5,8,9,10,11,13,15,16,19,20,25
2	ζ_2	A_5	2,6,13,21
3	ζ_3	A_8+A_{10}	8,11,12,16,18,20,22,23,25
4	ζ_4	$A_{11}+A_{12}$	1,2,5,6,8,9,11,12,13,15,16,19,20,21,22,23
5	ζ_5	A_{14}	14,17

Here is the progression of the student's points throughout the test. "o" = final point.



X = ability level, Y = unusualness of response pattern
Press HELP for more information

Figure 1 An Example of the Rule Space configuration

Possible Score Reports Based on the Rule-Space Results

Potential Audiences/Usages and Types of Reports

Audience	*Usage*	*Type of Report* *
Higher education institutions	selection, placement of applicants	per examinee: 1,3
High schools		
a. Test takers	vocational decisions; skills to be improved	per examinee: 1,2, 3/4, 5
b. Teachers	remediation/future instruction planning	per class: 8, 9
c. Principals	teacher/curriculum evaluation	entires school+ class comparisons: 6, 7
d. District Administration	school/curriculum evaluation, educational policy	entire district + school comparisons: 6, 7
e. State Administration	district/curriculum evaluation, educational policy	entire state + district comparisons: 6, 7
Item developers	test evaluation: items to be improved/ added/deleted	per item: 10, 11, 12 13, 14

* For key see attached list of report's components

Report's Components

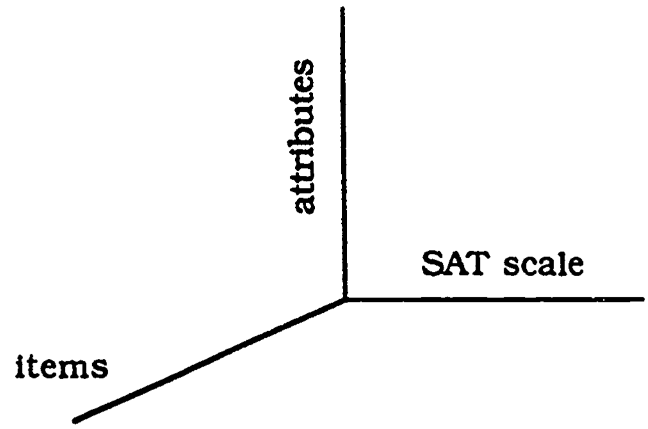
a. Individual performance
1. SAT scale score
2. SAT percentile score (relative standing school-wise/nation-wise)
3. Attribute probability profile per student
4. Attribute probability profile on a 6 point scale, where: A=90-100; B=70-89; C=50-69; D=30-49; E=10-29; F=0-9.
5. Detailed diagnosis (narrative) [description of attribute- mastery profile; appropriateness scores; recommendations . . .)

b. Group performance
6. Attribute profile -means
7. SAT scores - means
8. S-A (Student-Attribute) chart
9. S-I (Student -Item) chart

c. Item performance
10. IRT item difficulty index
11. IRT item discrimination index
12. Attribute pattern per item (Q matrix)
13. Reliability indices
14. Results of regressing item difficulties on attribute veactors.

The Database for the Retrieval System

a. Psychometric data



The basic information available for enhancing scoring reports is stored in the database. The database consists of four parts:

1. The score matrix which contains student ID, an item response pattern, a θ-value, a ζ-value (index for unusualness of a pattern), an attribute pattern for examinees.
2. The incidence matrix.
3. The probability matrix of indicating each item's success rate at various θ-levels (will be converted to SAT scale later),
4. The probability matrix of indicating each attribute's mastery rate at various θ levels.

b. Contextual data

1. Demographic data: student's gender, ethnicity, SES . . .
2. Student's classroom/school/district/state affiliation
3. Test format . . .

The information stored in the database will be available for creating a variety of combinations, according to a request by a user.

A mapping sentence that containing content, process and context facets areas will be available to help choosing any combination of variables. Some users may chose content variables for making a summary statistics of item and attribute performance on a test while others may select context (forms and settings of tests) variables to see their effect on performance differences.

The retrieval system extracts any combination of information on the variables from the database and prepares a summary for the required report.

Distribution List

Dr Terry Ackerman
Educational Psychology
260C Education Bldg
University of Illinois
Champaign IL 61801

Dr Terry Allard
Code 3422
Office of Naval Research
800 N Quincy St
Arlington VA 22217-5660

Dr Nancy Allen
Educational Testing Service
Mail Stop 02-T
Princeton NJ 08541

Dr Gregory Anrig
Educational Testing Service
Mail Stop 14-C
Princeton NJ 08541

Dr Phipps Arabie
Graduate School of Management
Rutgers University
92 New Street
Newark NJ 07102-1895

Dr ˙ aac I Bejar
Educational Testing Service
Mail Stop 11-R
Princeton NJ 08541

Dr William O Berry
Director of Life and
Environmental Sciences
AFOSR/NL N1
Bldg 410
Bolling AFB DC 20332-6448

Dr Thomas G Bever
Department of Psychology
University of Rochester
River Station
Rochester NY 14627

Dr Menucha Birenbaum
Educational Testing Service
Princeton NJ 08541

Dr Bruce Bloxom
Defense Manpower Data Center
99 Pacific St
Suite 155A
Mouterey CA 93943-3231

Dr Gwyneth Boodoo
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Dr Richard L Branch
HQ USMEPCOM/MEPCT
2500 Green Bay Road
North Chicago IL 60064

Dr Robert Brennan
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Dr David V Budescu
Department of Psychology
University of Haifa
Mount Carmel Haifa 31999
ISRAEL

Dr Gregory Candell
CTB/MacMillan/McGraw-Hill
2500 Garden Road
Monterey CA 93940

Dr Paul R Chatelier
Perceptronics
1911 North Ft Myer Drive
Suite 1100
Arlington VA 22209

Dr Susan Chipman
Coguitive Science Program
Office of Naval Research
800 North Quincy Street
Code 3422
Arlington VA 22217-5660

Dr Raymond E Christal
UES LAMP Science Advisor
AL/HRMIL
Brooks AFB TX 78235

Dr Norman Cliff
Department of Psychology
University of Southern California
Los Angeles CA 90089-1061

Director
Life Sciences
Code 3420
Office of Naval Research
Arlington VA 22217-5660

Commanding Officer
Naval Research Laboratory
Code 4827
Washington DC 20375-5000

Dr John M Cornwell
Department of Psychology
I/O Psychology Program
Tulane University
New Orleans LA 70118

Dr William Crano
Department of Psychology
Texas A&M University
College Station TX 77843

Dr Linda Curran
Defense Manpower Data Center
Suite 400
1600 Wilson Blvd
Rosslyn VA 22209

Professor Clément Dassa
Faculté des sciences de l'éducation
Département d'études en éducation
et d'administration de l'éducation
CP 6128 succursale A
Montéal Québec
CANADA H3C 3J7

Dr Timothy Davey
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Dr Charles E Davis
Educational Testing Service
Mail Stop 16-T
Princeton NJ 08541

Dr Ralph J DeAyala
Measurement Statistics and
Evaluation
Benjamin Bldg Room 1230F
University of Maryland
College Park MD 20742

Dr Sharon Derry
Florida State University
Department of Psychology
Tallahassee FL 32306

Hei-Ki Dong
Bellcore
6 Corporate Place
RM: PYA-1K207
PO Box 1320
Piscataway NJ 08855-1320

Dr Neil Dorans
Mail Stop 07-E
Educational Testing Service
Princeton NJ 08541

Dr Fritz Drasgow
University of Illinois
Department of Psychology
603 E Daniel Street
Champaign IL 61820

Defense Tech Information Center
Cameron Station Bldg 5
Alexandria VA 22314
(2 Copies)

Dr Richard Duran
Graduate School of Education
University of California
Santa Barbara CA 93106

Dr Susan Embretson
University of Kansas
Psychology Department
426 Fraser
Lawrence KS 66045

Dr George Engelhard Jr
Division of Educational Studies
Emory University
210 Fishburne Bldg
Atlanta GA 30322

ERIC Facility-Acquisitions
2440 Research Blvd
Suite 550
Rockville MD 20850-3238

Dr Marshall J Farr
Farr-Sight Co
2520 North Vernon Street
Arlington VA 22207

Dr Leonard Feldt
Lindquist Center for Measurement
University of Iowa
Iowa City IA 52242

Dr Richard L Ferguson
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Dr Gerhard Fischer
Liebiggasse 5
A 1010 Vienna
AUSTRIA

Dr Myron Fischl
US Army Headquarters
DAPE-HR
The Pentagon
Washington DC 20310-0300

Mr Paul Foley
Navy Personnel R&D Center
San Diego CA 92152-6800

Chair
Department of Computer Science
George Mason University
Fairfax VA 22030

Dr Robert D Gibbons
University of Illinois at Chicago
NPI 909A M/C 913
912 South Wood Street
Chicago IL 60612

Dr Janice Gifford
University of Massachusetts
School of Education
Amherst MA 01003

Dr Robert Glaser
Learning Res & Development Cntr
University of Pittsburgh
3939 O'Hara Street
Pittsburgh PA 15260

Dr Susan R Goldman
Peabody College
Box 45
Vanderbilt University
Nashville TN 37203

Dr Timothy Goldsmith
Department of Psychology
University of New Mexico
Albuquerque NM 87131

Dr Sherrie Gott
AFHRL/MOMJ
Brooks AFB TX 78235-5601

Dr Bert Green
Johns Hopkins University
Department of Psychology
Charles & 34th Street
Baltimore MD 21218

Professor Edward Haertel
School of Education
Stanford University
Stanford CA 94305-3096

Dr Ronald K Hambleton
University of Massachusetts
Lab of Psychom & Eval Res
Hills South Room 152
Amherst MA 01003

Dr Delwyn Harnisch
University of Illinois
51 Gerty Drive
Champaign IL 61820

Dr Patrick R Harrison
Computer Science Department
US Naval Academy
Annapolis MD 21402-5002

Ms Rebecca Hetter
Navy Personnel R&D Center
Code 13
San Diego CA 92152-6800

Dr Thomas M Hirsch
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Professor Paul W Holland
Div of Educ Psych & Quant
Methods Prog
Graduate School of Education
4511 Tolman Hall
University of California-Berkeley
Berkeley CA 94720

Professor Lutz F Hornke
Institut fur Psychologie
RWTH Aachen
Jaegerstrasse 17/19
D-5100 Aachen
WEST GERMANY

Ms Julia S Hough
Cambridge University Press
40 West 20th Street
New York NY 10011

Dr William Howell
Chief Scientist
AFHRL/CA
Brooks AFB TX 78235-5601

Dr Huynh Huynh
College of Education
University of South Carolina
Columbia SC 29208

Dr Martin J Ippel
Center for the Study of
Education and Instruction
Leiden University
PO Box 9555
2300 RB Leiden
THE NETHERLANDS

Dr Robert Jannarone
Elec and Computer Eng Dept
University of South Carolina
Columbia SC 29208

Dr Kumar Joag-dev
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright Street
Champaign IL 61820

Professor Douglas H Jones
Grad Sch of Management
Rutgers The State University NJ
Newark NJ 07102

Dr Brian Junker
Carnegie-Mellon University
Department of Statistics
Pittsburgh PA 15213

Dr Marcel Just
Carnegie-Mellon University
Department of Psychology
Schenley Park
Pittsburgh PA 15213

Dr J L Kaiwi
Code 442/JK
Naval Ocean Systems Center
San Diego CA 92152-5000

Dr Michael Kaplan
Office of Basic Research
US Army Research Institute
5001 Eisenhower Avenue
Alexandria VA 22333-5600

Dr Jeremy Kilpatrick
Department of Mathematics
Education
105 Aderhold Hall
University of Georgia
Athens GA 30602

Ms Hae-Rim Kim
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright Street
Champaign IL 61820

Dr. Jwa-keun Kim
Department of Psychology
Middle Tennessee State University
Murfreesboro TN 37132

Dr Sung-Hoon Kim
KEDI
92-6 Umyeon-Dong
Seocho-Gu
Seoul
SOUTH KOREA

Dr G Gage Kingsbury
Portland Public Schools
Res & Eval Department
501 North Dixon Street
PO Box 3107
Portland OR 97209-3107

Dr William Koch
Box 7246
Meas & Eval Center
University of Texas-Austin
Austin TX 78703

Dr James Kraatz
Computer-based Education
Research Laboratory
University of Illinois
Urbana IL 61801

Dr Patrick Kyllonen
AFHRL/MOEL
Brooks AFB TX 78235

Ms Carolyn Laney
1515 Spencerville Rod
Spencerville MD 20868

Richard Lanterman
Commandant (G-PWP)
US Coast Guard
2100 Second Street SW
Washington DC 20593-0001

Dr Michael Levine
Educational Psychology
210 Education Building
1310 South Sixth Street
Univ of IL at Urbana-Champaign
Champaign IL 61820-6990

Dr Charles Lewis
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541-0001

Mr Hsin-hung Li
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright Street
Champaign IL 61820

Library
Naval Training Systems Center
12350 Research Parkway
Orlando FL 32826-3224

Dr Marcia C Linn
Graduate School of Education
EMST
Tolman Hall
University of California
Berkeley CA 94720

Dr Robert L Linn
Campus Box 249
University of Colorado
Boulder CO 80309-0249

Logicon Inc (Attn: Library)
Tactical & Training Systems Div
PO Box 85158
San Diego CA 92138-5158

Dr Richard Luecht
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Dr George B. Macready
Dept of Meas Stat & Eval
College of Education
University of Maryland
College Park MD 20742

Dr Evans Mandes
George Mason University
4400 University Drive
Fairfax VA 22030

Dr Paul Mayberry
Center for Naval Analysis
4401 Ford Avenue
PO Box 16268
Alexandria VA 22302-0268

Dr James R McBride
HumRRO
6430 Elmhurst Drive
San Diego CA 92120

Mr Christopher McCusker
University of Illinois
Department of Psychology
603 E Daniel Street
Champaign IL 61820

Dr Robert McKinley
Educational Testing Service
Princeton NJ 08541

Dr Joseph McLachlan
Navy Pers Res & Dev Cntr
Code 14
San Diego CA 92152-6800

Alan Mead
c/o Dr Michael Levine
Educational Psychology
210 Education Bldg
University of Illinois
Champaign IL 61801

Dr Timothy Miller
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Dr Robert Mislevy
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Dr Ivo Molenar
Faculteit Sociale Wetenschappen
Rijksuniversiteit Groningen
Grote Kruisstraat 2/1
9712 TS Groningen
The NETHERLANDS

Dr Eiji Muraki
Educational Testing Service
Mail Stop 02-T
Princeton NJ 08541

Dr Ratna Nandakumar
Educational Studies
Willard Hall Room 213E
University of Delaware
Newark DE 19716

Acad Prog & Research Branch
Naval Technical Training
Command
Code N-62
NAS Memphis (75)
Millington TN 30854

Dr W Alan Nicewander
University of Oklahoma
Department of Psychology
Norman OK 73071

Head
Personnel Systems Department
NPRDC (Code 12)
San Diego CA 92152-6800

Director
Training Systems Department
NPRDC (Code 14)
San Diego CA 92152-6800

Library NPRDC
Code 041
San Diego CA 92152-6800

Librarian
Naval Center for Applied
Research in Artificial Intelligence
Naval Research Laboratory
Code 5510
Washington DC 20375-5000

Office of Naval Research
Code 3422
800 N Quincy Street
Arlington VA 22217-5660
(6 Copies)

ONR Resident Representative
New York City
33 Third Avenue - Lower Level
New York NY 10003-9998

Special Asst for Res Management
Chief of Naval Personnel
(PERS-O1JT)
Department of the Navy
Washington DC 20350-2000

Dr Judith Orasanu
Mail Stop 239-1
NASA Ames Research Center
Moffett Field CA 94035

Dr Peter J Pashley
Law School Admission Services
PO Box 40
Newtown PA 18940-0040

Wayne M Patience
American Council on Education
GED Testing Service Suite 20
One Dupont Circle NW
Washington DC 20036

Dept of Administrative Sciences
Code 54
Naval Postgraduate School
Monterey CA 93943-5026

Dr Peter Pirolli
School of Education
University of California
Berkeley CA 94720

Dr Mark D Reckase
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Mr Steve Reise
Department of Psychology
University of California
Riverside CA 92521

Mr Louis Roussos
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright Street
Champaign IL 61820

Dr Donald Rubin
Statistics Department
Science Center Room 608
1 Oxford Street
Harvard University
Cambridge MA 02138

Dr Fumiko Samejima
Department of Psychology
University of Tennessee
310B Austin Peay Bldg
Knoxville TN 37966-0900

Dr Mary Schratz
4100 Parkside
Carlsbad CA 92008

Mr Robert Semmes
N218 Elliott Hall
Department of Psychology
University of Minnesota
Minneapolis MN 55455-0344

Dr Valerie L Shalin
Department of Industrial
Engineering
State University of New York
342 Lawrence D Bell Hall
Buffalo NY 14260

Mr Richard J Shavelson
Graduate School of Education
University of California
Santa Barbara CA 93106

Kathleen Sheehan
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Dr Kazuo Shigemasu
7-9-24 Kugenuma-Kaigan
Fujisawa 251
JAPAN

Dr Randall Shumaker
Naval Research Laboratory
Code 5500
4555 Overlook Avenue SW
Washington DC 20375-5000

Dr Judy Spray
American College Testing
2201 North Dodge Street
PO Box 168
Iowa City IA 52243

Dr Martha Stocking
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Dr William Stout
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright St
Champaign IL 61820

Dr Kikumi Tatsuoka
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Dr David Thissen
Psychometric Laboratory
CB# 3270 Davie Hall
University of North Carolina
Chapel Hill NC 27599-3270

Mr Thomas J Thomas
Federal Express Corporation
Human Resource Development
3035 Director Row Suite 501
Memphis TN 38131

Mr Gary Thomasson
University of Illinois
Educational Psychology
Champaign IL 61820

Dr Howard Wainer
Educational Testing Service
15-T Rosedale Road
Princeton NJ 08541

Elizabeth Wald
Office of Naval Technology
Code 227
800 North Quincy Street
Arlington VA 22217-5000

Dr Michael T Waller
Univ of Wisconsin-Milwaukee
Educ Psychology Department
Box 413
Milwaukee WI 53201

Dr Ming-Mei Wang
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Dr Thomas A Warm
FAA Academy
PO Box 25082
Oklahoma City OK 73125

Dr David J Weiss
N660 Elliott Hall
University of Minnesota
75 E River Road
Minneapolis MN 55455-0344

Dr Douglas Wetzel
Code 15
Navy Personnel R&D Center
San Diego CA 92152-6800

German Military Representative
Personalstammamt
Koelner Str 262
D-5000 Koeln 90
WEST GERMANY

Dr David Wiley
School of Education and Social
Policy
Northwestern University
Evanston IL 60208

Dr Bruce Williams
Department of Educational
Psychology
University of Illinois
Urbana IL 61801

Dr Mark Wilson
School of Education
University of California
Berkeley CA 94720

Dr Eugene Winograd
Department of Psychology
Emory University
Atlanta GA 30322

Dr Martin F Wiskoff
PERSEREC
99 Pacific Street
Suite 4556
Monterey CA 93940

Mr John H Wolfe
Navy Personnel R&D Center
San Diego CA 92152-6800

Dr Kentaro Yamamoto
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Duanli Yan
Educational Testing Service
Mail Stop 03-T
Princeton NJ 08541

Dr Wendy Yen
CTB/McGraw Hill
Del Monte Research Park
Monterey CA 93940

Dr Joseph L Young
National Science Foundation
Room 320
1800 G Street NW
Washington DC 20550