DOCUMENT RESUME

ED 282 898		TM 870 279
AUTHOR	Lawrence, Ida M.; Dorans, Neil J.	
TITLE	An Assessment of the Dimensionality of SAT-Mathematical.	
PUB DATE	Apr 87	
NOTE	48p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (Washington, DC, April 21-23, 1987).	
PUB TYPE	Speeches/Conference Papers (150) -- Reports - Research/Technical (143)	
EDRS PRICE	MF01/PC02 Plus Postage.	
DESCRIPTORS	*College Entrance Examinations; *Factor Analysis; Factor Structure; Guessing (Tests); Higher Education; *Item Analysis; Least Squares Statistics; Mathematical Models; *Mathematics Tests; Test Theory	
IDENTIFIERS	Confirmatory Factor Analysis; Exploratory Factor Analysis; Item Parcels; LISREL Computer Program; *Scholastic Aptitude Test; TESTFACT (Computer Program); *Unidimensionality (Tests)	

ABSTRACT

Six editions of Scholastic Aptitude Test-Mathematical (SAT-M) were factor analyzed using confirmatory and exploratory methods. Confirmatory factor analyses (using the LISREL VI program) were conducted on correlation matrices among item parcels--sums of scores on a small subset of items. Item parcels were constructed to yield correlation matrices amenable to linear factor analyses. The items constituting a parcel measured the same dimension, and parcels measuring the same construct were parallel to each other. Content area (arithmetic, algebra, geometry, and miscellaneous) defined the parcel. Confirmatory factor analyses of item parcel data indicated that the SAT-M editions were unidimensional across different ability populations. Full-information factor analysis (using TESTFACT) was also used to assess dimensionality within item parcels. Analyses assuming a two-parameter model and a three-parameter item response model suggested that all parcels were unidimensional. A third set of analyses involved least-squares factor analyses of a smoothed positive definite matrix of tetrachorics adjusted for guessing. Despite the corrections for guessing, difficulty factors emerged, suggesting that factor analysis of adjusted tetrachorics suffers from the same problems that have plagued other attempts to factor analyze item data. Results supported the unidimensionality of SAT-M, with no empirical justification for reporting subscores based on content. (GDC)

An Assessment of the Dimensionality of SAT-Mathematical[1,2,3]

Ida M. Lawrence
Neil J. Dorans

Educational Testing Service

[1] Paper presented at the Annual Meeting of the National Council on Measurement in Education, Washington, DC, April, 1987.

[2] The authors benefited from thorough reviews by Daniel Eignor, Peter Pashley and Rebecca Zwick. The authors are indebted to Nancy Feryok for her assistance with getting TESTFACT to run on the SAT, and Georgiana Thurston for her valuable secretarial support. Advice from Robert Mislevy, Ledyard Tucker, and Cathy Wendler was also appreciated.

[3] Funding for this research was provided by the ETS-College Board Joint Staff Research and Development Committee.

ABSTRACT

Six editions of SAT-Mathematical (SAT-M) were factor analyzed using confirmatory and exploratory methods. Confirmatory factor analyses (using the LISREL VI program) were conducted on correlation matrices among item parcels for each test edition. An item parcel is a sum of scores on a small subset of items. Item parcels were constructed to yield correlation matrices that were amenable to linear factor analyses.

A critical assumption made in the parcel approach was that the items constituting a parcel measured the same dimension. Another requirement was that parcels measuring the same construct were parallel to each other. In this study, content area (arithmetic, algebra, geometry, and miscellaneous) defined the parcel. Within each content area, several parallel parcels, of 4-7 items each, were constructed. Parallel parcels were constructed by forming parcels of approximate equal difficulty and variability.

The primary method used in this study, confirmatory factor analyses of item parcel data, indicated that the SAT-M editions were unidimensional. Simultaneous confirmatory factor analyses of the same equating section, administered to different ability populations, were also conducted. These analyses were undertaken to test for the hypothesis of factorial invariance across populations. Results indicated that the unidimensional structure of SAT-M was consistent across different ability populations.

Some additional exploratory analyses were also conducted on item-level data. Full-information factor analysis (using TESTFACT) was used to assess dimensionality within item parcels for one test edition. Analyses were performed assuming a two-parameter model and a three-parameter item response model; results suggested that all of the parcels were unidimensional. The results for the two-parameter model were more affected by methodological artifacts than those for the three-parameter model, demonstrating the need for the correction for guessing.

Full-information factor analysis was also used to assess the dimensionality of the 25-item section in one edition of SAT-M, under the assumption of a three-parameter model. Results of this analysis suggest that a slight departure from unidimensionality might be attributed to geometry items.

A third set of exploratory item-level analyses involved least-squares factor analyses of a smoothed positive definite matrix of tetrachorics adjusted for guessing. Three sixty-item editions of SAT-M were analyzed, under two types of scoring: right/wrong/missing and right/wrong. Despite the corrections for guessing and omitting, difficulty factors emerged, and held up across all three content areas, suggesting that factor analysis of adjusted tetrachorics suffers from the same problems that have plagued most other attempts to factor analyze item data.

The results of this study are seen as evidence that SAT-M is unidimensional. There appears to be no empirical justification for reporting subscores based on content. A method for factor analyzing test data, at the item parcel level, was presented as a possible way to examine dimensionality of test data, while avoiding some of the problems associated with item level factor analyses.

An Assessment of the Dimensionality of SAT-Mathematical

Ida M. Lawrence
Neil J. Dorans

Educational Testing Service

The goal of this research was to obtain a fuller understanding of what is being measured by the mathematical portion of the Scholastic Aptitude Test (SAT-M). Because SAT-M covers a variety of content areas, based on different mathematics curriculum, it is important to understand the degree to which the test conforms to a unidimensional model.

This research sought to answer three specific questions:

1) Is SAT-M measuring more than one dimension, and if so, how might these dimensions be characterized?

2) To what extent is the dimensional structure of SAT-M invariant across test editions?

3) To what extent is the dimensional structure of SAT-M invariant across populations of examinees differing in ability?

In addition to learning more about the dimensionality of SAT-M, another objective of this research was to evaluate possible techniques for factor analyzing item data. A variety of methods have been advanced for assessing the dimensionality of binary-scored data. Comprehensive reviews of several procedures are found in papers by Hattie (1984, 1985) and Mislevy (1986). The procedures fall into two general categories, IRT-only approaches and factor model approaches. Applications of both kinds of approaches are summarized in Dorans and Lawrence (1987).

A complete review of the literature documenting the theoretical and practical problems involved in the linear factor analysis of binary scored data

is beyond the scope of this paper. Dorans and Lawrence (1987) provide a brief review of which the most salient points are summarized here. The existence of additional artifactual or "difficulty" factors when phi coefficients are factor analyzed via a linear model has been a well discussed phenomenon in the literature. Factor analysis of tetrachoric correlation coefficients theoretically can circumvent the problem of "difficulty" factors for free-response items that are right/wrong scored. However, other problems can occur when tetrachorics are factor analyzed and the tetrachoric correlation coefficients are based on binary scored multiple-choice items where guessing is possible. In this context, failure to take guessing effects into account will again produce artifactual factors and misleading information as to the number of factors needed to account for the data. Given the practical problems involved in the linear factor analysis of binary scored item response data using tetrachorics (i.e., the matrices are often non-positive definite) and the assumptions that must be met in order for the procedure to be viable (no or correctable guessing and normally distributed traits), other approaches that provide viable options to the problem of assessing item level dimensionality have been developed.

One set of approaches involves a blending of factor analytic and item response theory techniques. These procedures involve a generalized least squares approach attributable to Christoffersson (1975) and marginal maximum likelihood full information factor analysis (used in this research) based on the work of Bock and Aitkin (1981). Mislevy (1986) has provided an excellent review of these approaches, along with the closely related procedure attributable to Muthen (1978, 1984).

This research also used an approach which involves the linear factor analysis of item parcel data, or mini-tests, made up of small collections of non-overlapping items thought to measure the same underlying dimension or dimensions. Data on individual items are no longer used directly in deriving the correlation matrix. Cattell (1956; 1974) was an early advocate of this approach. Dorans and Lawrence (1987) discuss some of the critical issues involved in parcel construction.

Methods Used to Assess SAT Dimensionality

Three approaches to dimensionality assessment were employed to assess the dimensionality of the SAT-Mathematical tests. The primary approach involved using the LISREL VI (Jöreskog and Sörbom, 1984) computer program to test specific models for the structure underlying item parcel data. This approach has been used with success on SAT-Verbal data and Mathematics Level II Achievement Test data in a pilot test mode by Cook, Dorans, Eignor and Petersen (1985). In Dorans and Lawrence (1987), this linear confirmatory factor analytic approach was used on parcels composed of both final form SAT item data and parcels composed of items used for score equating.

The maximum likelihood full information factor analysis approach (Bock, Gibbons, and Muraki, 1986), implemented in the computer program TESTFACT, was used primarily to assess the dimensionality of items within a given parcel. The full information factor analysis model was used in this way as a check on the unidimensionality assumption of parcels that is explicitly made by the parcel approach. The excessive cost associated with running TESTFACT on an entire test precluded using this program on intact SAT test forms.

A third approach was also employed, namely the use of TESTFACT to produce a least squares solution to a smoothed positive definite matrix of tetrachoric

correlations that had been corrected for guessing. As will be seen in the results section, this use of TESTFACT as a more traditional approach to the factor analysis of item data does not seem to avert the identification of difficulty factors.

In the remainder of this section, the models employed in our analyses are formally stated. We start with the LISREL models for item parcel data, move on to the full information factor analysis models and finish with the analysis of adjusted tetrachorics.

LISREL Analysis of Item Parcels

The use of LISREL on item parcel data to assess test dimensionality is a cost-effective way of assessing how well postulated structural models fit the data. The parcel approach attempts to circumvent the problems associated with factoring item data by factoring item parcel scores, i.e., sums of scores on a small subset of items, which are more amenable to analysis by a linear factor model than item data. A critical assumption made by the parcel approach is that the items constituting a parcel or mini-test measure the same dimension. It was hoped that TESTFACT could be used to test this within-parcel unidimensionality assumption.

Parcel construction principles. As noted earlier, it is well documented, e.g., Carroll (1945, 1983) that linear factor analysis of a matrix of phi coefficients based on binary item data produced by a unidimensional model for continuous data, will be viewed as multidimensional with a second dimension clearly related to item difficulty. As McDonald and Ahlawat (1974) argue, part of the problem is that a linear regression model is inappropriate for the

item/factor regression, which has to be nonlinear given the bounded nature of dichotomous data and the unbounded metric assumed for the underlying factor.

Mislevy (1986) summarizes the problems with analyzing phi coefficients quite succinctly:

> When binary variables are produced by dichotomizing continuous variables, then, the choice of cutting points materially affects the values of the expected phi coefficients. Factor analysis of phi coefficients of binary variables produced by the same underlying correlational structure but dichotomized at different points can conform to factor models with different structures and possibly different numbers of factors. (pp. 9-10).

For the parcel approach to avoid the problems of factoring phi coefficients, the parcels must be constructed in a fashion that is sensitive to these problems. The major reason for constructing parcel scores is to achieve a matrix of correlations or covariances that is not affected by item difficulty and the nonlinearity of the item/factor regression. Parcel construction should attempt to "linearize" the data by attempting to remove the effects of nonlinearity and differences in item difficulty.

To mitigate the effects of differences in item difficulty and nonlinearity, parcel scores should have approximately equal means and variances. In the terminology of classical test theory, the parcels should be constructed to be parallel to each other. To achieve parallel parcels, it is essential to place approximately equal numbers of easy, middle difficulty and hard items within each parcel such that each parallel parcel is composed of several nonparallel items.

A critical question that needs to be addressed is how many items are needed for a parcel. Experience (Drasgow and Dorans, 1982) indicates that a minimum of at least three is needed and that six or seven is clearly enough <u>provided</u> that

the items within a parcel are adequately spaced to achieve a situation in which the probabilities associated with the parcel score distribution is approximately normal. A statistical justification for the parallel parcelling approach might be drawn from the work of Drasgow and Dorans (1982) where they introduce the notion of a categorization attenuation factor that reduces the correlation between two continuous variables when one variable is polychotomized. Parallel parcelling can be viewed as a heuristic approach to converting dichotomous data into polychotomous data with an eye toward minimizing the size of the categorization attenuation factor.

LISREL VI: First-order and second-order models. The LISREL VI computer program (Joreskog and Sorbom, 1984) fits and tests models for linear structural relationships among quantitative variables. As mentioned earlier, the primary reason for developing item parcels was to yield variance-covariance matrices that were amenable to a linear factor analysis. Both first-order factor analysis and second-order factor analysis are special cases of the LISREL VI model.

One goal of a factor analysis is to identify the number of common factors needed to fit the off-diagonal elements of the variance/covariance matrix. This is known as the number of factors problem. LISREL VI was used to assess the number of factors problem in the following fashion. For each test edition studied, the fit of a one-factor model to the correlation matrix among item parcels (correlation matrices were used to simplify proportion of variance interpretations and reduce the impact of variable length parcels on the multifactor solutions), was examined[1]. Next, the fit of a two common factor model to the same data was examined.

[1] For the cross-population analysis of equating section parcels, covariance matrices were factor analyzed to assess the equality of factor structures, as will be seen later.

Finally, a second-order factor model was used. A second-order factor analysis can be thought of as a factor analysis of the first-order factors (See Schmidt and Leiman, 1957, for a discussion of one approach to hierarchical or second-order factor analysis.) It is a particularly fruitful approach to employ a second-order model when one suspects that correlations among the first-order factors can be explained by a single second-order common factor. Such a model is particularly applicable to item parcel data that one suspects is essentially unidimensional. Drasgow and Parsons (1983) suggested a second-order factor model that influenced the choice of the model and approach used in the Cook, Dorans and Eignor (in press) study. That same approach was used here.

This second-order factor model decomposes each first-order factor into a second-order common factor that influences all first-order factors, and a second-order group factor which influences performance only on that first-order factor. Another way of stating this is that second-order group factors are uncorrelated with each other and with the second-order general factor. If the contribution of the second-order common factor to every first-order factor is large, the correlations among the first-order factors will be close to unity. If the second-order group factor for a particular first-order factor is relatively large, then the correlations of that first-order factor with other first-order factors will be among the lowest in the first-order factors correlation matrix.

To summarize, both first-order factor analyses and second-order factor analyses were employed. The first-order analyses focused on the number of factors issue. Both the first-order and the second-order analyses were focused on assessing hypothesized structures suggested by the item types and content areas measured by the tests. Fit of the model to the data was the dominant

concern in the first-order analyses. Decomposition of first-order factor variance into second-order common and group specific components was the main concern of the second-order analyses.

LISREL VI's indices of fit. LISREL VI provides several indices of fit that are described by Joreskog and Sorbom (1984). When LISREL VI provides maximum likelihood estimates of free parameters, it also provides the likelihood ratio χ^2 statistic with associated degrees of freedom and probability level. Ideally, this index should be helpful in assessing competing models for the data because, under certain conditions, the difference in χ^2 values is itself chi square distributed with degrees of freedom equal to the difference in degrees of freedom associated with the two competing models. However, it is important to keep in mind that this difference in χ^2 values is asymptotically distributed as chi square only if one model is a special case of the other model and the more general model is true. This difference in χ^2 values indicates whether the parameters that are estimated in the more general model add anything to the fit of the model for the data. It should be noted that Joreskog and Sorbom also cite several other reasons why the χ^2 indices should be used with caution.

Another goodness of fit index provided by LISREL VI is the root mean square residual,

$$\text{(1)} \qquad RMSR = [2 \sum_{i=1}^{n} \sum_{j=1}^{n} (c_{ij} - \hat{c}_{ij})^2 / (k+1)k]^{1/2} ,$$

where k is the number of observed variables, and c_{ij} and c_{ij} are elements of the observed and fitted covariance matrices. The RMSR is a useful descriptive index for comparing the fit of two different models for the data.

In addition to these indices of global fit, LISREL VI provides individual residuals in both raw and normalized forms. The normalized residuals are taken

from standard asymptotics based on normality, i.e., the raw residual divided by an estimate of its standard error; hence the normalized residual is assumed to be asymptotically a standard normal variable. The formula for a normalized residual is

$$NR = \frac{c_{ij} - \hat{c}_{ij}}{(c_{ii}c_{jj} + c_{ij}^{2})/N} \tag{2}$$

where c_{ij} and $\hat{c}_{ij}$ are elements of the observed and fitted covariance matrices. Joreskog and Sorbom (1984) suggest that normalized residuals with values greater than two in absolute value merit close examination. In assessing model fit, primary attention was paid to the pattern of normalized residuals (referenced to hereafter as NR).

Full Information Factor Analysis of Item Data

The TESTFACT program (Wilson, Wood, and Gibbons, 1984) was used to obtain full information factor analysis solutions for selected item parcels. Eock, Gibbons and Muraki (1986) describe the theory behind the TESTFACT program. Mislevy (1986) also describes the theory in his description of recent developments in the factor analysis of categorical variables.

This factor analytic model operates on information contained in the joint frequencies of the 2^p contingency tables of response counts on a p-item test. Observed performance on an item is presumed to be obtained via a dichotomizing process performed on the unobserved continuous variable measured by the item. For each item, there is an assumed threshold parameter which identifies the location along the continuous variable at which the dichotomous "chop" occurs. The probability of a correct response to an item is a function of examinee

ability with respect to one or more latent factors and the location of the threshold parameter along the continuous variable.

The full-information factor analysis solution is applied to a matrix of distinct response patterns of rights and wrongs to obtain estimates of factor loadings and thresholds for each item. Stated differently, TESTFACT is used to estimate discrimination and difficulty parameters for item data based on a multidimensional IRT model. The model allows for input of lower asymptotes for each item, as a way to take into account the effects of guessing.

The full information solution is a maximum likelihood solution in which estimates of the loadings and threshold for each item are obtained. The orientation of the factors is orthogonal. In an effort to achieve interpretable results, TESTFACT allows both orthogonal and oblique rotations through use of the VARIMAX (Kaiser, 1958) and PROMAX (Hendrickson and White, 1964) rotational procedures, to be applied to the orthogonal solution.

The TESTFACT program provides standard errors of estimation and statistical tests of fit. In particular, the program automatically produces a test of differences in chi-squares for nested models, which is used to determine the number of factors. Zwick (1987) states that the test statistic is not distributed as a chi-square with applications of TESTFACT that use Bayesian priors to constrain parameter estimations because the solution does not involve maximization of a likelihood function under these circumstances. Unfortunately, the program does not provide raw or normalized residuals for a fitted correlation or covariance matrix, as does LISREL, for assessing model fit. Instead, the residual matrix reported in TESTFACT output is computed as the difference between the initial tetrachoric matrix and the smoothed one, which does not indicate how well the factor model fits the data.

The program also produces estimates of amounts and proportions of variances accounted for by the underlying factors and communality estimates for the response process variables. Hence, a scree test can be easily computed to assess dimensionality. In addition, reasonableness checks on communality estimates can be performed. For example, overfactoring is probably indicated if an item's communality exceeds the total test reliability. In other words, determination of the number of factors issue can be approached from many different vantage points when using TESTFACT. Over reliance on a sometimes suspect statistical test is not the only course of action available to TESTFACT users.

Adjusted Tetrachorics Analyses

Factor analysis of a matrix of tetrachoric correlations was, and still is, a more traditional approach taken to circumvent the problems associated with factoring a matrix of phi coefficients. If one assumes that the observed counts in a 2-by-2 table of corrects and incorrects for items j and k arose through dichotomization of two normally-distributed continuous variables, then one can estimate from these 2-by-2 tables of counts the correlation among these underlying item response process variables. In theory, these correlations can be factor analyzed to produce results in which "difficulty" factors are no longer present.

In practice, the factoring of tetrachorics is fraught with many difficulties. First, guessing can have a differential impact on the 2-by-2 counts depending on the item's difficulty. In addition, extreme values of the underlying correlation are poorly estimated resulting in many arbitrary +1.0's and -1.0's. Also, since the tetrachorics are estimated pairwise, inconsistencies can occur across different pairs. Finally, as Mislevy (1986)

notes, there are neither standard errors nor statistical tests associated with traditional tetrachoric analysis, and tetrachorics, estimation problems aside, do not describe all the information contained in the 2^p contingency table of response counts on a p item test. In fact, the full information factor analysis approach that was just described attempts to use all the information in the 2^p table to assess the dimensionality in a set of items, while the generalized least squares approaches of Christofferson (1975) and Muthen (1978), which Mislevy (1986) ably describes and aptly refers to as "partial information" approaches also use more data than standard tetrachoric analyses. Relative to the analysis of tetrachorics, the full information and partial information approaches are quite expensive. In other words, tetrachoric analysis is attractive because it is inexpensive.

TESTFACT appears to be able to deal with many of the problems that have plagued traditional tetrachoric analysis. Zwick (1986) has found a way to use TESTFACT to produce a relatively inexpensive unweighted least squares analysis of a smoothed positive definite tetrachoric correlation matrix that can be adjusted for omits and guessing. We used this approach on entire test forms because Zwick's experience indicated that results closely mirrored the full information solution at a fraction of the cost. Bock, Gibbons and Muraki (1986) describe the particulars of the various adjustments made to the tetrachorics.

Procedures

Data Source

The data analyzed in this study were obtained from six editions of SAT-Mathematical. Each edition contains two operational SAT-Mathematical sections that produce a SAT-Mathematical score based on a total of 60 items (40 five-choice items and 20 four-choice items). The mathematical questions require

applications of mathematical techniques to solve items from three content areas: arithmetic, algebra, and geometry. The three content areas are represented by approximately the same number of questions. Items which cannot be classified by content are placed in a group referred to as "miscellaneous" (in general, seven or eight items per test form are placed in this category).

Choice of Editions for Factor Analyses

The six editions analyzed in this study were selected from two previous equating data collection designs. With the basic SAT equating data collection design, one edition (Z) is administered to one group of examinees, a second edition (X) is administered to a second group of examinees, and a third edition (Y) is administered to a third group of examinees. In general, the examinee groups taking editions X and Z represent populations of similar ability, and the group taking edition Y represents either a less able or more able candidate population. Thus, factor analysis of SAT data from equating samples provides a means for assessing the dimensionality of editions administered to examinee groups of varying ability that are representative of actual SAT populations.

Two of the six editions were from January administrations of the SAT, where SAT-Mathematical means tend to be below the yearly average: In January 1982 the SAT-Mathematical mean was 435, while in January 1983 it was 431. Two of the six editions were from June administrations, where the preponderance of test-takers are high school juniors: In June 1985, the mean was 477, while the corresponding mean in June 1986 was 477. The last two editions were from November administrations, which are predominantly high school senior populations: The November 1983 mean was 477, while the corresponding mean was 485 in November 1985.

The January 1983 edition was part of the June 1986 equating calibration design, which also included the June 1985 edition. The January 1982 edition was part of the November 1985 equating calibration design, which also included the November 1986 edition.

In addition to examining the dimensional structure across different editions of SAT-Mathematics, another goal of this research was to determine the extent to which dimensional structure of the test is invariant across populations differing in ability. While operational sections of the SAT are generally administered to a large population only once, the same equating test is administered to several large populations. As part of the equating data collection design, edition Z is linked to edition X via one equating test, and to edition Y via a different equating test, as indicated in Figure 1. The common anchor tests from this design provide an opportunity to assess the stability of equating test factor structure across different populations. All samples in these analyses involved approximately 3,000 examinees.

Formation of Item Parcels

Expanding upon the methodology used by Cook, Dorans, and Eignor (in press), items from each SAT-Mathematical edition were separated into parallel item subsets, referred to in this research as item parcels, using the principles described earlier in this report. Parcels of approximate equivalent difficulty and standard deviation were formed by selecting items based on their observed p-values (computed as the number of examinees answering the item correctly, divided by the number of examinees in the sample). Following the formation of item parcels, scores on the parcels were computed for each examinee, using a binary right/wrong scoring of the item data. Item parcel scores within each edition were then intercorrelated, and the resulting correlation matrices served

as input for linear confirmatory factor analyses (covariance matrices were used in the cross-population analyses).

For SAT-Mathematical operational tests, and equating sections, content area defined the parcel. Within each content area (arithmetic, algebra, geometry, and miscellaneous) items were placed into parcels of four to seven items each.

Hypothesized Factor Structures for Confirmatory LISREL Analyses

Eleven item parcels for five of the mathematical test forms were constructed, as follows: three arithmetic parcels, three algebra parcels, three geometry parcels, and two miscellaneous parcels. One of the editions (administered in January 1982) was separated into only ten, rather than eleven, parcels because it had fewer miscellaneous items and only one parcel in that category was needed. In forming item parcels, four-choice items and five-choice items were distributed across the parcels in equal numbers.

The structure for the factor pattern matrices and factor correlation matrices for the SAT-Mathematical analyses are presented in Figure 2. Underlying the three-factor solution depicted in Figure 2 is a second-order model, which assumes that a general mathematical factor and three content-related first-order factors explain the common portion component of item parcel correlations. For this model, and the other models, item parcels for miscellaneous items are assumed to load on all of the first-order factors. The one-factor model assumes that SAT-Mathematical is a unidimensional test.The two-factor solution is a first-order model which assumes that algebra and arithmetic item parcels load on one factor, and geometry item parcels load on a second factor.

Factorial Invariance of Equating Sections. In constructing item parcels for the equating sections, the same parcel definitions were used, i.e., content

area. For the mathematical equating sections, which contain 25 five-choice items, seven item parcels were constructed (two for each content area, and one for the miscellaneous items).

Dimensional structure of equating sections was assessed using the same factor structures as were found to underlie the mathematical operational tests, i.e., as displayed in Figure 1, although each factor is defined by fewer parcels. Factorial invariance across populations on a particular equating section was assessed by evaluating the fit of a factor model which assumes equivalent factor pattern matrices for the two samples taking the same equating section. Parcel covariance matrices, rather than parcel correlation matrices, were analyzed in the factorial invariance analyses for reasons cited in Joreskog (1971) and Meredith (1964).

Factorial invariance of factor structures across populations was assessed by applying separate simultaneous factor analyses to each of four verbal equating sections and four mathematical equating sections. LISREL was used to examine the hypothesis that the factor pattern underlying parcel covariances for a particular equating section is the same in two populations. To study factorial invariance, it was necessary, for each equating section, to estimate a model that constrains the same common factor pattern matrix over the two populations of interest. Distributions of NRs and RMSR for this constrained model were compared to fit indices resulting from a model that does not place an equality constraint on the factor pattern in each population (i.e., the factor pattern is estimated separately within each population). If the constrained model is found to fit the data as well as the unconstrained model, we may conclude that the factor structure underlying the parcel covariances is the same in each population receiving the same equating section.

Exploratory Factor Analysis

One of the major criticisms lodged against the parcel approach is that it assumes that items are unidimensional within parcels. To deal with this criticism, our original intention was to use the full information factor analysis model to assess dimensionality within parcels. As we attempted to use TESTFACT in this manner, we encountered several obstacles, including TESTFACT's "user-unfriendliness" and its cost. In a spirit consonant with exploratory factor analysis, Dorans and Lawrence (1987) began to experiment with various options available in the TESTFACT program. The goal of our exploration was to find a cost-effective way of using TESTFACT that could be used in conjunction with the relatively inexpensive use of LISREL to analyze parcel data. In the process, we examined a cost effective alternative to the parcel approach. All our exploration involved data from the June 1986 equating data collection design, which included the June 1985 and January 1983 editions. We performed three major types of analyses.

First, for the June 1986 SAT-Mathematical, we used the full information factor analysis approach to assess dimensionality within parcels. For these analyses, item data was scored right/wrong/missing and the adjustment for missing data described in Bock, Gibbons and Muraki (1986) was used. On a formula-scored test like the SAT, examinees tend to omit very difficult items. In addition, some examinees do not reach all test items. Hence, there exists missing data that needs to be treated differently than right/wrong. Analyses were performed with and without Carroll's (1945) correction for guessing which is also described in Bock et al. (1986). Estimates of the lower asymptote from IRT item calibrations under the three-parameter logistic model were used for the correction for guessing.

Second, as an alternative to the parcel approach, we used TESTFACT to obtain least squares factor analyses of a smoothed positive definite matrix of tetrachorics corrected for guessing on the full SAT-Mathematical editions administered in June 1986, June 1985 and January 1983. This analysis is a cost-effective way of using TESTFACT to factor analyze item data. It was performed under two item scoring conditions: omits and not reached treated as wrong, i.e., the right/wrong condition, and omits and not reached treated as missing and adjusted via the procedure described in Bock, Gibbons and Muraki (1986).

Third, the full information factor analysis approach was applied to the 25-item M1 section of the January 1983 edition.

Results and Discussion

Confirmatory Factor Analyses of SAT-Mathematical Item Parcels

Table 1 contains distributions of normalized residuals (NRs) for the factor analyses of SAT-Mathematical item parcels. Each panel in Table 1 (one for each of the six editions) contains a distribution of NRs associated with the three solutions of interest (see Figure 2): (1) a one-factor solution, (2) a two-factor solution, which assumes a second factor defined by geometry item parcels, and (3) a three-factor solution which hypothesizes a separate factor for each content area. For each solution, the root mean square raw residual (RMSR), which provides a summary index of the fit of the model, is displayed in Table 2.

The information contained in these tables reveals that SAT-Mathematical is clearly unidimensional. For the six editions studied, a solution with a single factor provides an excellent fit to the item parcel data. Out of a possible 55 NRs associated with each of five of the test editions, the number of NRs greater

than 2.0 standard deviation units is one or two per edition, typically involving the correlation between two miscellaneous parcels. With respect to the sixth test edition (January 1982), 2 out of a possible 45 NRs are greater than 2.0. Thus, with the exception of one or two data points, virtually all of the inter-parcel correlations fitted by a one-factor model have NRs within a band which ranges between -2.0 and + 2.0.

The data contained in these tables also indicates that addition of a second factor, defined by geometry item parcels, provides little in terms of improvement in fit to the data. As can be seen in Table 2, differences in RMSR for a one-factor solution and a two-factor solution are slight, ranging between .02 and .06. Finally, looking at results for the three-factor solution, we see that improvement in fit to the data is trivial. In fact, for two of the editions the item parcels are so collinear that LISREL VI could not extract a third factor.

In addition to assessing model fit, another, more substantive approach to determining whether SAT-Mathematical is unidimensional is to examine the intercorrelations among the factors defined by the mathematical content areas. These results are displayed in Table 3, which presents factor intercorrelations by edition for the three-factor solution, and for the two-factor solution in cases where a third factor was not extracted. For the solution with three factors, correlations are all above .92, and several are as high as .99. Within this limited range, the correlation between algebra and arithmetic is always higher than the correlations between algebra or arithmetic with geometry. The consistency of this finding for four editions suggests that geometry parcels may be measuring a construct which differs slightly from what is being measured by algebra and arithmetic parcels. However, for the two editions where it was not

possible to extract a third factor, the correlations between factors in the two-factor solution are exceptionally high (.96 and .95).

Table 4 displays the relative contributions of one general factor and three content specific factors to the variance of the first-order factors based on a second-order factor solution. Again, data are presented for four of the six editions, as it was not possible to fit a second-order solution for all of the editions. Comparing across editions, the general factor accounts for typically 99 percent of the algebra parcel variance, about 98 percent of the arithmetic parcel variance, and about 90 percent of the geometry parcel variance. From this we may conclude that the general factor is slightly less related to the geometry factor than it is to the algebra and arithmetic factors.

In conclusion, confirmatory factor analyses of item parcel data for SAT-Mathematical provide evidence that the test is essentially unidimensional. This conclusion is buttressed by findings which are consistent across several editions administered to populations of varying ability.

Confirmatory Factor Analyses of Mathematical Equating Section Item Parcels

Distributions of NRs for the factor analyses of mathematical equating section item parcels are displayed in Table 5. Each panel in the table focuses on model fit for a single equating section that was administered to two different large populations. LISREL was used to factor analyze item parcel covariance matrices from two populations simultaneously.

For each mathematical equating section, two analyses were done. In the first analysis, referred to in the table as "Within-population", a separate factor structure is assumed to explain the item parcel covariance data associated with each population. In the second analysis, referred to in the table as "Between-populations", an equivalent factor structure is assumed to

underlie the item parcel covariance data within each population. The within-population distributions of NRs portray the overall fit of each model within each population. The between-population distributions of NRs indicate the fit of the hypothesized model when factor loadings are constrained to be equal across both populations.

The RMSR for each within-population and between-population solution is presented in Table 6. The last column in the table, "RMSR Difference", indicates the loss in model fit as a result of imposing the equal factor loadings constraint on item parcel covariance data from the two populations.

Distributions of NRs and RMSR for the within-population analyses indicate that a one-factor solution provides an excellent fit to the item parcel covariance data. A similar finding was found with respect to the six editions of SAT-Mathematical (see earlier section).

Inspection of between-population NRs and RMSR suggest that a common loading matrix fits the data reasonably well in each of the two populations taking equating sections gx, iv, and jp, respectively. The exception to this finding is with respect to equating section il, which was administered to a January population and a June population. The normalized residual matrices (not shown) for these analyses show larger residuals associated with geometry and miscellaneous item parcels. One possibility is that curriculum experience differences between a June administration (primarily juniors) and a January administration (mostly seniors) may be responsible for the apparent lack of model fit when the loading matrix is constrained equal across the two populations. This does not appear to be the case for equating section gx, which was administered to a January population and a November population, both primarily senior populations where differences in curriculum would be expected

to have a smaller effect. However, relative to both of these equating sections, better between-population fit is observed when the factor structure is assumed equal over populations of similar ability (i.e., equating sections iv and jp).

Synthesis of Confirmatory Factor Analysis Results

The confirmatory factor analysis results for the six editions of SAT-Mathematical and the four Mathematical equating sections strongly indicate that the SAT-Mathematical Test is unidimensional. Of the three content areas, geometry content seems to exhibit the most unique variance. The average loading of the geometry factor on the general factor is .95, which is the expected correlation between an infinitely long total mathematics score and an infinitely long geometry score. Hence, there is little empirical justification for reporting subscores for SAT-Mathematical on the basis of content.

Exploratory Factor Analyses Results for SAT-Mathematical

Three types of exploratory analyses were performed: (1) Within-parcel full information factor analyses; (2) analyses of adjusted tetrachorics; and (3) full information factor analysis of an intact section. The results for each of these solutions are presented in order.

Within-parcel analyses. Each of the 11 parcels for the June 1986 edition of the SAT-Mathematical were subjected to two types of full information factor analyses: one involving a correction for guessing, which we refer to as the 3PNO solution (for three parameter normal ogive), and one involving no correction for guessing, the 2PNO solution. Both solutions used item data scored as right/wrong/missing.

Table 7 contains a summary of the results of these 22 TESTFACT runs. Running down the middle of the table are each parcel's label, the number of

items in each parcel and the parcel's KR-20 reliability estimate. With the exception of the last miscellaneous parcel (MIS2), parcels were composed of five or six items and had reliabilities ranging from .40 to .60.

Testing the unidimensionality within each parcel was the purpose of these TESTFACT analyses. Two criteria were used for assessing unidimensionality: the number of latent roots of the adjusted tetrachorics matrix that are greater than one, and the number of factors significant at the .05 level.

For the 3PNO model, the left-hand portion of Table 7, the root criterion indicated that all 11 parcels were unidimensional In contrast, five of the 11 full information solutions indicated a second sign. . ant factor. Four of these five solutions seemed to suffer from overfactoring in that the PROMAX rotation produced an orientation of factors in which the second factor was marked by a single item only, and several of the items had communality estimates that were higher than the parcel KR-20, suggesting that items were more reliable than their composite, an unreasonable result. The fifth solution had the markings of a solution that contained difficulty factors. The easier items on GE01 loaded on one factor, while the harder items loaded on the second factor. The significance test criterion seemed to lead to overfactoring.

The results for the 2PNO model were more affected by methodological artifacts than those for the 3PNO model, demonstrating the need for the correction for guessing.

Adjusted tetrachorics analyses. The full 60-item editions of SAT-Mathematical that were administered in January 1983, June 1985 and June 1986 were subjected to least square analyses of a smoothed positive definite matrix of tetrachorics adjusted for guessing under two types of item scoring: right/wrong/missing and right/wrong. Table 8 contains details about the number

of factors, number of tetrachorics set to 1 (for reasons of sparse data in some of the cells of the 2x2 table of corrects and incorrects) prior to smoothing, and correlations between the first and second PROMAX factors.

The proportions of variance accounted for by the first four roots of the tetrachoric matrix are listed in Table 8. With the possible exception of the January 1983 right/wrong solution, two factors seem to be indicated. Note that one effect of treating missing data as wrong is to make the solution appear more unidimensional. One reason for this more unidimensional appearance is the large number of tetrachorics set equal to one under right/wrong scoring. Note that 17.3% of the tetrachorics for the January 1983 solution, which appears most unidimensional, are set to one. In contrast, under right/wrong/missing scoring, less than 1% of the tetrachorics are set equal to 1 for the January 1983 solution.

The PROMAX solutions for all six factor analyses of adjusted tetrachorics result in two factors that may be labelled "easy" and "hard." Cross-tabulations of the loadings on factors I and II versus the difficulty of the item are presented for all six solutions in Table 9. The rules for classifying items onto factor I, factor II or both I and II (I/II) and in terms of difficulty (E, M or H for easy, middle, or hard, respectively) are given at the foot of the table. These cross-tabulations clearly justify the "easy" label for factor I and the "hard" label for factor II. Note that the effects of the right/wrong scoring are two-fold: it shifts items from easy towards hard and tends to disguise the difficulty factors a bit. Tables 10, 11 and 12 present cross-tabulations for each of the three content areas, arithmetic, algebra and geometry. The difficulty factors hold up across all three content areas.

In sum, the tetrachoric analyses seem to be susceptible to the "difficulty" factor bugaboo that has plagued factor analysis of item data, despite the corrections for guessing and missing and the smoothing that went into the production of the final matrix of tetrachorics.

Full information factor analysis of an intact section. The 25-item M1 section of the January 1983 edition of the SAT-Mathematical Test was analyzed under the 3PNO full information model which operated on items scored as right, wrong or missing. Two factors were extracted at a cost of about $250. We stopped at two factors for cost reasons and because differences in proportions of variance accounted for by successive factors suggested a two-factor solution was adequate. The second factor offered a statistically significant improvement over the one-factor solution. The cross-tabulations at the bottom of Table 13 indicate that the PROMAX loadings are not as related to difficulty as they were for the tetrachoric solutions. In fact, the table at the right suggests that the first factor might be a geometry factor, which is consistent with the LISREL analyses which suggested that any departure from unidimensionality, albeit slight, might be attributed to the geometry items. This application of full information factor analyses seems to be somewhat successful in that difficulty factors were avoided and an interpretable two-factor solution was achieved.

Discussion

Dimensionality analyses of SAT-Mathematical indicate that the test is unidimensional. Confirmatory factor analyses of item parcel data provide evidence that a single-factor solution provides an excellent fit to the data. This conclusion of unidimensionality is partially borne out by exploratory factor analysis at the item-level, using full-information factor analysis with a three-parameter normal ogive model on items scores as right/wrong/missing. In

sum, it is apparent that subscores based on content are not needed for SAT-Mathematical.

The analyses conducted in this study have served to underscore the value of using confirmatory factor analysis of item parcel data to study the dimensional structure of test data. This approach is computationally inexpensive, and appears to provide meaningful and consistent results. The use of parallel parcels makes item data amenable to linear factor analysis. The method seems circumvent the problems associated with directly factor analyzing item data, namely, the propagation of artifactual "difficulty" factors. The method can also avoid the problem of observing a "speed" factor, as items from later positions in the test can be balanced across parcels. In sum, the parallel parcel approach can be used to dispense with difficulty and speed factors and, hence, obtain a clearer look at the substantive factor structure of the test.

One criticism of confirmatory factor analysis is that the approach enables one to "find what one is looking for". This criticism does not bear out in this study, as can be seen from the results of factor analyzing SAT-Mathematical. SAT-Mathematical is an example where item content might be expected to emerge as a factor, yet the one-factor model fits the data as well as a model hypothesizing separate factors for different content areas.

The formation of parallel item parcels can be time-consuming. Routine use of the parcel approach to assess test dimensionality would be facilitated if a computerized algorithm for building parallel parcels were developed. Investigation of the categorization attenuation factor (Drasgow & Dorans, 1982) might prove fruitful.

The within-parcel analyses conducted in this research indicated that assuming a three-parameter logistic model for the items provides more meaningful

results than assuming a two-parameter model. For SAT data it was possible to assume a three-parameter model by using as guessing parameters the c-values estimated in previous IRT calibrations of the items (using LOGIST). Bock and colleagues (1986) argue for using actual c-values with TESTFACT.

In addition to, or instead of, assessing dimensionality within item parcels, it might be useful to apply full-information factor analysis to a subset of items comprising parcels of the same item type or content area (i.e., factor analyze all algebra items). Unidimensionality of the subset of items would be evidence for unidimensionality of parcels comprising these items. However, in order to curtail computer costs, full-information factor analysis should be restricted to fifteen or fewer items.

The use of TESTFACT to assess test dimensionality by factor analyzing a smoothed positive definite matrix of tetrachorics seems to suffer from the "difficulty" factor problem that has plagued most previous attempts at factor analyzing item data. It appears that the correction for missing data (omitted and not-reached items) used by TESTFACT leads to cleaner "difficulty" factors than the treatment of missing data as an incorrect response. This enhancement of interpretability of the "difficulty" factor might be due to the fact that the correction for missing data used in TESTFACT is essentially consistent with a correction that would follow from a missing at random assumption. A conditional missing at random assumption that takes examinee ability as well as item difficulty into account might yield a correction that is less likely to extract "difficulty" factors in a tetrachoric analysis. Given that the tetrachoric approach to directly factor analyzing item data is relatively inexpensive, research into more appropriate adjustments for formula-scored tests seems warranted.

One criticism of the parcel approach is that it does not address dimensionality at the item level because an item may measure a property that is lost in the analysis of parcels. This is an important criticism that warrants further study. One question such research would address is whether dimensionality at the test score level is the same as dimensionality at the item score level. One might argue that a linear factor analysis of parallel parcel data is more likely to provide a better picture of test score dimensionality than dimensionality analysis of item level data, because (1) parallel parcel data is more like test data, (2) item level data is fraught with noise due to the unreliability of a single item, and (3) variation due to differences in item difficulty and examinee item responding strategies are likely to dominate item level analyses. For these reasons, the parallel parcel approach may be a reasonable way of dampening statistical effects to focus on substantive findings.

References

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

Bock, R. D., Gibbons, R. D., and Muraki, E. (1986). Full-information item factor analysis (MRC Report No. 85-1 revised). Chicago: National Opinion Research Center.

Carroll, J. B. (1945). The effect of difficulty and chance success on correlations between items or between tests. Psychometrika, 10, 1-19.

Carroll, J. B. (1983). The difficulty of a test and its factor composition revisited. In S. Messick and H. Wainer (Eds.), Principals of modern psychological measurement: A festschrift for Frederic M. Lord. Hillsdale, NJ: Erlbaum.

Cattell, R. B. (1956). Validation and intensification of the Sixteen Personality Factor Questionnaire. Journal of Clinical Psychology, 12, 205-214.

Cattell, R. B. (1974). Radical parcel factoring versus item factoring in defining personality structure in questionnaires: Theory and experimental checks. Australian Journal of Psychology, 26, 103-119.

Christoffersson, A. (1975). Factor analysis of dichotomized variables. Psychometrika, 40, 5-32.

Cook, L. L., Dorans, N. J., and Eignor, D. R. (in press). An assessment of the dimensionality of three SAT-Verbal test editions. Journal of Educational Statistics.

Cook, L. L., Dorans, N. J., Eignor, D. R., and Petersen, N. S. (1985). An assessment of the relationship between the assumption of unidimensionality and the quality of IRT true-score equating (RR-85-30). Princeton, NJ: Educational Testing Service.

Dorans, N., and Lawrence, I. (1987). The internal construct validity of the SAT. Report submitted to the ETS-College Board Joint Staff Research and Development Committee.

Drasgow, F., and Dorans, N. J. (1982). Robustness of estimates of the squared multiple correlation and squared cross-validity coefficient to violations of multivariate normality. Applied Psychological Measurement, 6, 185-200.

Drasgow, F., and Parsons, C. K. (1983). Application of unidimensional item response theory models to multidimensional data. Applied Psychological Measurement, 7, 189-199.

Hattie, J. A. (1984). An empirical study of various indices for determining unidimensionality. Multivariate Behavorial Research, 19, 49-78.

Hattie, J. A. (1985). Methodology review: Assessing unidimensionality of test and items. Applied Psychological Measurement, 9, 139-164.

Hendrickson, A. E., and White, P. O. (1964). PROMAX: A quick method for rotation to oblique simple structure. British Journal of Mathematical and Statistical Psychology, 17, 65-70.

Joreskog, K. G. (1971). Simultaneous factor analysis in several populations. Psychometrika, 36, 409-426.

Joreskog, K. G., and Sorbom, D. (1984). LISREL VI-Analysis of linear structural relationships by the method of maximum likelihood. Chicago, IL: International Educational Services.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. Psychometrika, 23, 187-200.

McDonald, R. P., and Ahlawat, K. S. (1974). Difficulty factors in binary data. British Journal of Mathematical and Statistical Psychology, 27, 82-99.

Meredith, W. (1964). Notes on factorial invariance. Psychometrika, 29, 177-185.

Mislevy, R. J. (1986). Recent developments in the factor analysis of categorical variables. Journal of Educational Statistics, 11, 3-31.

Muthen, B. (1978). Contributions to factor analysis of dichotomous variables. Psychometika, 43, 551-560.

Muthen, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latest variable indicators. Psychometicka, 49, 115-132.

Schmid, J., and Leiman, J. (1957). The development of hierarchical factor solutions. Psychometrika, 22, 53-61.

Wilson, D., Wood, R. L., and Gibbons, R. (1984). TESTFACT: Test scoring and item factor analysis. Chicago: Scientific Software.

Zwick, R. (1986, December). Personal communication.

Zwick, R. (1987). Assessing the dimensionality of NAEP reading data. Unpublished manuscript.

/gt
il\rrncme87

Table 1

Distributions of Normalized Residuals
for the Various Factor Analytic Solutions:
SAT Mathematical Parcel Data

Normalized Residuals	January 1982 (N=4238)			January 1983 (N=3094)		
	1-Factor[a]	2-Factor[b]	3-Factor[c]	1-Factor[a]	2-Factor[b]	3-Factor[c]
NR≤-3	0	0	0	0	0	
-3<NR≤-2	0	0	0	1	1	
-2<NR< 2	43	45	44	53	53	d
2≤NR< 3	2	0	1	1	1	
NR≥ 3	0	0	0	0	0	

Normalized Residuals	June 1985 (N=3081)			June 1986 (N=3102)		
	1-Factor[a]	2-Factor[b]	3-Factor[c]	1-Factor[a]	2-Factor[b]	3-Factor[c]
NR≤-3	0	0	0	0	0	0
-3<NR≤-2	0	0	0	0	1	0
-2<NR< 2	53	55	55	53	53	55
2≤NR< 3	1	0	0	1	0	0
NR≥ 3	1	0	0	1	1	0

Normalized Residuals	November 1983 (N=4660)			November 1985 (N=3602)		
	1-Factor[a]	2-Factor[b]	3-Factor[c]	1-Factor[a]	2-Factor[b]	3-Factor[c]
NR≤-3	0	0		0	0	0
-3<NR≤-2	0	0		0	0	0
-2<NR< 2	54	55	d	54	54	55
2≤NR< 3	1	0		1	1	0
NR≥ 3	0	[illegible]		0	0	0

Note: See Figure 2 for pictorial representation of models.

[a]The one-factor solution assumes that all parcels load on a single factor.

[b]The two-factor solution assumes that arithmetic and algebra parcels load on one factor, geometry parcels load on a second factor, and miscellaneous parcels load on both factors.

[c]The three-factor solution assumes that algebra parcels load on first factor, arithmetic parcels load on second factor, geometry parcels load on third factor, and miscellaneous parcels load on all three factors.

[d]Not possible to extract a third factor.

Table 2

Root Mean Square Raw Residual by
SAT-Mathematical Edition and Factor Analytic Solution

Admin.	1 factor[a]	2 factor[b]	3 factor[c]
1/82	.014	.011	.009
1/83	.016	.013	-
6/85	.017	.011	.010
6/86	.021	.017	.009
11/83	.011	.007	-
11/85	.012	.009	.007

[a]The one-factor solution assumes that all parcels load on a single factor.

[b]The two-factor solution assumes that arithmetic and algebra parcels load on one factor, geometry parcels load on a second factor, and miscellaneous parcels load on both factors.

[c]The three-factor solution assumes that algebra parcels load on first factor, arithmetic parcels load on second factor, geometry parcels load on third factor, and miscellaneous parcels load on all three factors.

Table 3

Intercorrelations of Content Area Factors in the Three-factor and Two-factor Solution for SAT-Mathematical Editions

Admin.		Factors		
		I	II	III
		ALGEBRA	ARITH.	GEOMETRY
	I	1.0		
1/82	II	.969	1.0	
	III	.951	.938	1.0
	I	1.0		
6/85	II	.996	1.0	
	III	.954	.951	1.0
	I	1.0		
6/86	II	.994	1.0	
	III	.922	.940	1.0
	I	1.0		
11/85	II	.998	1.0	
	III	.953	.946	1.0

Admin.		I ALGB, ARITH.	II GEOMETRY
1/83	I	1.0	
	II	.957	1.0
11/83	I	1.0	
	II	.953	1.0

Table 4

Relative Contributions of One General Factor and Three Content Area Factors to Variance of First Order Parcel Factors for SAT Mathematical Editions

Admin.		First Order Factors		
		Algebra	Arithmetic	Geometry
1/82	general factor	.98	.96	.92
	content area factor	.02	.04	.08
6/85	general factor	1.00	.99	.91
	content area factor	.00	.01	.09
6/86	general factor	.98	1.00	.87
	content area factor	.02	.00	.13
11/85	general factor	1.00	.99	.90
	content area factor	.00	.01	.10

Table 5

Distributions of Normalized Residuals for the Within- and Between-Population One-Factor Solution: Mathematical Equating Sections

Section: il

January 1983 Within Population	January 1983 Between Population	Normalized Residuals	June 1986 Within Population	June 1986 Between Population
0	0	NR≤-3	0	0
0	3	-3<NR≤-2	0	0
28	24	-2<NR< 2	27	25
0	1	2≤NR< 3	1	3
0	0	NR≥ 3	0	0

Section: jp

June 1985 Within Population	June 1985 Between Population	Normalized Residuals	June 1986 Within Population	June 1986 Between Population
0	0	NR≤-3	0	0
0	0	-3<NR≤-2	0	0
27	26	-2<NR< 2	28	28
1	2	2≤NR< 3	0	0
0	0	NR≥ 3	0	0

Section: gx

January 1982 Within Population	January 1982 Between Population	Normalized Residuals	November 1985 Within Population	November 1985 Between Population
0	0	NR≤-3	0	0
0	1	-3<NR≤-2	0	0
28	26	-2<NR< 2	28	28
0	0	2≤NR< 3	0	0
0	1	NR≥ 3	0	0

Section: iv

November 1983 Within Population	November 1983 Between Population	Normalized Residuals	November 1985 Within Population	November 1985 Between Population
0	0	NR≤-3	0	0
0	0	-3<NR≤-2	1	1
28	27	-2<NR< 2	26	26
0	1	2≤NR< 3	1	1
0	0	NR≥ 3	0	0

Table 6

Root Mean Square Residual by Mathematical Equating Section and Factor Analytic Solution

Admin.	Equating Section	1-Factor Within Pop.	1-Factor Between Pop.	RMSR Diff.
1/83	il	.016	.026	-.010
6/86	il	.016	.026	-.010
6/85	jp	.017	.019	-.002
6/86	jp	.016	.018	-.002
11/83	iv	.010	.011	-.001
11/85	iv	.013	.015	-.002
1/82	gx	.007	.018	-.011
11/85	gx	.006	.019	-.013

Table 7

Within-Parcel Full Information Factor Analysis of SAT-Mathematical Edition Administered in June 1986

3PNO Model					2PNO Model (c=0)	
Number of Factors					Number of Factors	
Root >1	Sig. (.05)	Parcel	# Items	KR-20	Root >1	Sig. (.05)
1	1	ALG1	6	.40	1	2[c]
1	2[a]	ALG2	6	.55	1	2[c]
1	2[a]	ALG3	5	.41	1	1
1	1	ARI1	6	.55	1	1
1	1	ARI2	6	.57	1	1
1	2[a]	ARI3	6	.50	1	2[c]
1	2[b]	GEO1	6	.56	2[d]	2[d]
1	1	GEO2	5	.60	1	2[c]
1	1	GEO3	5	.62	1	2[c]
1	2[a]	MIS1	5	.54	1	2[c]
1	1	MIS2	4	.37	1	1

Comments

[a] Second factor is a specific factor. Over 60% of item communality estimates for these four two-factor solutions exceed their respective parcel KR-20 coefficients.

[b] First factor marked by easy items. Second factor marked by hard items.

[c] Second factor is a specific factor. Over 45% of item communality estimates for these five two-factor solutions exceed their respective KR-20 coefficients.

[d] First factor marked by easy items. Second factor marked by hard items.

Table 8

Analyses of Tetrachorics Adjusted for Guessing
SAT-Mathematical

	Right/Wrong/Missing Item Scoring	Right/Wrong Item Scoring
	January 1983	
PVAF	47.7%, 4.5%, 2.7%, 2.4%	53.6%, 3.4%, 2.7%, 2.2%
TETRA-1	.8%	17.3%
R_{12}	.78	.84
	June 1985	
PVAF	43.4%, 5.5%, 2.6%, 2.3%	48.5%, 3.7%, 2.6%, 2.5%
TETRA-1	1.8%	15.6%
R_{12}	.80	.80
	June 1986	
PVAF	43.5%, 5.0%, 2.6%, 2.4%	47.2%, 4.1%, 3.0%, 2.5%
TETRA-1	3.6%	12.3%
R_{12}	.75	.76

Note: PVAF = proportion of variance accounted for by the first four roots
TETRA = 1 = number of tetrachorics set equal to one
R_{12} = correlation between factors

Table 9

Crosstabulation of Factor Loading[a] and Difficulty[b] Classifications for Tetrachoric Solution of the SAT-Mathematical Test

January 1983

Right/Wrong/Missing Item Scoring

	E	M	H
I	17	16	1
I/II	3	5	0
II	0	1	17

Right/Wrong Item Scoring

	E	M	H
I	12	8	0
I/II	3	11	2
II	0	6	18

June 1985

Right/Wrong/Missing Item Scoring

	E	M	H
I	21	5	0
I/II	3	7	2
II	1	8	13

Right/Wrong Item Scoring

	E	M	H
I	15	2	0
I/II	3	10	2
II	3	5	20

June 1986

Right/Wrong/Missing Item Scoring

	E	M	H
I	24	9	0
I/II	2	7	2
II	0	4	12

Right/Wrong Item Scoring

	E	M	H
I	18	14	2
I/II	3	6	3
II	0	1	13

[a]Items classified as I if loading on Factor I >.4 and loading on Factor II <.4, Items classified as II if loading on Factor I <.4 and loading on Factor II >.4, or Items classified as I/II otherwise.

[b]Easy (E): $p \geq .7$; Hard (H): $p \leq .4$; Middle (M): otherwise.

Table 10

Crosstabulation of Factor Loading[a] and Difficulty[b] Classifications for Tetrachoric Solution of the SAT-Mathematical Test

Arithmetic

Right/Wrong/Missing Item Scoring | Right/Wrong Item Scoring

January 1983

Right/Wrong/Missing Item Scoring

	E	M	H
I	7	5	1
I/II	0	1	0
II	0	1	4

Right/Wrong Item Scoring

	E	M	H
I	6	2	0
I/II	1	4	1
II	0	1	4

June 1985

Right/Wrong/Missing Item Scoring

	E	M	H
I	6	0	0
I/II	2	2	0
II	1	0	7

Right/Wrong Item Scoring

	E	M	H
I	4	0	0
I/II	1	2	0
II	3	1	7

June 1986

Right/Wrong/Missing Item Scoring

	E	M	H
I	11	0	0
I/II	0	3	0
II	0	1	3

Right/Wrong Item Scoring

	E	M	H
I	7	3	2
I/II	1	2	0
II	0	0	3

[a]Items classified as I if loading on Factor I $>.4$ and loading on Factor II $<.4$, Items classified as II if loading on Factor I $<.4$ and loading on Factor II $>.4$, or Items classified as I/II otherwise.

[b]Easy (E): $p \geq .7$; Hard (H): $p \leq .4$; Middle (M): otherwise.

Table 11

Crosstabulation of Factor Loading[a] and Difficulty[b] Classifications for Tetrachoric Solution of the SAT-Mathematical Test

Algebra

January 1983

Right/Wrong/Missing Item Scoring

	E	M	H
I	5	5	0
I/II	0	1	0
II	0	0	6

Right/Wrong Item Scoring

	E	M	H
I	3	4	0
I/II	1	0	0
II	0	2	7

June 1985

Right/Wrong/Missing Item Scoring

	E	M	H
I	5	3	0
I/II	0	4	2
II	0	1	2

Right/Wrong Item Scoring

	E	M	H
I	4	1	0
I/II	1	3	2
II	0	2	4

June 1986

Right/Wrong/Missing Item Scoring

	E	M	H
I	8	1	0
I/II	1	2	0
II	0	1	4

Right/Wrong Item Scoring

	E	M	H
I	8	1	0
I/II	1	2	1
II	0	0	4

[a]Items classified as I if loading on Factor I >.4 and loading on Factor II <.4, Items classified as II if loading on Factor I <.4 and loading on Factor II >.4, or Items classified as I/II otherwise.

[b]Easy (E): $p \geq .7$; Hard (H): $p \leq .4$; Middle (M): otherwise.

Table 12

Crosstabulation of Factor Loading[a] and Difficulty[b] Classifications for Tetrachoric Solution of the SAT-Mathematical Test

Geometry

January 1983

Right/Wrong/Missing Item Scoring

	E	M	H
I	2	5	0
I/II	2	2	0
II	0	0	5

Right/Wrong Item Scoring

	E	M	H
I	1	1	0
I/II	1	5	1
II	0	2	5

June 1985

Right/Wrong/Missing Item Scoring

	E	M	H
I	5	2	0
I/II	1	0	0
II	0	6	3

Right/Wrong Item Scoring

	E	M	H
I	5	0	0
I/II	0	3	0
II	0	2	7

June 1986

Right/Wrong/Missing Item Scoring

	E	M	H
I	4	5	0
I/II	0	1	1
II	0	2	3

Right/Wrong Item Scoring

	E	M	H
I	2	8	0
I/II	0	1	0
II	0	1	4

[a]Items classified as I if loading on Factor I >.4 and loading on Factor II <.4, Items classified as II if loading on Factor I <.4 and loading on Factor II >.4, or Items classified as I/II otherwise.

[b]Easy (E): $p \geq .7$; Hard (H): $p \leq .4$; Middle (M): otherwise.

Table 13

Full Information 3PNO Factor Analyses Under Right/Wrong/Missing Item Scoring of the 25-item M1 SAT-Mathematical Edition Administered in January 1983

Number of Factors

PVAF: 50.3%, 5.9%, 3.9%, 3.7%, 3.5%, . . .

TETRA=1: 1.3%

Significant: at least 2

r_{12}: .78

Classifications[a]

	E	M	H	GEO	ALG	ARI	MIS
I	5	7	3	7	3	3	2
I/II	2	0	0	0	1	0	1
II	4	1	3	0	3	5	0

[a]See Table 9 for classification schemes for E, M, H and I, I/II, II.

Note: PVAF = proportion of variance accounted for by the first five roots
TETRA = 1 = number of tetrachorics set equal to one
R_{12} = correlation between factors

Figure 1

SAT Equating Data Collection Design

	New Form	Equating Block to Old Form X	Old Form X	Equating Block to Old Form Y	Old Form Y
New Form Z Sample 1	X	X			
New Form Z Sample 2	X			X	
Old Form X Sample		X	X		
Old Form Y Sample				X	X

Figure 2

Factor Pattern Matrices for SAT-Mathematical Editions

Parcel	Content	3-Factor	2-Factor	1-Factor
1	Algebra	1 0 0	1 0	1
2	Algebra	X 0 0	X 0	X
3	Algebra	X 0 0	X 0	X
4	Arithmetic	0 1 0	X 0	X
5	Arithmetic	0 X 0	X 0	X
6	Arithmetic	0 X 0	X 0	X
7	Geometry	0 0 1	0 1	X
8	Geometry	0 0 X	0 X	X
9	Geometry	0 0 X	0 X	X
10	Miscellaneous	X X X	X X	X
11	Miscellaneous	X X X	X X	X

Factor intercorrelations matrices for SAT-Mathematical Editions

1	1	1
X 1	X 1	
X X 1		

1 - loading fixed to equal one
X - parameter to be estimated
0 - loading fixed to equal zero

Note: For one edition (January, 1982), there was only one parcel for miscellaneous items.