DOCUMENT RESUME

ED 462 423 TM 033 687

| | |
|---|---|
| AUTHOR | Meara, Kevin; Sireci, Stephen G. |
| TITLE | Appraising the Dimensionality of the Medical College Admission Test. MCAT Monograph. |
| SPONS AGENCY | Association of American Medical Colleges, Washington, DC. |
| REPORT NO | MCAT-2 |
| PUB DATE | 2000-08-00 |
| NOTE | 66p. |
| AVAILABLE FROM | Association of American Medical Colleges, Section for the Medical College Admission Test, 2450 N Street, NW, Washington, DC 20037. Tel: 202-828-0400; Fax: 202-828-1125; Web site: http://www.aamc.org/mcat. |
| PUB TYPE | Reports - Evaluative (142) |
| EDRS PRICE | MF01/PC03 Plus Postage. |
| DESCRIPTORS | *College Applicants; College Entrance Examinations; *Factor Structure; Higher Education; Medical Education; Reliability; Sciences; Thinking Skills; Verbal Tests |
| IDENTIFIERS | *Dimensionality (Tests); Item Parcels; *Medical College Admission Test |

ABSTRACT

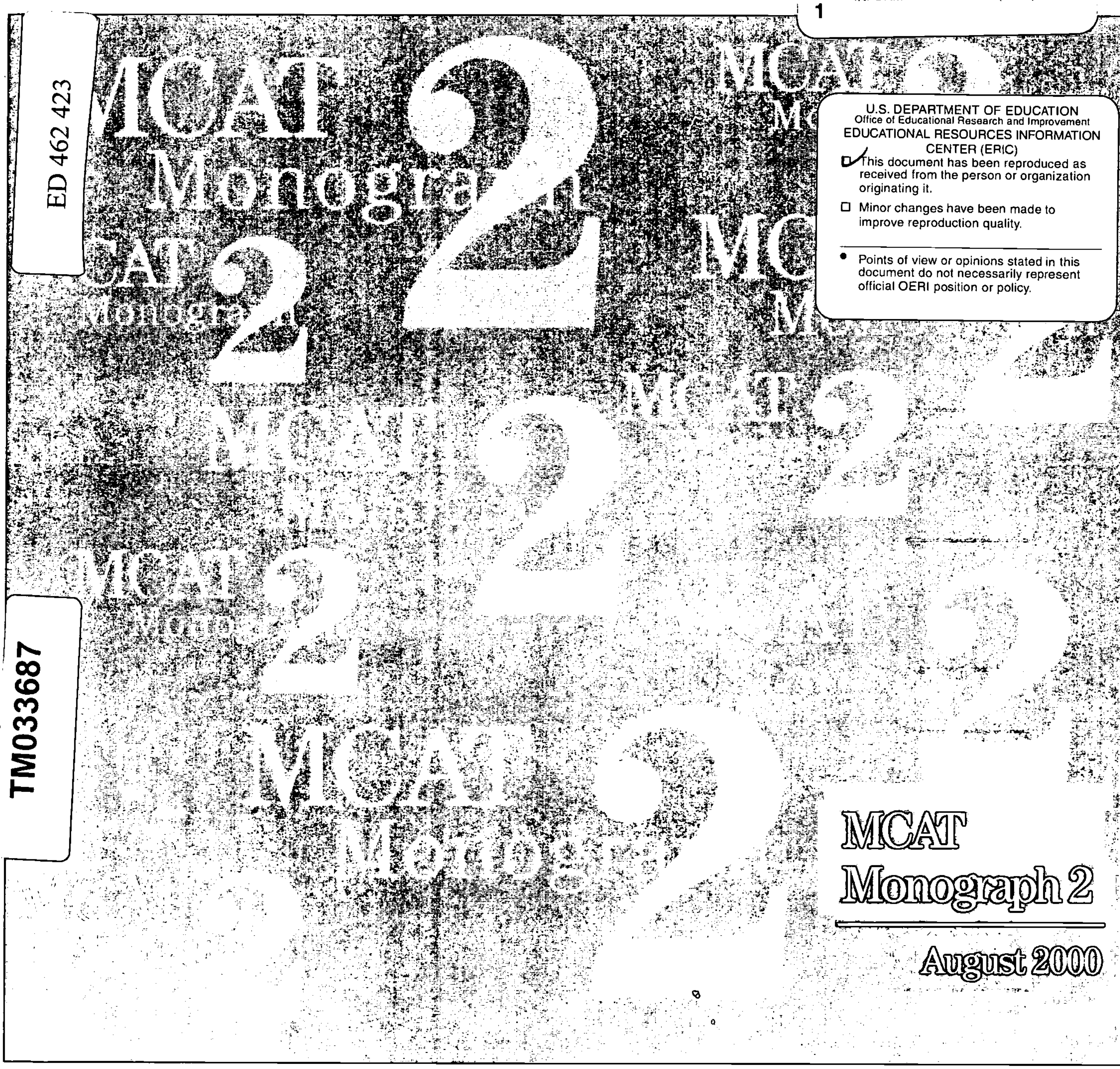
To provide a better understanding of the structure of the Medical College Admission Test (MCAT) and to determine if there are structural differences across selected groups of MCAT examinees, several dimensionality analyses were conducted on data from recent administrations of the MCAT. The first set of analyses focused on the global structure of the MCAT, and the second set appraised the consistency of the structure of data across groups of testtakers that differed with respect to sex, repeater/nonrepeater status, orientation to the English language, and race/ethnicity. Data from two forms of the MCAT were used. Forms 15A and 15B were administered in 1994 to 16,520 examinees, and Forms 23A and 23B were administered in 1996 to 12,625 examinees. Results suggest that appraisals of the MCAT structure should be conducted at the parcel level rather than at the item level. Parcel-level results suggest that a dominant factor underlies the MCAT. This is probably a "general intelligence" factor. The results also suggest additional factors that represent the principal disciplines measured on the MCAT. After the general factor, the next structural layer of the MCAT separates test material measuring science from test material measuring verbal reasoning and writing skills. The next structural level depicts three factors: science, verbal reasoning, and writing skills. These three factors were supported by all analyses. Results also support the distinction between the science disciplines, and, in general, analyses support the current content structure of the MCAT reported in the test blueprint. From a statistical perspective, results suggest it might be possible to scale the biological and physical sciences along a single continuum. With respect to the consistency of the MCAT structure across selected groups of testtakers, results supported the hypothesis of structural invariance across groups. In general, multidimensional scaling analyses indicated that all dimensions were relevant for accounting for the variation in the data for each group. Some exceptions are discussed. An appendix contains depictions of the confirmatory factor analysis models and parceling schemes. (Contains 13 figures, 30 tables, and 20 references.) (SLD)

ASSOCIATION OF AMERICAN MEDICAL COLLEGES



# Appraising the Dimensionality of the Medical College Admission Test





# Appraising the Dimensionality of the Medical College Admissions Test

Kevin Meara and Stephen G. Sireci

University of Massachusetts—Amherst

---

Report submitted to the Association of American Medical Colleges (AAMC), MCAT Division, September, 1998. 

## Introduction

As part of its 1998 Summer Graduate Student Research Project, a series of dimensionality studies was commissioned by the Association of American Medical Colleges (AAMC) to Kevin Meara of the University of Massachusetts at Amherst. This report presents the results of these studies. The primary purposes of this research were to better understand the structure of the Medical College Admission Test (MCAT) and to determine if structural differences exist across selected groups of MCAT examinees.

To accomplish these goals, several dimensionality analyses were conducted on data from recent administrations of the MCAT. The first set of analyses focused on appraising the global structure of the MCAT. Global in this case means that the total population of test takers was the studied sample. The second set of analyses appraised the consistency of the structure of these data across groups of test takers that differed with respect to sex, repeater/non-repeater status, orientation to the English language, and race/ethnicity.

The term "dimensionality" can be a confusing one. In fact, Brennan (1998) recently exclaimed: "the terms *unidimensional* and *multidimensional* have so many conflicting connotations that their unqualified use is little more than a Tower of Babel" (p. 6). In this study, we use the word dimensionality to refer to the structural aspects of the MCAT that correspond to systematic sources of variation in the responses of examinees to MCAT items. These aspects are often called "dimensions" or "factors" in the psychometric literature. The intended structure of the MCAT stipulates four distinct dimensions, which are characterized as test sections in the MCAT battery: Verbal Reasoning, Biological Sciences, Physical Sciences, and the Writing Sample. The analyses conducted in this study sought to uncover this intended structure and determine if it is consistent across selected groups of the examinee population.

## Method

The analyses conducted in this study were comprehensive. At the most general level, the analyses can be classified into one of two groups: global or group analyses. The global analyses investigated the dimensionality of MCAT data for the entire MCAT examinee population (from two recent administrations of the MCAT). These analyses were conducted on both item-level data and parcel-level (groups of items) data. The group analyses were conducted only using parcel-level data. In this section, first we will describe characteristics of the data, and then we will describe the global and group analyses.

### Data

Data from two forms of the MCAT were used in this study. Forms 15A and 15B were administered in 1994 to 16,520 examinees (8,494 and 8,026 were administered forms 15A and 15B respectively). Forms 23A and 23B were administered in 1996 to 12,625 examinees (8,147 and 4,478 were administered forms 23A and 23B respectively). The only difference between test forms A and B, which are identical in content, is item order. All global analyses were done using data from each of the four forms, whereas group analyses were performed by combining forms A and B for each test administration.

For each examinee, we had responses to 181 dichotomously scored items and two essay prompts. Raw essay scores ranged from 2 to 12 points. The 181 dichotomously scored items were aggregated from the following three test sections: Verbal Reasoning (55 items), Physical Science (63 items), and Biological Science (63 items). The science items were further broken into four *disciplines*: Physics, General Chemistry, Biology, and Organic Chemistry. In addition to *discipline*, items were classified by these characteristics: *content* categories (48), *cognitive* classifications (11), and *passage type* (6). These characteristics were used mainly to aid factor interpretation of the item-level principal component analyses (PCA) results. For analyses conducted using parcel-level data, parcels were created by bundling groups of items based on content and difficulty.

Both confirmatory factor analysis (CFA) and weighted multidimensional scaling (WMDS) were used to analyze the factor structure of the groups. The variables of interest included: 1) sex (analyzed using CFA only), 2) test taking status (first-timers/repeaters), 3) English as a second language (ESL), and 4) race. Different sampling methods were used for each procedure. For the CFA analyses, samples consisted of all available subjects in each group. For example, there were 8,820 first-timers and 7,661 repeaters who tested with form 15. All examinees were used. Sample sizes were problematic for some racial groups. First, scores for Native Americans (form 15: n=123, form 23: n=85) were too few to yield stable estimates; therefore, results for them were not reported. Second, sample sizes for Other Hispanic test takers (form 15: n=373, form 23: n=287) were also small, although the results for them are reported. Sample sizes for the group CFA analyses are shown in Table 1a.

**Data: Sample Sizes**

Table 1a. Data Used for This Study -- Group Sample Sizes

| | | Number of Observations | |
|---|---|---|---|
| **Variable** | **Group** | **Form 15** | **Form 23** |
| **Each Exam** | All | 16,520 | 12,625 |
| **Each Form** | Form A | 8,494 | 8,147 |
| | Form B | 8,026 | 4,478 |
| **Sex** | Female | 7,651 | 5,952 |
| | Male | 8,714 | 6,662 |
| **Repeater** | First-time test taker | 8,820 | 7,297 |
| | Repeater | 7,661 | 5,320 |
| **ESL** | English is Native Language | 13,337 | 10,245 |
| | Learned English between 6-10 | 1,578 | 1,163 |
| | Learned English after age 10 | 1,477 | 1,088 |
| **Race** | Asian American | 3,844 | 2,853 |
| | African American | 1,393 | 1,085 |
| | Other Hispanic | 373 | *287* |
| | Native American | *123* | *85* |
| | Mexican American or PR | 967 | 909 |
| | Caucasian | 8,880 | 6,711 |

For the WMDS analyses, data in each group were split into two samples based on sex, and separate inter-parcel correlation matrices were derived for each one. Instead of having two groups for first-timers/repeaters, there were four. For form 15, there were 4,649 male and 4,188 female first-timers, and 4,084 male and 3,483 female repeaters. Because the sample sizes were small for Other Hispanic examinees, they were not divided into two groups. Sample sizes for the group WMDS analyses are shown in Table 1b.

Table 1b, Group WMDS Sample Sizes

| | | Number of Observations | | | |
|---|---|---|---|---|---|
| | | Form 15 | | Form 23 | |
| **Variable** | **Group** | **Male** | **Female** | **Male** | **Female** |
| **Repeater** | First-time test taker | 4,649 | 4,188 | 3,826 | 3,472 |
| | Repeater | 4,084 | 3,483 | 2,838 | 2,486 |
| **ESL** | English is Native Language | 6,955 | 6,289 | 5,317 | 4,928 |
| | Learned English between 6-10 | 832 | 738 | 639 | 525 |
| | Learned English after age 10 | 876 | 587 | 631 | 460 |
| **Race** | Asian American | 2,138 | 1,712 | 1,586 | 1,267 |
| | African American | 526 | 868 | 387 | 698 |
| | Other Hispanic | *203* | *171* | *155* | *132* |
| | Native American | *65* | *58* | *35* | *51* |
| | Mexican American or PR | 467 | 505 | 445 | 469 |
| | Caucasian | 4,910 | 3,988 | 3,657 | 3,056 |

## Analyses

### Global Item-level Analyses

The first step in this dimensionality study was to perform a series of item-level analyses on the student response data using principal component analyses (PCA). Generally, work done with PCA is exploratory in nature. The purpose is to explore the data, to discover and detect characteristic features and interesting relationships, without imposing any definite model on the data (Joreskog & Sorbom, 1989). In this study, PCA was used for two reasons. First, it served as a preliminary check on dimensionality of the four forms (15A, 15B, 23A and 23B). Second, it was done to help make decisions about parceling items for the next series of analyses.

Only dichotomously scored items were used for these analyses. PCAs were conducted on each of the four complete forms (4 analyses), and on each of the three test sections (Verbal Reasoning, Biological Science, and Physical Science) for each form (12 analyses). The results yielded both an unrotated factor solution and an orthogonally rotated solution. Due to the small size of the rotated factor loadings for the complete forms, efforts were made to interpret the factors of the unrotated solutions only. Interpretations were carried out both visually and statistically. Visual analysis involved identifying patterns among the characteristics and the factor loadings. Statistical analysis involved calculating correlations among known item characteristics and the factor loadings from the PCA solutions. This was done by dummy coding characteristics such as *discipline* (which consists of 5 classifications) as separate, dichotomous variables. Each dummy variable was then correlated with the factor loadings. The characteristics correlated with the loadings included: a) *corrected item difficulty*, b) *discipline*, c) *passage type*, d) *cognitive classification*, e) *content classification*, and f) *biserial correlation.*

Global Parcel-level Analyses

The next step in this study was global parcel-level analysis. Performing factor analyses on dichotomous items has been criticized for several reasons (e.g., Cattell 1956; Cattell & Burdsal, 1975; Dorans & Lawrence, 1987, 1991; Green, 1983). Paramount among these reasons is the finding that spurious factors may emerge due to "noise" in the item-level data (i.e., due to the unreliability of a single item). In addition, when a linear model (such as the PCA model) is used with dichotomous (non-linear) data, the model has a tendency to overestimate dimensionality. For these reasons, dichotomously scored items are often bundled together into parcels to yield more stable representations of factor structure.

Two parceling strategies were used in this study. The first strategy grouped items into parcels based on content at the *discipline* level. Items were parceled largely based on passages (or item sets), which ensured that items of similar *discipline* were grouped together. This strategy was supported after considering the results from the item-level PCA. Using this method, 31 and 32 parcels were created for test forms 15 and 23 respectively. The second strategy considered the difficulty level of the items within discipline area. This scheme ignored sets of items and parceled largely by *difficulty* within *discipline*. Items of similar *discipline* categories were grouped together, and the parcels were balanced for difficulty. Using this method, 35 and 34 parcels were created. Appendix A shows how the items were parceled using each strategy for test form 15.

PCAs were performed on all four forms, for each parceling method. For form 23, the results were similar regardless of how the items were bundled. For form 15, on the other hand, the two bundling strategies yielded slightly different results. The factors bundled using the second scheme (i.e. difficulty-within-discipline) were more easily interpreted; thus, only the results for items bundled using the second scheme are presented. There are two advantages of focusing on the difficulty-within-discipline parceling scheme. First, when items are parceled in this way, item dependence is reduced because item sets are broken up. Second, the effect of item order is reduced because the items are not parceled according to test order. Additional analyses using multidimensional scaling (MDS) and confirmatory factor analysis (CFA) were performed only on the items bundled using the second parceling scheme.

MDS is a technique similar to factor analysis that is used to study and describe the structure of multivariate data. Unlike factor analysis, MDS does not specify a linear model and is considered to be more appropriate for evaluating test structure (Davison, 1985). A strength of MDS is that it's a "spatial distance model," which enables one to plot the stimuli (e.g., items or item parcels) and visually inspect the relationships that emerge. In this study, dissimilarities among the parcels were calculated from the subdiagonal polychoric correlation matrix for each form. Like factor analysis, MDS requires a user to obtain solutions in several dimensions and then choose among them. The fit statistics used to evaluate the results are STRESS and $R^2$. Although there are no universally accepted "rules of thumb" regarding the best fitting solution, many consider a solution to display adequate fit to the data if STRESS is less than or equal to .10, and $R^2$ is greater than or equal to .90. However, simulations conducted by Kruskal and Wish (1978) suggest that if a STRESS value of .15 or less is obtained when fitting an unidimensional model to the data, the data should be considered unidimensional. Like factor analysis, the relative improvement in fit from one solution to the next is also considered helpful in determining dimensionality. However, although data-model fit is important in determining the most appropriate MDS model, the interpretability of the solution is typically the most important factor (Davison & Sireci, in press). High-dimensional solutions that cannot be interpreted are typically discarded, even if they exhibit better fit than lower-dimensional solutions.

CFA was used to test the fit of parcel-level data to seven different models, ranging from one- to eight-factors. Following is a list of the specifications of each multidimensional model. Diagrams of each model can be found in Apendix A. The one-factor model is self-explanatory. The two-factor model specified constructs measuring: a) all VR and essay parcels on factor 1, and b) all science items on factor 2. The three-factor model specified constructs measuring: a) all VR parcels on factor 1, b) all science parcels on 2, and c) essay score items on factor 3. Two four factor models were fit to the data. The first model (model 4a) specified constructs measuring: a) all VR parcels on factor 1, b) all physics and general chemistry parcels on 2, c) all biology and organic chemistry parcels on 3, and d) essay items on factor 4. This model mirrored the discipline specifications of the MCAT. The second four-factor model (model 4b) specified constructs measuring: a) all VR parcels on factor 1, b) all non-biology science parcels on 2, c) biology parcels on 3, and d) essay items on factor 4. This structure was supported by the exploratory PCA and MDS analyses. The six-factor model specified constructs measuring: a) all VR parcels on factor 1, b) physics on 2, c) general chemistry on 3, d) biology on 4, e) organic chemistry on 5, and f) essay scores on factor 6. Finally, the eight-factor model specified constructs measuring: a) humanities parcels on factor 1, b) natural sciences on 2, c) social sciences on 3, d) physics on 4, e) general chemistry on 5, f) biology on 6, g) organic chemistry on 7, and h) essay scores on factor 8. LISREL 7.2 (Joreskog & Sorbom, 1989) was used to carry out the CFA analyses. The goodness of fit index (GFI), adjusted goodness of fit index (AGFI), root mean square residual (RMSR), and change in chi-square were used to evaluate data-model fit. Generally, a model was considered to fit the data when the GFI and AGFI were greater than or equal to .90, and the RMSR was less than .10.

### Group Parcel-level Analyses

The final step in this study was to evaluate the consistency of the MCAT structure across diverse groups of test takers, using weighted multidimensional scaling (WMDS) and confirmatory factor analysis. Both procedures allow for multiple groups within an analysis. The four groups of interest were as follows: 1) sex (females and males), 2) repeaters (first-timers and repeaters), 3) ESL (English as a first language, English learned between ages 6-10, and English learned after age 10), and 4) race (Asian Americans, African Americans, Other Hispanics, Native Americans, Mexican Americans or Puerto Ricans, and Caucasians). WMDS was used on all of the above groups except sex.

Weighted MDS models, also called "individual differences" models are appropriate for evaluating test structure across groups because a common representation of test structure (called the stimulus space) is derived simultaneously for all groups. Differences in dimensional structure among the groups are reported using "subject weights." These weights are used to adjust the stimulus space so it can be "stretched" or "shrunk" to best fit the data for one or more groups. For example, the INDSCAL model proposed by Carroll and Chang (1970) uses a weighted Euclidean distance formula to scale stimuli:

$$d_{ijk} = \sqrt{\sum_{a=1}^{r} w_{ka} (x_{ia} - x_{ja})^2}$$

where: $d_{ijk}$= the Euclidean distance between stimuli (e.g., test items) $i$ and $j$ for group $k$; $w_{ka}$ is the weight for group $k$ on dimension $a$; $x_{ia}$= the coordinate of stimulus $i$ on dimension $a$; and $r$= the dimensionality of the model. A common stimulus space is derived for the stimuli. The "personal" distances for each group are related to the common stimulus space by the equation:

$$x_{kia} = \sqrt{w_{ka}}\, x_{ia}$$

where $x_{kia}$ represent the coordinate for stimulus $i$ on dimension $a$ in the personal space for group $k$; $w_{ka}$ represents the weight of group $k$ on dimension $a$; and $x_{ia}$ represents the coordinate of stimulus $i$ on dimension $a$ in the common stimulus space.

Although weighted MDS models can evaluate test structure simultaneously across all groups, most MDS models do not provide statistical tests of structural equivalence (cf. Ramsay, 1981). Rather, descriptive fit indices are used to evaluate data-model fit. The STRESS index represents the square root of the normalized residual variance of the monotonic regression of the MDS distances on the transformed proximities. Thus, lower values of STRESS indicate better fit. The $R^2$ index reflects proportion of variance of the transformed proximities accounted for by the MDS distances. Thus, higher values of $R^2$ indicate better fit. Recent applications of weighted MDS have illustrated its advantages for evaluating structural equivalence across cultural groups (Day & Rounds, 1998; Day, Rounds, & Swaney, 1998) and across different language versions of a test (Sireci, Fitzgerald, & Xing, 1998).

For the weighted MDS analyses in this study, groups of test takers were split first by sex and then by one of the other grouping variables of interest. For example, the analysis of "repeaters" versus "first-time" test takers involved creating four polychoric matrices for the following four groups: male repeaters, female repeaters, male first-time test takers, and female first-time test takers. The WMDS analysis for race involved deriving ten matrices. The matrices were derived for males and females from each of the following groups: African Americans, Asian Americans, Caucasians, and Mexican Americans. One matrix was derived from the responses of Spanish and South American test takers, and another was derived from the responses of Native American test takers. There were too few test takers in these last two groups to derive separate matrices for each sex. All matrices were fit using the weighted MDS model INDSCAL (Carroll & Chang, 1970).

Confirmatory factor analysis is becoming an increasingly popular technique for evaluating structural equivalence across different groups (e.g., Reise & Widaman, & Pugh, 1993; Robie & Ryan, 1996; Sireci, Fitzgerald, & Xing, 1998). CFA is attractive in this situation because it can handle multiple groups simultaneously, statistical tests of model fit are available, and descriptive indices of model fit are provided. In multi-group CFA analyses, the hypothesized structure of an assessment is incorporated into a structural equation model, and the structure is constrained to be equal across all groups. A typical hypothesis tested using CFA is whether the factor loading matrix is equivalent across all groups. The structure of the factor loadings is usually an "independent clusters structure" (MacDonald, 1985), which specifies that: 1) each measured variable has a nonzero loading on only the factor it was designated to measure, 2) correlations among the factors (i.e., lower diagonal of the phi matrix) are freely estimated, and 3) the errors associated with the factor loadings (i.e., theta delta matrix) are uncorrelated (Marsh, 1994). Sireci, et al. (1998) used the term "invariant independent clusters structure" to refer to a model that constrains this structure to be equal across two or more groups. For the CFA analyses, matrices of polychoric correlations were analyzed using LISREL 7.2 and LISREL 8.0 (Joreskog & Sorbom, 1989; 1996). One-, two-, three-, $four_a$-, and six-factor models were fit to the data. These models are identical to those fit in the global parcel-level PCA analyses. The GFI, RMSR, and Chi-square statistic were used to evaluate data-model fit (Marsh, Balla, & MacDonald, 1988). Using simulated data, Sireci et al. (1998) found that the RMSR was the best index for detecting departure from an invariant independent clusters structure.

## Results

### Global Analyses

#### Global Item-level PCA Results for Complete Forms

As expected, the analyses done on all items composing a test form all showed multidimensionality. For all forms, the eigenvalues and the percentage of variance accounted for by each factor are presented in Tables 2a - 3b. For form 15A, 47 factors had eigenvalues > 1.0. The first principal component (eigenvalue=11.4) accounted for only 6.3% of the variance; and, the second factor (eigenvalue=7.8) accounted for 4.3% of the variance. Similarly, for form 15B, 47 factors had eigenvalues > 1.0. The first principal component (eigenvalue= 12.6) accounted for only 6.9% of the variance; and, the second factor (eigenvalue=7.8) accounted for 4.3% of the variance. Although slightly more variance (8.9%) was accounted for by the first factor (eigenvalue=16.1) for form 23A, and less variance (2.2%) by the second factor (eigenvalue= 4.0), PCA extracted 54 factors. Finally, for form 23B, 58 factors were extracted. The first principal component (eigenvalue=16.6) accounted for 9.1% of the variance; and, the second factor (eigenvalue= 4.0) accounted for 2.2% of the variance.

**Summary of Item-level PCA Results for:
Complete Test Forms (i.e. all disciplines considered)**

Table 2a. Complete Form, 1994, 15A

| Factor | Eigenvalue | % of Var | Cum. % |
|---|---|---|---|
| 1 | 11.45 | 6.3 | 6.3 |
| 2 | 7.81 | 4.3 | 10.6 |
| 3 | 4.47 | 2.5 | 13.1 |
| 4 | 3.43 | 1.9 | 15.0 |
| 5 | 3.09 | 1.7 | 16.7 |
| 6 | 2.93 | 1.6 | 18.3 |
| 7 | 2.27 | 1.3 | 19.6 |
| 8 | 2.14 | 1.2 | 20.8 |
| 9 | 1.93 | 1.1 | 21.8 |
| 10 | 1.86 | 1.0 | 22.9 |
| 11 | 1.82 | 1.0 | 23.9 |
| 12 | 1.79 | 1.0 | 24.9 |

Table 2b. Complete Form, 1994, 15B

| Factor | Eigenvalue | % of Var | Cum. % |
|---|---|---|---|
| 1 | 12.56 | 6.9 | 6.9 |
| 2 | 7.81 | 4.3 | 11.3 |
| 3 | 4.51 | 2.5 | 13.7 |
| 4 | 3.04 | 1.7 | 15.4 |
| 5 | 2.81 | 1.6 | 17.0 |
| 6 | 2.60 | 1.4 | 18.4 |
| 7 | 2.29 | 1.3 | 19.7 |
| 8 | 2.06 | 1.1 | 20.8 |
| 9 | 2.02 | 1.1 | 21.9 |
| 10 | 1.91 | 1.1 | 23.0 |
| 11 | 1.84 | 1.0 | 24.0 |
| 12 | 1.74 | 1.0 | 25.0 |

Table 3a. Complete Form, 1996, 23A

| Factor | Eigenvalue | % of Var | Cum. % |
|---|---|---|---|
| 1 | 16.09 | 8.9 | 8.9 |
| 2 | 4.03 | 2.2 | 11.1 |
| 3 | 2.24 | 1.2 | 12.4 |
| 4 | 1.83 | 1.0 | 13.4 |
| 5 | 1.68 | .9 | 14.3 |
| 6 | 1.51 | .8 | 15.1 |
| 7 | 1.44 | .8 | 15.9 |
| 8 | 1.40 | .8 | 16.7 |
| 9 | 1.33 | .7 | 17.4 |
| 10 | 1.31 | .7 | 18.1 |
| 11 | 1.28 | .7 | 18.9 |
| 12 | 1.27 | .7 | 19.6 |

Table 3b. Complete Form, 1996, 23B

| Factor | Eigenvalue | % of Var | Cum. % |
|---|---|---|---|
| 1 | 16.55 | 9.1 | 9.1 |
| 2 | 3.99 | 2.2 | 11.3 |
| 3 | 2.31 | 1.3 | 12.6 |
| 4 | 1.89 | 1.0 | 13.7 |
| 5 | 1.67 | .9 | 14.6 |
| 6 | 1.55 | .9 | 15.4 |
| 7 | 1.52 | .8 | 16.3 |
| 8 | 1.49 | .8 | 17.1 |
| 9 | 1.42 | .8 | 17.9 |
| 10 | 1.37 | .8 | 18.6 |
| 11 | 1.36 | .7 | 19.4 |
| 12 | 1.34 | .7 | 20.1 |

To determine if there were similar structures across forms and tests, efforts were made to interpret the first 10 factors for each form using visual inspection and correlation analysis. Only the unrotated solutions were interpreted. Across all four forms the first factor correlated highly with the *biserials* (.60 to .89), suggesting that this factor is a kind of general test factor. The second factor across all four forms related to verbal reasoning (correlations ranged from .56 to .72 across the four forms), but also had significant negative loadings related to organic chemistry (-.21 to -.38). In addition, the second factor (for 23A and 23B), had significant negative correlations with general chemistry (-.39 to -.40). More complete information about the factor interpretations of the first 10 factors for all four forms, can be found in Table 10.

**Summary of Item-level Analysis: Factor Interpretation**

Table 10, Factor interpretation for the four complete forms.

| Factor | Category | 15A | 15B | 23A | 23B |
|---|---|---|---|---|---|
| 1 | Biserial: | Biserial (.61) | Biserial (.62) | Biserial (.89) | Biserial (.88) |
| | Discipline: | BLG. (.31) | BLG. (.41) | BLG (.20) | BLG (.18) |
| | Discipline: | | | VR (-.33) | VR (-.30) |
| 2 | Discipline: | VR (.62) | VR (.56) | VR (.71) | VR (.72) |
| | Discipline: | ORG (-.21) | ORG (-.22) | ORG (-.38) | ORG (-.38) |
| | Discipline: | | | GCH (-.39) | GCH (-.40) |
| 3 | Difficulty: | CD (.37) | | CD (-.70) | CD (-.71) |
| | Discipline: | VR (.47) | GCH (.55) | PHY (.23) | PHY (.18) |
| | Discipline: | GCH (-.33) | PHY (.27) | ORG (.20) | ORG (.23) |
| | Discipline: | PHY (-.56) | VR (-.46) | BLG (-.20) | BLG (-.20) |
| 4 | Difficulty: | | CD (.30) | CD (-.33) | |
| | Discipline: | VR (.49) | VR (.52) | VR (.41) | VR (.36) |
| | Discipline: | BLG (-.20) | BLG (-.21) | ORG (.38) | ORG (.31) |
| | Discipline: | ORG (-.25) | ORG (-.32) | BLG (-.42) | BLG (-.60) |
| | Discipline: | | | PHY (-.33) | PHY (-.19) |
| 5 | Discipline: | VR (.16) | VR (.20) | ORG (.22) | VR (.33) |
| 6 | Discipline: | GCH (.31) | VR (.23) | VR (.25) | GCH (.26) |
| | Discipline: | PHY (.20) | PHY (-.20) | PHY (.27) | |
| | Discipline: | BLG (-.46) | BLG (-.34) | BLG (-.34) | BLG (-.25) |
| | Passage Type: | D (-.39) | | | |
| 7 | Difficulty: | | | | CD (-.33) |
| | Discipline: | GCH (.47) | PHY (.48) | BLG (.50) | BLG (.44) |
| | Discipline: | PHY (-.33) | BLG (-.48) | PHY (-.35) | PHY (-.38) |
| | Discipline: | | | GCH (-.22) | GCH (-.17) |
| 8 | Discipline: | GCH (.23) | PHY (.21) | GCH (.48) | ORG (.40) |
| | Discipline: | | BLG (-.28) | PHY (-.29) | PHY (-.20) |
| 9 | Discipline: | ORG (.20) | | ORG (.19) | |
| | Passage Type: | | D (.36) | | D (.26) |
| 10 | Discipline: | BLG (.17) | GCH (-.15) | GCH (.20) | |
| | Discipline: | PHY (-.21) | | ORG (.18) | |
| | Passage Type: | | | | D (.18) |

In general, the results of these analyses indicated that the factors loaded consistently across the scrambled version of each test form (A & B). Loadings across test forms (15 and 23), were similar but not identical. One distinction is that a larger percentage of items loaded on the first factor for form 23. If we look at the results of the correlations across all four forms we see some interesting things. First, correlations between loadings and *discipline* classification were generally large. In other words, the five disciplines (verbal reasoning, biology, organic chemistry, physics, and general chemistry) tended to load on separate factors. Next, correlations between loadings and *content* classification tended to be small. This is probably due to the small number of items in each *content* category. The exception is verbal reasoning (VR) which has only three *content* categories for 55 items. As a result, VR *content* categories tended to load on the same factors that loaded significantly on verbal reasoning *discipline* factors. Finally, correlations between loadings and *cognitive* classifications also tended to be small.

Global Item-level PCA Results for Test Sections

In general, the item-level PCA results at the test section level suggest that all test sections are multidimensional. The eigenvalues and the percentage of variance accounted for by each factor for all test sections are presented in Tables 4a through 9b. For the verbal reasoning section, form 15A, 12 factors had eigenvalues > 1.0. The first principal component (eigenvalue= 8.4) accounted for 15.2% of the variance; and, the second factor (eigenvalue= 4.3) accounted for 7.8% of the variance. Similarly, for 15B a 10 factor solution was extracted. The first principal component (eigenvalue= 8.8) accounted for 15.9% of the variance; and, the second factor (eigenvalue= 4.2) accounted for 7.6% of the variance. Analysis of the factor loadings revealed that items tended to load as sets based on membership to a particular passage. For example, items 42-47, which are based on a humanities passage, loaded on factor 2; whereas, items 48-55, which are based on a natural science passage, loaded on factor 1. The loadings then, tended to cluster items according to *discipline/content*.

For verbal reasoning, form 23A, 15 factors had eigenvalues > 1.0. The first principal component (eigenvalue= 5.9) accounted for 10.8% of the variance; and, the second factor (eigenvalue= 1.7) accounted for 3.1% of the variance. Similarly, for 23B, PCA extracted 16 factors. The first principal component (eigenvalue= 6.2) accounted for 11.3% of the variance; and, the second factor (eigenvalue= 1.6) accounted for 2.9% of the variance. In this case, analysis of the factor loadings for the unrotated solution revealed that most items loaded on the first factor. This is consistent with the high correlations between the biserials and the first factors (.97 and .88). In addition, the items did not load as sets on other factors, although they did load distinctly by *content* category according to the correlational analysis.

Table 4a. Verbal Reasoning, 1994, 15A

| Factor | Eigenvalue | % of Var | Cum. % |
|---|---|---|---|
| 1 | 8.38 | 15.2 | 15.2 |
| 2 | 4.27 | 7.8 | 23.0 |
| 3 | 3.05 | 5.5 | 28.5 |
| 4 | 2.08 | 3.8 | 32.3 |
| 5 | 1.89 | 3.4 | 35.8 |
| 6 | 1.69 | 3.1 | 38.8 |
| 7 | 1.58 | 2.9 | 41.7 |
| 8 | 1.52 | 2.8 | 44.5 |
| 9 | 1.17 | 2.1 | 46.6 |
| 10 | 1.05 | 1.9 | 48.5 |

Table 4b. Verbal Reasoning, 1994, 15B

| Factor | Eigenvalue | % of Var | Cum. % |
|---|---|---|---|
| 1 | 8.76 | 15.9 | 15.9 |
| 2 | 4.20 | 7.6 | 23.6 |
| 3 | 2.60 | 4.7 | 28.3 |
| 4 | 2.28 | 4.1 | 32.4 |
| 5 | 2.06 | 3.8 | 36.2 |
| 6 | 1.77 | 3.2 | 39.4 |
| 7 | 1.52 | 2.8 | 42.2 |
| 8 | 1.42 | 2.6 | 44.7 |
| 9 | 1.12 | 2.0 | 46.8 |
| 10 | 1.05 | 1.9 | 48.7 |

Table 7a. Verbal Reasoning, 1996, 23A

| Factor | Eigenvalue | % of Var | Cum. % |
|---|---|---|---|
| 1 | 5.94 | 10.8 | 10.8 |
| 2 | 1.71 | 3.1 | 13.9 |
| 3 | 1.35 | 2.5 | 16.3 |
| 4 | 1.30 | 2.4 | 18.7 |
| 5 | 1.19 | 2.2 | 20.9 |
| 6 | 1.16 | 2.1 | 23.0 |
| 7 | 1.14 | 2.1 | 25.1 |
| 8 | 1.09 | 2.0 | 27.0 |
| 9 | 1.07 | 2.0 | 29.0 |
| 10 | 1.06 | 1.9 | 30.9 |

Table 7b. Verbal Reasoning, 1996, 23B

| Factor | Eigenvalue | % of Var | Cum. % |
|---|---|---|---|
| 1 | 6.22 | 11.3 | 11.3 |
| 2 | 1.58 | 2.9 | 14.2 |
| 3 | 1.46 | 2.7 | 16.9 |
| 4 | 1.26 | 2.3 | 19.1 |
| 5 | 1.21 | 2.2 | 21.3 |
| 6 | 1.17 | 2.1 | 23.5 |
| 7 | 1.13 | 2.0 | 25.5 |
| 8 | 1.10 | 2.0 | 27.5 |
| 9 | 1.09 | 2.0 | 29.5 |
| 10 | 1.06 | 1.9 | 31.5 |

For the physical sciences section, results were consistent between the scrambled versions of each test form, but not across test forms. For all forms, an average of 14 factors had eigenvalues > 1.0. The median first principal component (median eigenvalue= 7.7) accounted for an average of 12.2% of the variance; and, the median second factor (median eigenvalue= 3.3) accounted for an average of 4.5% of the variance. For 23 A&B, there were large correlations between the first factors and *biserials* (.95 and .96), and most of the items loaded on the first factor for the unrotated solution. For 15 A&B, on the other hand, the correlations between the factors and *biserials* were small (.29 and .26) and fewer items loaded on the first factor. In the rotated solutions, for 15 A&B, the items tended to load in groups based on item set affiliation. For all four forms, there were positive correlations between *discipline* and factor loadings (.42 to .47).

Table 5a. Physical Sciences, 1994, 15A

| Factor | Eigenvalue | % of Var | Cum. % |
|---|---|---|---|
| 1 | 7.79 | 12.4 | 12.4 |
| 2 | 3.83 | 6.1 | 18.4 |
| 3 | 2.30 | 3.7 | 22.1 |
| 4 | 2.23 | 3.5 | 25.6 |
| 5 | 1.78 | 2.8 | 28.5 |
| 6 | 1.58 | 2.5 | 31.0 |
| 7 | 1.48 | 2.3 | 33.3 |
| 8 | 1.44 | 2.3 | 35.6 |
| 9 | 1.24 | 2.0 | 37.6 |
| 10 | 1.20 | 1.9 | 39.5 |

Table 5b. Physical Sciences, 1994, 15B

| Factor | Eigenvalue | % of Var | Cum. % |
|---|---|---|---|
| 1 | 7.78 | 12.4 | 12.4 |
| 2 | 3.80 | 6.0 | 18.4 |
| 3 | 2.81 | 4.5 | 22.9 |
| 4 | 1.99 | 3.2 | 26.0 |
| 5 | 1.67 | 2.7 | 28.7 |
| 6 | 1.51 | 2.4 | 31.1 |
| 7 | 1.38 | 2.2 | 33.3 |
| 8 | 1.35 | 2.1 | 35.4 |
| 9 | 1.24 | 2.0 | 37.4 |
| 10 | 1.14 | 1.8 | 39.2 |

Table 8a. Physical Sciences, 1996, 23A

| Factor | Eigenvalue | % of Var | Cum. % |
|---|---|---|---|
| 1 | 7.50 | 11.9 | 11.9 |
| 2 | 1.82 | 2.9 | 14.8 |
| 3 | 1.25 | 2.0 | 16.8 |
| 4 | 1.24 | 2.0 | 18.7 |
| 5 | 1.12 | 1.8 | 20.5 |
| 6 | 1.11 | 1.8 | 22.3 |
| 7 | 1.09 | 1.7 | 24.0 |
| 8 | 1.06 | 1.7 | 25.7 |
| 9 | 1.05 | 1.7 | 27.4 |
| 10 | 1.04 | 1.7 | 29.0 |

Table 8b. Physical Sciences, 1996, 23B

| Factor | Eigenvalue | % of Var | Cum. % |
|---|---|---|---|
| 1 | 7.65 | 12.2 | 12.2 |
| 2 | 1.76 | 2.8 | 15.0 |
| 3 | 1.29 | 2.0 | 17.0 |
| 4 | 1.24 | 2.0 | 19.0 |
| 5 | 1.17 | 1.8 | 20.8 |
| 6 | 1.14 | 1.8 | 22.6 |
| 7 | 1.11 | 1.8 | 24.4 |
| 8 | 1.09 | 1.7 | 26.1 |
| 9 | 1.08 | 1.7 | 27.8 |
| 10 | 1.07 | 1.7 | 29.6 |

For the biological sciences section, results were consistent across form versions (A & B), and across test forms (15 & 23). For all forms, an average of 12 factors had eigenvalues > 1.0. The median first principal component (median eigenvalue= 7.3) accounted for an average of 11.5% of the variance; and, the median second factor (median eigenvalue= 1.8) accounted for an average of 2.9% of the variance. Similar to the physics test section, most items loaded on the first factor which correlated highly with the *biserials* (range .90 to .96). The second factor correlated positively with biology (range .79 to .90) and negatively with organic chemistry (-.79 to -.90). This indicates that these factors distinguished between biology and organic chemistry items.

Table 6a. Biological Sciences, 1994, 15A

| Factor | Eigenvalue | % of Var | Cum. % |
|---|---|---|---|
| 1 | 6.86 | 10.9 | 10.9 |
| 2 | 1.70 | 2.7 | 13.6 |
| 3 | 1.41 | 2.2 | 15.8 |
| 4 | 1.30 | 2.1 | 17.9 |
| 5 | 1.24 | 2.0 | 19.8 |
| 6 | 1.13 | 1.8 | 21.6 |
| 7 | 1.11 | 1.8 | 23.4 |
| 8 | 1.08 | 1.7 | 25.1 |
| 9 | 1.05 | 1.7 | 26.8 |
| 10 | 1.04 | 1.7 | 28.4 |

Table 6b. Biological Sciences, 1994, 15B

| Factor | Eigenvalue | % of Var | Cum. % |
|---|---|---|---|
| 1 | 7.52 | 11.9 | 11.9 |
| 2 | 1.75 | 2.8 | 14.7 |
| 3 | 1.44 | 2.3 | 17.0 |
| 4 | 1.31 | 2.1 | 19.1 |
| 5 | 1.20 | 1.9 | 21.0 |
| 6 | 1.14 | 1.8 | 22.8 |
| 7 | 1.09 | 1.7 | 24.5 |
| 8 | 1.08 | 1.7 | 26.3 |
| 9 | 1.05 | 1.7 | 27.9 |
| 10 | 1.03 | 1.6 | 29.6 |

Table 9a. Biological Sciences, 1996, 23A

| Factor | Eigenvalue | % of Var | Cum. % |
|---|---|---|---|
| 1 | 7.15 | 11.4 | 11.4 |
| 2 | 1.86 | 3.0 | 14.3 |
| 3 | 1.52 | 2.4 | 16.7 |
| 4 | 1.23 | 1.9 | 18.7 |
| 5 | 1.19 | 1.9 | 20.5 |
| 6 | 1.15 | 1.8 | 22.4 |
| 7 | 1.07 | 1.7 | 24.1 |
| 8 | 1.06 | 1.7 | 25.8 |
| 9 | 1.05 | 1.7 | 27.4 |
| 10 | 1.02 | 1.6 | 29.0 |

Table 9b. Biological Sciences, 1996, 23B

| Factor | Eigenvalue | % of Var | Cum. % |
|---|---|---|---|
| 1 | 7.45 | 11.8 | 11.8 |
| 2 | 1.87 | 3.0 | 14.8 |
| 3 | 1.46 | 2.3 | 17.1 |
| 4 | 1.22 | 1.9 | 19.1 |
| 5 | 1.21 | 1.9 | 21.0 |
| 6 | 1.15 | 1.8 | 22.8 |
| 7 | 1.12 | 1.8 | 24.6 |
| 8 | 1.11 | 1.8 | 26.3 |
| 9 | 1.09 | 1.7 | 28.1 |
| 10 | 1.07 | 1.7 | 29.8 |

The PCA results from the complete forms and test sections indicate that there is consistency across form versions (A & B). However, there are differences across test forms (15 & 23) in terms of loading patterns and the degree of multidimensionality. The eigenvalues and percentage of variance in the data accounted for by the first factor suggests that each test section is multidimensional. (To conclude each section is unidimensional, we would have liked to see the first factor accounting for at least 20% of the variance in the data.) In addition, the PCA results indicate that there were distinctions among the various *disciplines*. This finding was used to make parceling decisions for the next series of analysis. Items also tended to group by *passage* or by set. There were not many significant correlations between the factors and the cognitive classifications, therefore, item parcels were created based on the discipline classification of the items. As mentioned earlier, PCA and other methods of exploratory factor analysis have been criticized when applied to dichotomous data because spurious factors often emerge due to unreliability and non-linearity of item-level data. Thus, we move now to analysis of the parceled data, which obviates these problems.

Global Parcel-level PCA Results

The eigenvalues and the percentage of variance accounted for by each factor for all forms are presented in Tables 11a through 12b. For 15A and 15B (35 parcels) the results were nearly identical. Four factor solutions were extracted for both forms. For 15A, the first principal component (eigenvalue= 10.7) accounted for 30.5% of the variance; and, the second factor (eigenvalue= 2.3) accounted for 6.7% of the variance. Similarly, for 15B, the first principal component (eigenvalue=11.8) accounted for 33.7% of the variance; and, the second factor (eigenvalue= 2.3) accounted for 6.5% of the variance. Following are the factor loadings for the unrotated solution: a) VR parcels and all science parcels loaded on factor 1, and b) the essay parcels loaded on factor 3. The verbal reasoning items did have sizable loadings on factor 2, but they were smaller than the loadings on factor 1. The rotated solution tended to distinguish among the science parcels. It yielded the following factor loadings: a) physics, general chemistry and organic chemistry parcels loaded on factor 1, b) all VR parcels loaded on factor 2, c) biology parcels loaded on factor 3, and d) essay scores loaded on factor 4. This PCA solution makes distinctions between biology and all other sciences. These loadings were replicated with 15B. Tables 13a to 14b contain the loadings for the unrotated and rotated PCA solutions for 15A and 15B. In general, the results suggest a strong dominant factor with smaller factors related to *discipline* areas.

**Summary of Parcel-level PCA Results for Forms 15 A&B**

Table 11a. Parcel-level, Form 15A, 1994

| Factor | Eigenvalue | % of Var | Cum. % |
|---|---|---|---|
| 1 | 10.66 | 30.5 | 30.5 |
| 2 | 2.35 | 6.7 | 37.2 |
| 3 | 1.32 | 3.8 | 40.9 |
| 4 | 1.10 | 3.1 | 44.1 |
| 5 | .93 | 2.6 | 46.7 |
| 6 | .89 | 2.5 | 49.3 |
| 7 | .80 | 2.3 | 51.6 |
| 8 | .77 | 2.2 | 53.8 |
| 9 | .74 | 2.1 | 55.9 |
| 10 | .72 | 2.1 | 57.9 |

Table 11b. Parcel-level, Form 15B, 1994

| Factor | Eigenvalue | % of Var | Cum. % |
|---|---|---|---|
| 1 | 11.79 | 33.7 | 33.7 |
| 2 | 2.26 | 6.5 | 40.1 |
| 3 | 1.28 | 3.7 | 43.8 |
| 4 | 1.05 | 3.0 | 46.8 |
| 5 | .89 | 2.6 | 49.4 |
| 6 | .83 | 2.4 | 51.7 |
| 7 | .79 | 2.2 | 54.0 |
| 8 | .71 | 2.0 | 56.0 |
| 9 | .71 | 2.0 | 58.0 |
| 10 | .60 | 2.0 | 60.0 |

**Parcel-level Analysis**
**Factor Loadings for Both PCA Solutions, Form: 15A**

Table 13a. Unrotated Solution -- 15A

| Parcel | Factor 1 | Factor 2 |
|---|---|---|
| VAF: | 30.5% | 6.7% |
| VR1hum | .54 | |
| VR2hum | .50 | |
| VR3hum | .49 | |
| VR4nst | .43 | |
| VR5nst | .50 | |
| VR6nst | .51 | |
| VR7ssc | .50 | |
| VR8ssc | .45 | |
| VR9ssc | .54 | |
| VR10ssc | .51 | |
| phy1 | .60 | |
| phy2 | .57 | |
| phy3 | .57 | |
| phy4 | .61 | |
| phy5 | .66 | |
| gch1 | .55 | |
| gch2 | .58 | |
| gch3 | .57 | |
| gch4 | .58 | |
| gch5 | .66 | |
| gch6 | .65 | |
| blg1 | .61 | |
| blg2 | .65 | |
| blg3 | .59 | |
| blg4 | .58 | |
| blg5 | .59 | |
| blg6 | .55 | |
| blg7 | .56 | |
| blg8 | .58 | |
| org1 | .54 | |
| org2 | .57 | |
| org3 | .51 | |
| org3 | .47 | |
| e1score | | .73 |
| e2score | | .66 |

Table 13b, Rotated Solution -- 15A

| Parcel | Factor 1 | Factor 2 | Factor 3 | Factor 4 |
|---|---|---|---|---|
| VAF: | 30.5% | 6.7% | 3.8% | 3.1% |
| VR1hum | | .64 | | |
| VR2hum | | .60 | | |
| VR3hum | | .59 | | |
| VR4nst | | .52 | | |
| VR5nst | | .63 | | |
| VR6nst | | .63 | | |
| VR7ssc | | .62 | | |
| VR8ssc | | .53 | | |
| VR9ssc | | .59 | | |
| VR10ssc | | .60 | | |
| phy1 | .53 | | | |
| phy2 | .50 | | | |
| phy3 | .51 | | | |
| phy4 | .65 | | | |
| phy5 | .61 | | | |
| gch1 | .54 | | | |
| gch2 | .61 | | | |
| gch3 | .55 | | | |
| gch4 | .61 | | | |
| gch5 | .62 | | | |
| gch6 | .59 | | | |
| blg1 | | | .57 | |
| blg2 | | | .51 | |
| blg3 | | | .48 | |
| blg4 | | | .46 | |
| blg5 | | | .65 | |
| blg6 | | | .43 | |
| blg7 | | | .65 | |
| blg8 | | | .60 | |
| org1 | .62 | | | |
| org2 | .60 | | | |
| org3 | .45 | | | |
| org3 | .55 | | | |
| e1score | | | | .88 |
| e2score | | | | .84 |

**Parcel Level Analysis**
**Factor Loadings for Both PCA Solutions, Form: 15B**

Table 14a, Unrotated Solution -- 15B

| Parcel | Factor 1 | Factor 2 |
|---|---|---|
| VAF: | 33.7% | 6.5% |
| VR1hum | .58 | |
| VR2hum | .56 | |
| VR3hum | .53 | |
| VR4nst | .49 | |
| VR5nst | .57 | |
| VR6nst | .57 | |
| VR7ssc | .55 | |
| VR8ssc | .50 | |
| VR9ssc | .54 | |
| VR10ssc | .53 | |
| phy1 | .61 | |
| phy2 | .59 | |
| phy3 | .60 | |
| phy4 | .64 | |
| phy5 | .68 | |
| gch1 | .56 | |
| gch2 | .61 | |
| gch3 | .61 | |
| gch4 | .61 | |
| gch5 | .69 | |
| gch6 | .68 | |
| blg1 | .62 | |
| blg2 | .66 | |
| blg3 | .60 | |
| blg4 | .60 | |
| blg5 | .63 | |
| blg6 | .58 | |
| blg7 | .59 | |
| blg8 | .61 | |
| org1 | .54 | |
| org2 | .58 | |
| org3 | .51 | |
| org3 | .53 | |
| e1score | | .74 |
| e2score | | .66 |

Table 14b, Rotated Solution -- 15B

| Parcel | Factor 1 | Factor 2 | Factor 3 | Factor 4 |
|---|---|---|---|---|
| VAF: | 33.7% | 6.5% | 3.7% | 3.0% |
| VR1hum | | .66 | | |
| VR2hum | | .64 | | |
| VR3hum | | .61 | | |
| VR4nst | | .56 | | |
| VR5nst | | .64 | | |
| VR6nst | | .65 | | |
| VR7ssc | | .63 | | |
| VR8ssc | | .52 | | |
| VR9ssc | | .60 | | |
| VR10ssc | | .61 | | |
| phy1 | .57 | | | |
| phy2 | .52 | | | |
| phy3 | .56 | | | |
| phy4 | .65 | | | |
| phy5 | .64 | | | |
| gch1 | .51 | | | |
| gch2 | .61 | | | |
| gch3 | .56 | | | |
| gch4 | .62 | | | |
| gch5 | .63 | | | |
| gch6 | .61 | | | |
| blg1 | | | .56 | |
| blg2 | | | .52 | |
| blg3 | | | .51 | |
| blg4 | | | .49 | |
| blg5 | | | .64 | |
| blg6 | | | .46 | |
| blg7 | | | .66 | |
| blg8 | | | .60 | |
| org1 | .64 | | | |
| org2 | .59 | | | |
| org3 | .48 | | | |
| org3 | .56 | | | |
| e1score | | | | .87 |
| e2score | | | | .83 |

For 23A and 23B (34 parcels) the results were similar to each other and to 15A and 15B. Three and four factor solutions were extracted for each form respectively. For 23A, the first principal component (eigenvalue= 10.7) accounted for 31.4% of the variance; and, the second factor (eigenvalue= 2.2) accounted for 6.4% of the variance. For 23B, the first principal component (eigenvalue= 10.9) accounted for 32.3% of the variance; and, the second factor (eigenvalue= 2.2) accounted for 6.4% of the variance. The rotated solution for 23A, suggested a 3 factor solution: a) all science parcels loaded on factor 1, b) VR parcels loaded on factor 2, and c) essay scores loaded on factor 3. The rotated solution for 23B, like the solutions for 15A and 15B, distinguished between biology parcels (which loaded on factor 3) and all other science parcels (which loaded on factor 1). Tables 15a to 16b contain the loadings for the unrotated and rotated PCA solutions for 23A and 23B.

**Summary of Parcel-level PCA Results for:**
**Forms 23 A&B**

Table 12a. Parcel-level, Form 23A, 1996

| Factor | Eigenvalue | % of Var | Cum. % |
|---|---|---|---|
| 1 | 10.67 | 31.4 | 31.4 |
| 2 | 2.19 | 6.4 | 37.8 |
| 3 | 1.31 | 3.9 | 41.7 |
| 4 | 1.00 | 2.9 | 44.6 |
| 5 | .93 | 2.7 | 47.4 |
| 6 | .83 | 2.4 | 52.1 |
| 7 | .79 | 2.3 | 54.4 |
| 8 | .78 | 2.3 | 56.7 |
| 9 | .77 | 2.3 | 58.8 |
| 10 | .72 | 2.1 | 60.8 |

Table 12b. Parcel-level, Form 23B, 1996

| Factor | Eigenvalue | % of Var | Cum. % |
|---|---|---|---|
| 1 | 10.98 | 32.3 | 32.3 |
| 2 | 2.17 | 6.4 | 38.7 |
| 3 | 1.28 | 3.8 | 42.4 |
| 4 | 1.04 | 3.1 | 45.5 |
| 5 | .91 | 2.7 | 48.2 |
| 6 | .88 | 2.6 | 50.7 |
| 7 | .81 | 2.4 | 53.1 |
| 8 | .78 | 2.3 | 55.4 |
| 9 | .75 | 2.2 | 57.6 |
| 10 | .72 | 2.1 | 59.7 |

**Parcel Level Analysis**
**Factor Loadings for Both PCA Solutions, Form: 23A**

Table 15a. Unrotated Solution -- 23A

| Parcel | Factor 1 | Factor 2 | Factor 3 |
|---|---|---|---|
| VAF: | 31.4% | 6.4% | 3.9% |
| VR1hum | .44 | | |
| VR2hum | .48 | | |
| VR3hum | .47 | | |
| VR4nst | .43 | | |
| VR5nst | .46 | | |
| VR6nst | .57 | | |
| VR7ssc | .55 | | |
| VR8ssc | .58 | | |
| VR9ssc | .47 | | |
| VR10ssc | .48 | | |
| phy1 | .60 | | |
| phy2 | .57 | | |
| phy3 | .61 | | |
| phy4 | .52 | | |
| phy5 | .55 | | |
| phy6 | .61 | | |
| gch1 | .55 | | |
| gch2 | .60 | | |
| gch3 | .60 | | |
| gch4 | .64 | | |
| gch5 | .63 | | |
| blg1 | .58 | | |
| blg2 | .59 | | |
| blg3 | .63 | | |
| blg4 | .61 | | |
| blg5 | .61 | | |
| blg6 | .63 | | |
| blg7 | .62 | | |
| blg8 | .60 | | |
| org1 | .61 | | |
| org2 | .56 | | |
| org3 | .52 | | |
| e1score | | | .73 |
| e2score | | | .69 |

Table 15b. Rotated Solution -- 23A

| Parcel | Factor 1 | Factor 2 | Factor 3 |
|---|---|---|---|
| VAF: | 31.4% | 6.4% | 3.9% |
| VR1hum | | .58 | |
| VR2hum | | .61 | |
| VR3hum | | .54 | |
| VR4nst | | .50 | |
| VR5nst | | .51 | |
| VR6nst | | .55 | |
| VR7ssc | | .65 | |
| VR8ssc | | .66 | |
| VR9ssc | | .55 | |
| VR10ssc | | .57 | |
| phy1 | .56 | | |
| phy2 | .51 | | |
| phy3 | .60 | | |
| phy4 | .52 | | |
| phy5 | .50 | | |
| phy6 | .61 | | |
| gch1 | .62 | | |
| gch2 | .65 | | |
| gch3 | .65 | | |
| gch4 | .64 | | |
| gch5 | .64 | | |
| blg1 | .44 | | |
| blg2 | .47 | | |
| blg3 | .51 | | |
| blg4 | .50 | | |
| blg5 | .48 | | |
| blg6 | .52 | | |
| blg7 | .49 | | |
| blg8 | .41 | | |
| org1 | .65 | | |
| org2 | .61 | | |
| org3 | .63 | | |
| e1score | | | .84 |
| e2score | | | .81 |

**Parcel Level Analysis**
**Factor Loadings for PCA Solutions, Form: 23B**

Table 16a, Unrotated Solution -- 23B

| Parcel | Factor 1 | Factor 2 |
|---|---|---|
| VAF: | 32.3% | 6.4% |
| VR1hum | .42 | |
| VR2hum | .48 | |
| VR3hum | .47 | |
| VR4nst | .45 | |
| VR5nst | .43 | |
| VR6nst | .56 | |
| VR7ssc | .56 | |
| VR8ssc | .61 | |
| VR9ssc | .49 | |
| VR10ssc | .53 | |
| phy1 | .60 | |
| phy2 | .57 | |
| phy3 | .61 | |
| phy4 | .50 | |
| phy5 | .56 | |
| phy6 | .63 | |
| gch1 | .58 | |
| gch2 | .62 | |
| gch3 | .59 | |
| gch4 | .64 | |
| gch5 | .62 | |
| blg1 | .57 | |
| blg2 | .59 | |
| blg3 | .64 | |
| blg4 | .63 | |
| blg5 | .59 | |
| blg6 | .66 | |
| blg7 | .63 | |
| blg8 | .60 | |
| org1 | .63 | |
| org2 | .55 | |
| org3 | .51 | |
| e1score | | .67 |
| e2score | | .62 |

Table 16b, Rotated Solution -- 23B

| Parcel | Factor 1 | Factor 2 | Factor 3 | Factor 4 |
|---|---|---|---|---|
| VAF: | 32.3% | 6.4% | 3.8% | 3.1% |
| VR1hum | | .60 | | |
| VR2hum | | .65 | | |
| VR3hum | | .60 | | |
| VR4nst | | .50 | | |
| VR5nst | | .44 | | |
| VR6nst | | .52 | | |
| VR7ssc | | .59 | | |
| VR8ssc | | .63 | | |
| VR9ssc | | .55 | | |
| VR10ssc | | .59 | | |
| phy1 | .48 | | | |
| phy2 | .41 | | | |
| phy3 | .48 | | | |
| phy4 | .48 | | | |
| phy5 | .39 | | | |
| phy6 | .55 | | | |
| gch1 | .61 | | | |
| gch2 | .64 | | | |
| gch3 | .65 | | | |
| gch4 | .58 | | | |
| gch5 | .56 | | | |
| blg1 | | | .51 | |
| blg2 | | | .60 | |
| blg3 | | | .55 | |
| blg4 | | | .57 | |
| blg5 | | | .62 | |
| blg6 | | | .57 | |
| blg7 | | | .63 | |
| blg8 | | | .53 | |
| org1 | .63 | | | |
| org2 | .66 | | | |
| org3 | .72 | | | |
| e1score | | | | .85 |
| e2score | | | | .82 |

Global Parcel-level MDS Results

Like the PCA solutions, the MDS results were similar across all four forms of the test. For each form the statistics indicate a good fit of the data in three-dimensions (median statistics for the 3-dimensional solutions: RSQ = .96 and Stress = .10). The fit statistics are presented in Table 17. The first dimension clearly corresponds to verbal reasoning parcels versus science parcels. The second dimension is an essay (Writing Sample) score versus non-essay score dimension. The third dimension is not easily interpreted visually; although, it does seem to make small distinctions among groups of science items.

**Parcel Level Analysis**
**MDS Fit Statistics, All four forms**

Table 17, MDS fit statistics for all four forms.

| Form | Dimensional Solution | Stress | R-squared |
|---|---|---|---|
| 15A | 1 | .22 | .87 |
| | 2 | .13 | .94 |
| | 3 | .10 | .96 |
| | 4 | .08 | .97 |
| | 5 | .06 | .98 |
| | 6 | .06 | .98 |
| 15B | 1 | .22 | .89 |
| | 2 | .15 | .93 |
| | 3 | .10 | .96 |
| | 4 | .08 | .97 |
| | 5 | .07 | .97 |
| | 6 | .06 | .98 |
| 23A | 1 | .28 | .78 |
| | 2 | .13 | .93 |
| | 3 | .10 | .96 |
| | 4 | .07 | .97 |
| | 5 | .07 | .97 |
| | 6 | .06 | .98 |
| 23B | 1 | .28 | .79 |
| | 2 | .17 | .89 |
| | 3 | .11 | .94 |
| | 4 | .09 | .95 |
| | 5 | .08 | .96 |
| | 6 | .07 | .97 |

* shading indicates accepted solution

The relationships among the parcels are most easily interpreted by looking at the scatter plots of the stimulus coordinates (Figures 1a to 4). The two-dimensional plots for all four forms were very similar. In each case, the verbal reasoning parcels are polarized from the science parcels along the first dimension; and, the essay parcels separate from all other items along the second dimension. Clearly there are three clusters of different parcel types. Within the science item clusters there is evidence that some parcels are grouping by discipline. For example, at the farthest end of dimension one there is a grouping of organic chemistry parcels. Follow-up analysis using cluster analysis are planned. The MDS results, appear to be consistent with the parcel-level PCA solutions.

**Parcel Level Analysis**
**MDS 2-d Stimulus Plots for 15A & 15B**

Figure 1a, 2-d MDS Stimulus Plot for 15A

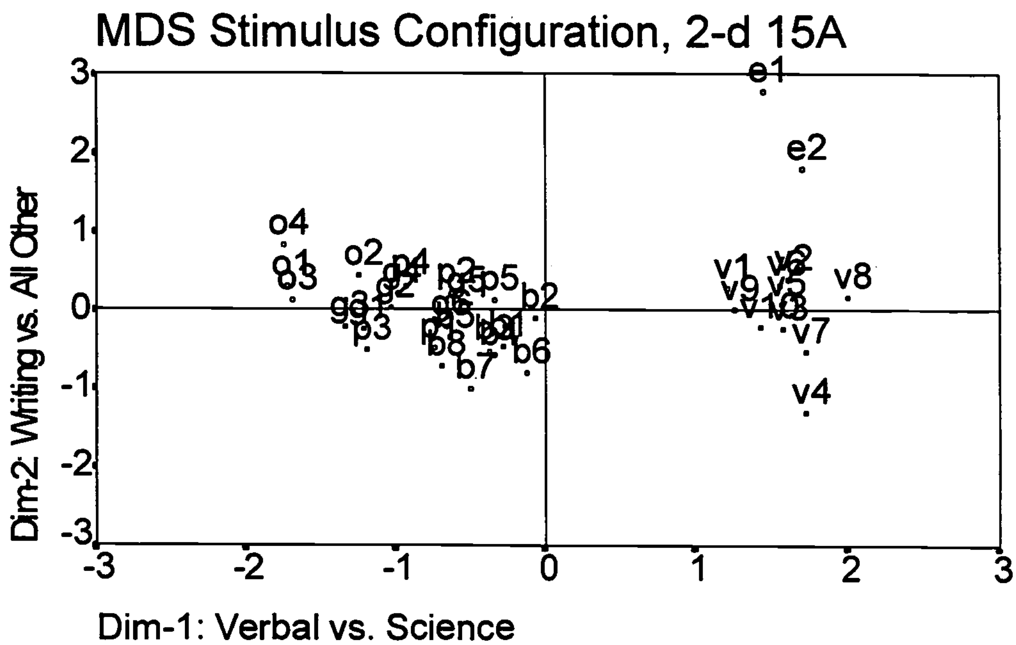
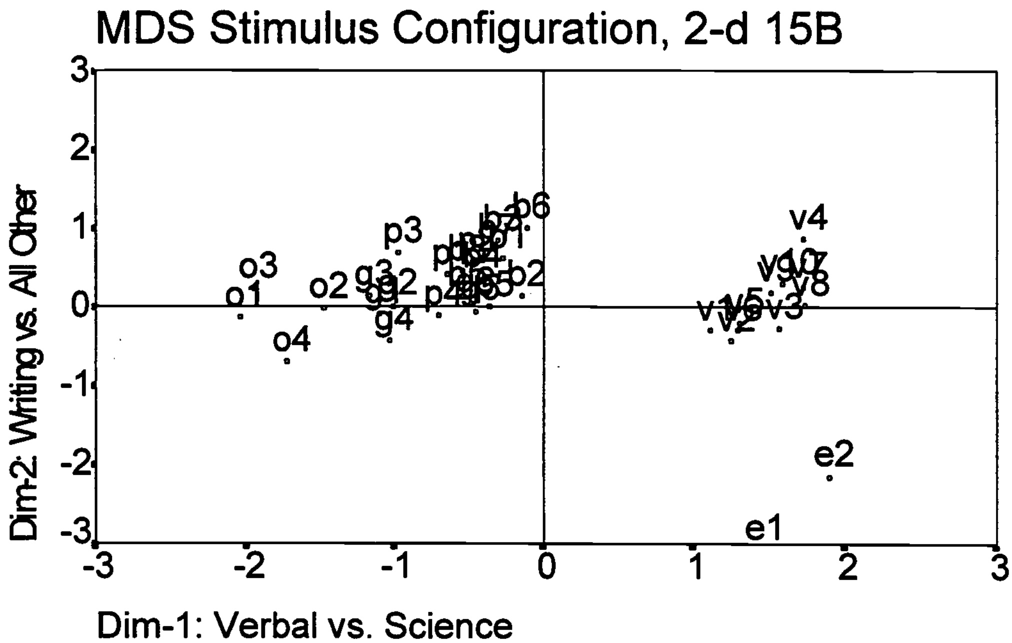


Figure 1b, 2-d MDS Stimulus Plot for 15B



**Parcel Level Analysis**
**MDS 2-d Stimulus Plots for 23A & 23B**

Figure 2a, 2-d MDS Stimulus Plot for 23A

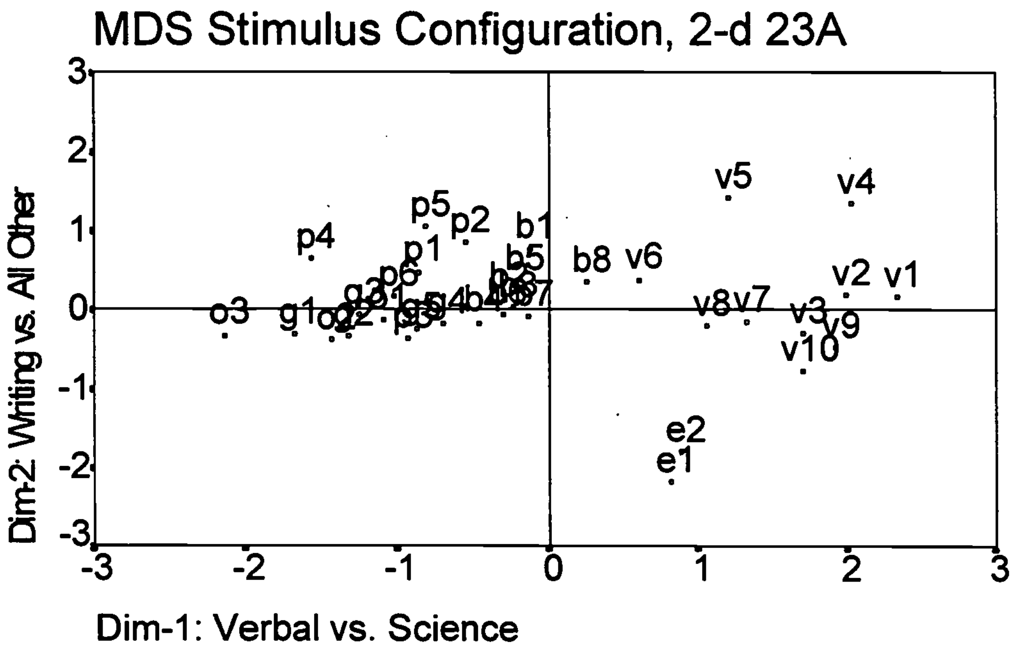
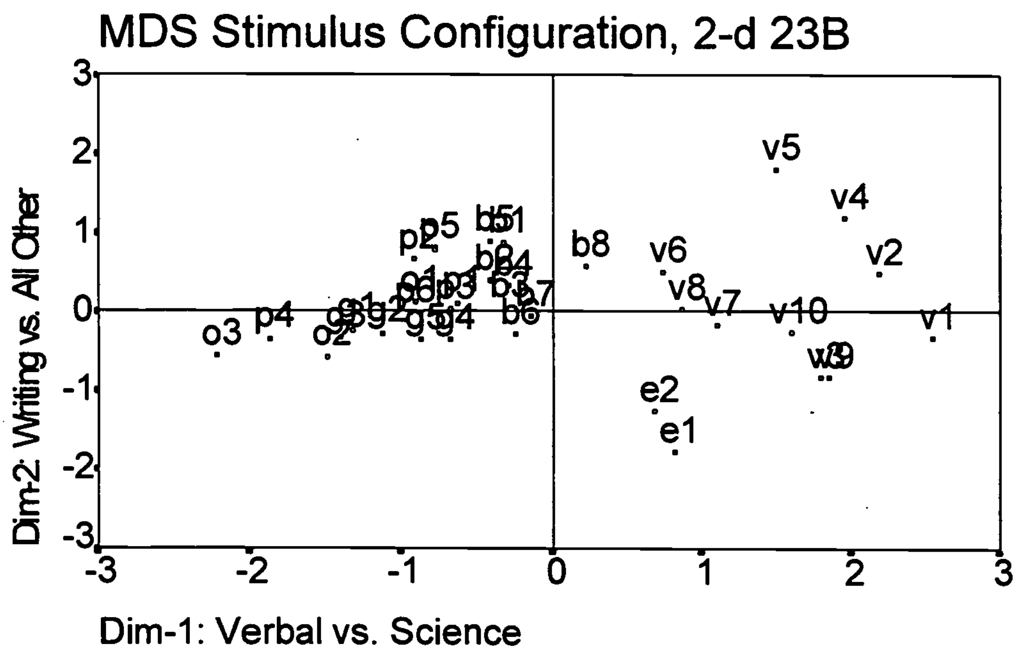



Figure 2b, 2-d MDS Stimulus Plot for 23B



**Parcel Level Analysis**
**MDS 3-d Stimulus Plots, 15A and 23A**

Figure 3, 3-d MDS Stimulus Plot for 15A

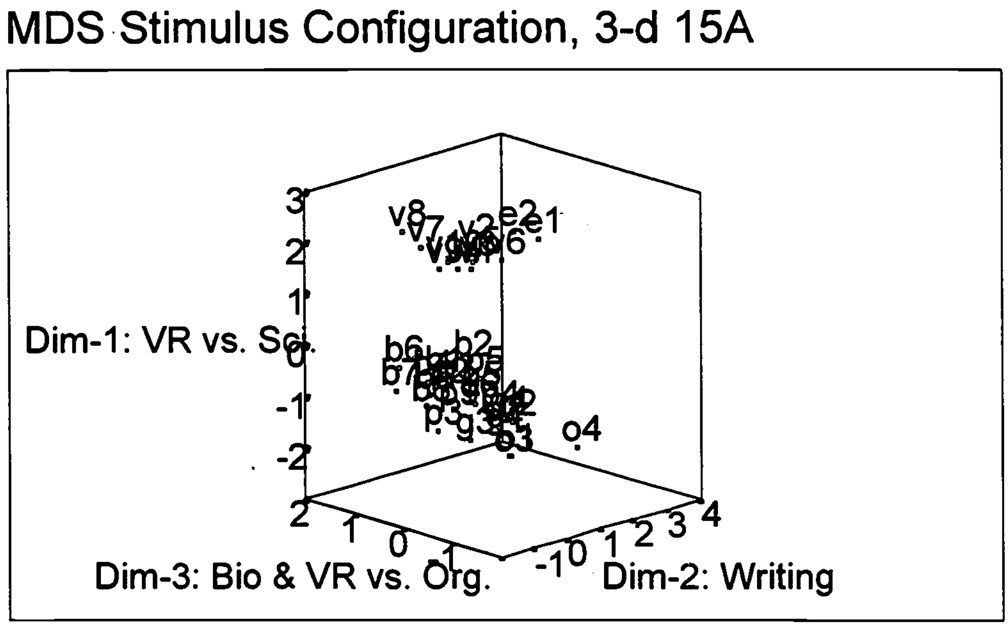
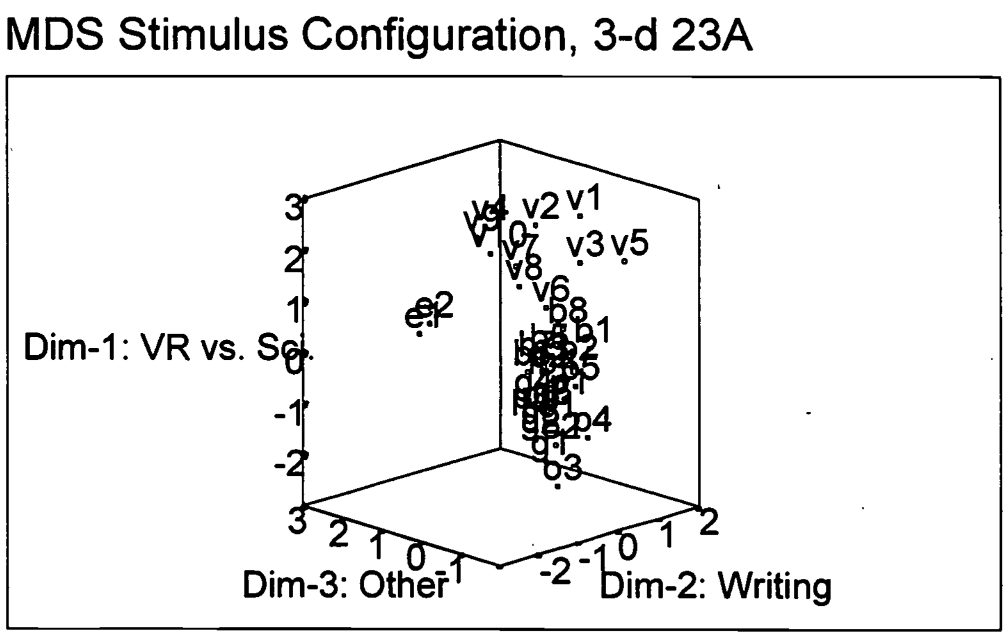



Figure 4, 3-d MDS Stimulus Plot for 23A



Global Parcel-level CFA Results

Table 18 summarizes the fit statistics for the CFAs. The results were identical across forms and exams. The one-factor models did not appear to fit the data well. The two-, three-, $four_a$-, $four_b$-, six- and eight-factor models, on the other hand, all appear to fit the data well. The goodness of fit index for the two-factor model ranged from .92 to.94 and the median RMSR was .036. This result suggests that one underlying construct can adequately account for all of the *science* items. There was some improvement, however, when science parcels were separated by discipline in the six-factor model. The GFI for the six-factor model were all equal to .98 and the median RMSR=.020. Also notable, the 4b-model fit the data slightly better (AGFI=.98, RMSR=.020) than the 4a-factor model (AGFI=.96, RMSR=.024). The 4a model conforms to the score reporting scheme currently used for the MCAT; whereas, model 4b separates biology from all other sciences. The improvement in fit from the six-factor model to the eight-factor model was negligible. Diagrams of above models can be found in Appendix A.

Which solution fits best? To draw some conclusions, we used the change in Chi-square values from model to model as an indicator of improvement in fit. For Form 15A, the largest change occurs from the one-factor model (which does not fit according to the GFI) to the two-factor model (change=6438). Among models that fit, there is a sizable change in the chi-square value between the two- and three-factor models (2539), and there is a sizable difference between the three- and six-factor models (1855). There is little improvement between the three- and $four_a$-factor models (635), and between the six- and eight-factor models (520). It appears, then, that the $four_a$-factor model offers little advantage over the three-factor model. Thus, the three- and six-factor models may be the best fitting choices (depending on how test scores are used). In summary, the results of the CFA showed consistency with the other parcel-level analysis and suggest that there is consistency across form variations (A & B) and across tests (15 & 23). In addition, they provide evidence that the test forms (15A, 15B, 23A and 23B) are measuring at least three underlying constructs: a) verbal reasoning, b) science, and c) writing.

Results of Group Analyses

Group WMDS Results

As described in the global analyses, two- and three-dimensional MDS models seemed appropriate for the MCAT data. The two-dimensional models distinguished the verbal reasoning test section and writing sample from the two science test sections. The three-dimensional model appeared to distinguish some of the science disciplines. In applying the weighted MDS models to the group data, we expected to uncover similar dimensions. However, in weighted MDS models, additional dimensions are typically needed if one or more dimensions are needed to account for systematic variation in one or more groups. Because form differences were not noted in the global analyses, and because some of the group sample sizes were small, the data were combined across the unscrambled (Form A) and scrambled (Form B) versions. All analyses were performed separately on Form 15 and 23. The criteria for selecting the best MDS solution were fit (STRESS and $R^2$), interpretability, and consistency across replications.

**Parcel Level Analysis**
**CFA Fit Statistics, All four forms**

Table 18, CFA fit statistics for all four forms.

| Form | No. of Factors | GFI / AGFI | RMSR | Chi-Square | DF |
|---|---|---|---|---|---|
| 15A | 1 | .84 / .82 | .055 | 12,810 | 560 |
| | 2 | .94 / .93 | .033 | 6,372 | 559 |
| | 3 | .96 / .95 | .027 | 3,833 | 557 |
| | 4a | .97 / .96 | .024 | 3,198 | 554 |
| | 4b | .98 / .97 | .020 | 2,506 | 554 |
| | 6 | .98 / .98 | .018 | 1,978 | 545 |
| | 8 | .99 / .98 | .028 | 1,458 | 532 |
| 15B | 1 | .83 / .82 | .053 | 12,248 | 560 |
| | 2 | .94 / .93 | .032 | 6,035 | 559 |
| | 3 | .96 / .95 | .026 | 3,662 | 557 |
| | 4a | .97 / .96 | .024 | 2,948 | 554 |
| | 4b | .98 / .97 | .020 | 2,340 | 554 |
| | 6 | .98 / .98 | .017 | 1,660 | 545 |
| | 8 | .99 / .98 | .016 | 1,375 | 532 |
| 23A | 1 | .85 / .84 | .052 | 12,094 | 527 |
| | 2 | .93 / .92 | .038 | 7,259 | 526 |
| | 3 | .95 / .95 | .033 | 4,225 | 524 |
| | 4a | .96 / .95 | .030 | 3,565 | 521 |
| | 4b | .97 / .97 | .025 | 2,602 | 521 |
| | 6 | .98 / .98 | .021 | 1,698 | 512 |
| | 8 | .99 / .99 | .017 | 1,287 | 499 |
| 23B | 1 | .85 / .83 | .051 | 6,996 | 527 |
| | 2 | .92 / .91 | .038 | 4,347 | 526 |
| | 3 | .95 / .94 | .033 | 2,650 | 524 |
| | 4a | .96 / .95 | .030 | 2,291 | 521 |
| | 4b | .97 / .97 | .026 | 1,669 | 521 |
| | 6 | .98 / .98 | .022 | 1,236 | 512 |
| | 8 | .98 / .98 | .019 | 1,053 | 499 |

First-time/repeaters. In reporting the results of the group analyses, we start with the repeater/non-repeater analysis, which involved four groups. Fit statistics and subject weights for Forms 15 and 23 can be found in Tables 19a-20c. For both test forms, the three-dimensional solution appeared best. The STRESS values were .15 and .14, and the $R^2$ values were .90 and .89, for Forms 15 and 23, respectively. The first dimension separated the verbal reasoning parcels and writing samples from the science parcels; the second dimension distinguished the organic chemistry parcels from the other parcels; and the third dimension separated the writing samples from all the parcels. The percentages of variance accounted for by the first through third dimensions were 42%, 28%, and 20%, respectively for Form 15, and 57%, 21%, and 11% respectively for Form 23. No differences among the subject weights across the four groups were observed. The three dimensions appeared to account for the variation in the data for all groups of male and female repeaters and non-repeaters in similar fashion. Figures 5a through 6b show the subject weights and stimulus configurations for the three-dimensional solution for Forms 15 and 23.

**Multi-group WMDS Analysis:**
**Repeaters Form 15**

Table 19a, MDS Fit Statistics for Repeaters, Averaged Over Matrices.

| Form | Dimensional Solution | Stress | R-squared |
|---|---|---|---|
| 15 | 2 | .18 | .88 |
| | 3 | .15 | .90 |
| | 4 | .13 | .91 |
| | 5 | .11 | .93 |
| | 6 | .09 | .94 |

* shading indicates accepted solution

Table 19b, MDS Fit Statistics for Each Repeater Group for the 3-Dimensional Solution.

| Form | Matrix | Stress | R-squared |
|---|---|---|---|
| 15 | 1. First-timer, Male | .14 | .92 |
| | 2. First-timer, Female | .16 | .88 |
| | 3. Repeater, Male | .14 | .91 |
| | 4. Repeater, Female | .16 | .88 |

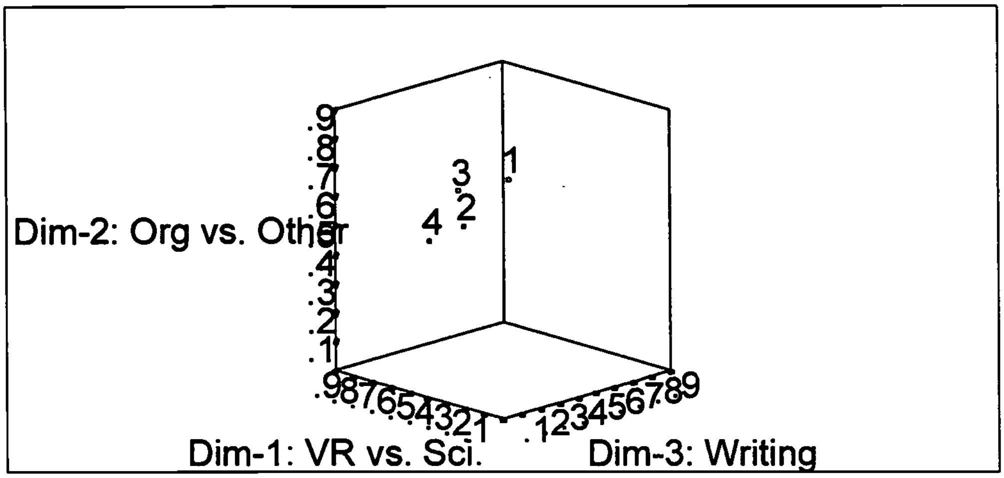
Table 19c, MDS Subject Weights for Each Repeater Group for the 3-Dimensional Solution.

| | | | | Dimension | |
|---|---|---|---|---|---|
| Form | Matrix | Weirdness | 1 | 2 | 3 |
| 15 | 1. First-timer, Male | .21 | .48 | .64 | .52 |
| | 2. First-timer, Female | .10 | .68 | .45 | .47 |
| | 3. Repeater, Male | .10 | .63 | .59 | .39 |
| | 4. Repeater, Female | .20 | 76 | .40 | .37 |
| | Importance of each dimension: | | .42 | .28 | .20 |

**Multi-group WMDS Analysis:**
**Repeaters Form 15**

Figure 5a,

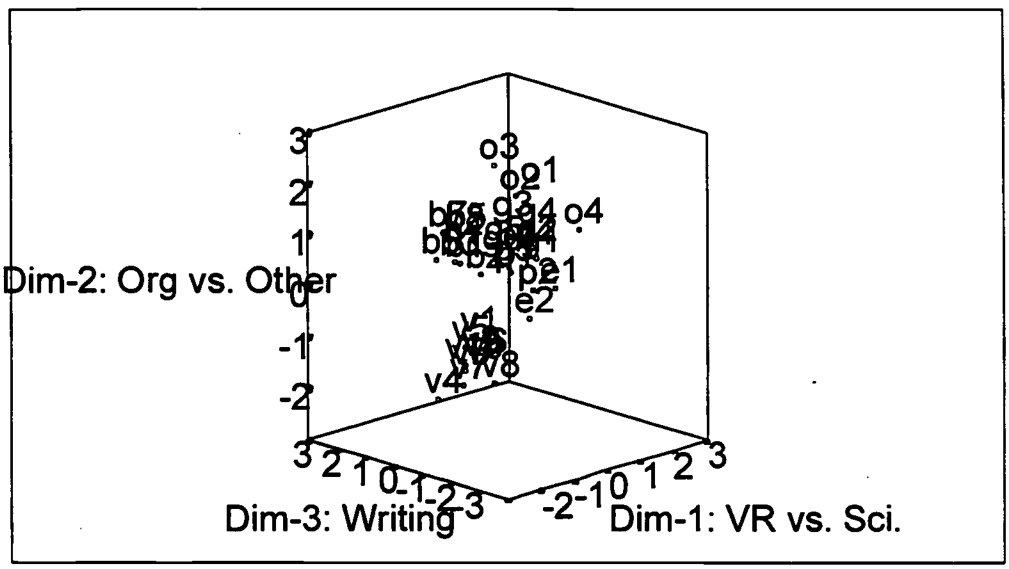
Derived Subject Weights: Repeaters, Form 15.




1) One-timer, Male 3) Repeater, Male
2) One-timer, Female 4) Repeater, Female

Figure 5b,

Stimulus Configuration: Repeaters, Form 15



**Multi-group WMDS Analysis:**
**Repeaters Form 23**

Table 20a, MDS Fit Statistics for Repeaters, Averaged Over Matrices.

| Form | Dimensional Solution | Stress | R-squared |
|---|---|---|---|
| 23 | 2 | .19 | .85 |
| | 3 | .14 | .89 |
| | 4 | .12 | .92 |
| | 5 | .11 | .92 |
| | 6 | .09 | .93 |

* shading indicates accepted solution

Table 20b, MDS Fit Statistics for Each Repeater Group for the Three-Dimensional Solution.

| Form | Matrix | Stress | R-squared |
|---|---|---|---|
| 23 | 1. First-timer, Male | .14 | .90 |
| | 2. First-timer, Female | .15 | .89 |
| | 3. Repeater, Male | .14 | .90 |
| | 4. Repeater, Female | .15 | .88 |

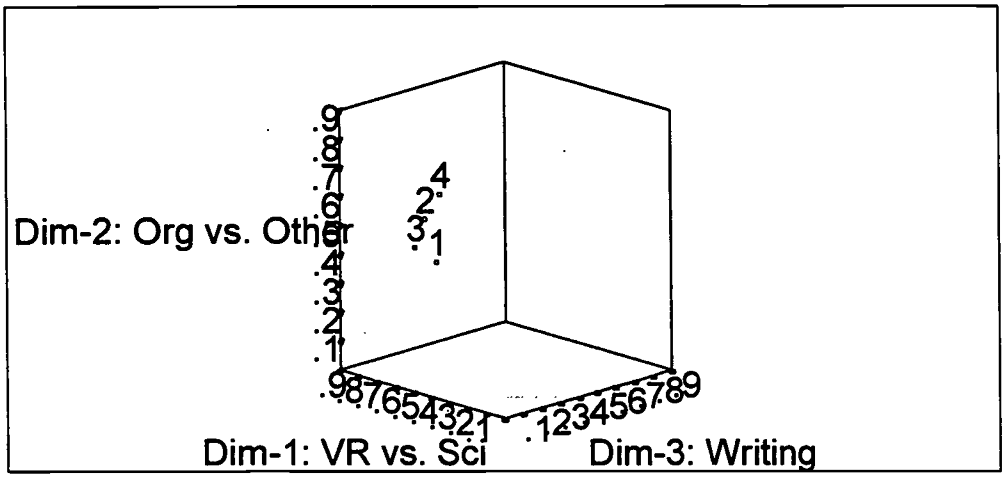
Table 20c, MDS Subject Weights for Each Repeater Group for the Three-Dimensional Solution.

| | | | Dimension | | |
|---|---|---|---|---|---|
| Form | Matrix | Weirdness | 1 | 2 | 3 |
| 23 | 1. First-timer, Male | .23 | .79 | .31 | .42 |
| | 2. First-timer, Female | .08 | .75 | .49 | .31 |
| | 3. Repeater, Male | .10 | .81 | .38 | .32 |
| | 4. Repeater, Female | .22 | .66 | .60 | .30 |
| | Importance of each dimension: | | .57 | .21 | .11 |

**Multi-group WMDS Analysis:**
**Repeaters Form 23**

Figure 6a,

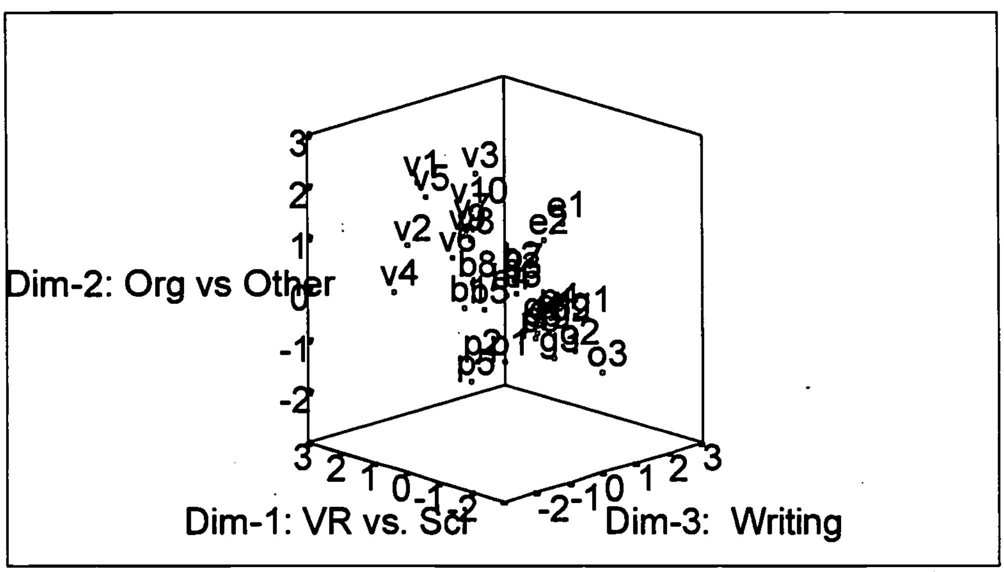
Derived Subject Weights: Repeaters, Form 23.




1) One-timer, Male 3) Repeater, Male
2) One-timer, Female 4) Repeater, Female

Figure 6b,

Stimulus Configuration: Repeaters, Form 23.



Native English/ESL. Six matrices were involved in the analysis of structure consistency across test takers with expected differences in English proficiency. The first two matrices were derived from males and females who self-reported themselves as native speakers of English. The third and fourth matrices were derived from male and female examinees who reported they learned English between the ages of 6 and 10. The fifth and sixth matrices were derived from males and females who reported they learned English after the age of ten. Fit statistics and subject weights for both forms can be found in Tables 21a-22c. For both test forms, a four-dimensional solution was deemed most appropriate. The STRESS values were .18 for both test forms, and the $R^2$ values were .80 and .79, for Forms 15 and 23, respectively. The verbal and essay dimensions noted in the repeater analyses re-emerged, as did separate dimensions for the non-biological sciences and the biological sciences. For Form 15, the percentages of variance accounted for by the first through fourth dimensions were 44%, 15%, 11%, and 9%, respectively. The results were similar for Form 23 (42%, 15%, 12% and 10%). Inspection of the group weights revealed one notable difference among the groups. Females who learned English between 6 and 10 years old had a relatively lower weight on the "essay" dimension for both forms. Thus, the writing samples accounted for less variation in the data for these females, relative to the other groups. This difference is illustrated in Figure 7b, which displays the subject weights from a two-dimensional subspace of this solution. Figures 7a through 10c show the subject weights and stimulus configurations for the four-dimensional solution for all forms.

**Multi-group WMDS Analysis:**
**ESL Form 15**

Table 21a, MDS Fit Statistics for ESL, Averaged Over Matrices.

| Form | Dimensional Solution | Stress | R-squared |
|---|---|---|---|
| 15 | 2 | .27 | .75 |
| | 3 | .21 | .78 |
| | 4 | .18 | .79 |
| | 5 | .16 | .80 |
| | 6 | .14 | .82 |

* shading indicates accepted solution

Table 21b, MDS Fit Statistics for Each ESL Group for the Four-Dimensional Solution.

| Form | Matrix | Stress | R-squared |
|---|---|---|---|
| 15 | 1. English first, Male | .14 | .90 |
| | 2. English first, Female | .15 | .87 |
| | 3. ESL Learned 6-10, Male | .20 | .77 |
| | 4. ESL Learned 6-10, Female | .23 | .68 |
| | 5. ESL Learned after 10, Male | .19 | .79 |
| | 6. ESL Learned after 10, Female | .20 | .77 |

Table 21c, MDS Subject Weights for Each ESL Group for the Four-Dimensional Solution.

| | | | Dimension | | | |
|---|---|---|---|---|---|---|
| **Form** | **Matrix** | **Weirdness** | **1** | **2** | **3** | **4** |
| **15** | **1. English first, Male** | .22 | .76 | .29 | .42 | .26 |
| | **2. English first, Female** | .25 | .65 | .42 | .48 | .22 |
| | **3. ESL Learned 6-10, Male** | .04 | .67 | .38 | .29 | .32 |
| | **4. ESL Learned 6-10, Female** | .42 | .48 | .55 | .10 | .36 |
| | **5. ESL Learned after 10, Male** | .12 | .70 | 33 | .28 | .35 |
| | **6. ESL Learned after 10, Female** | .08 | .71 | .34 | .27 | .28 |
| | **Importance of each dimension:** | | .44 | .15 | .11 | .09 |

## Multi-group WMDS Analysis:
## ESL Form 15

Figure 7a,

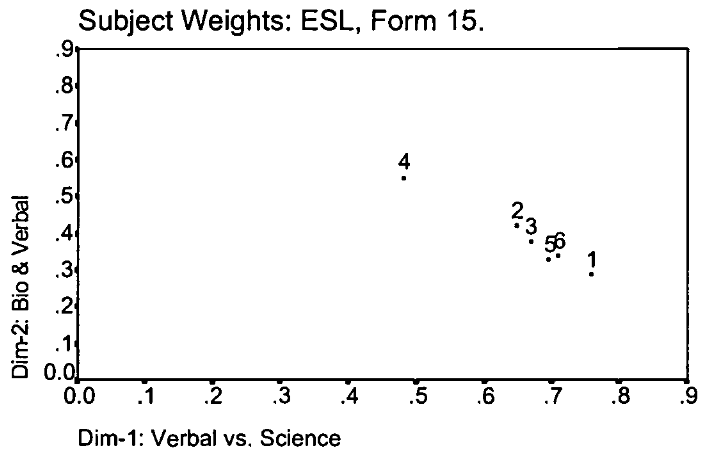



Figure 7b,

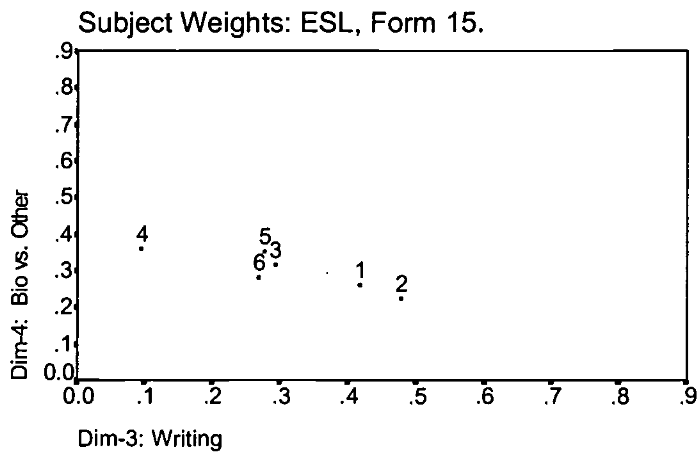
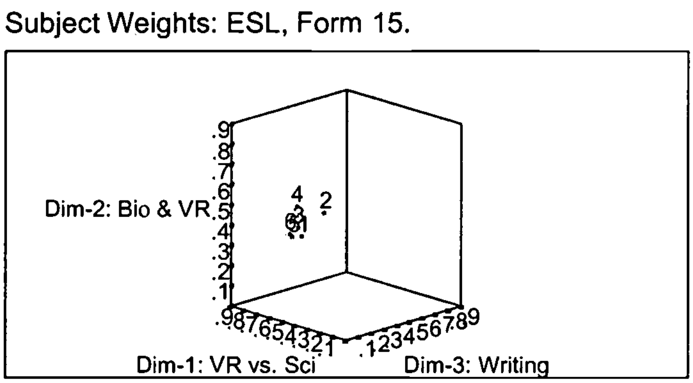



Figure 7c,




1) E 1st, Male 3) L 6-10, Male 5) L >10, Male
2) E 1st, Female 4) L 6-10, Female 6) L >10, Female

**Multi-group WMDS Analysis:**
**ESL Form 15**

Figure 8a,

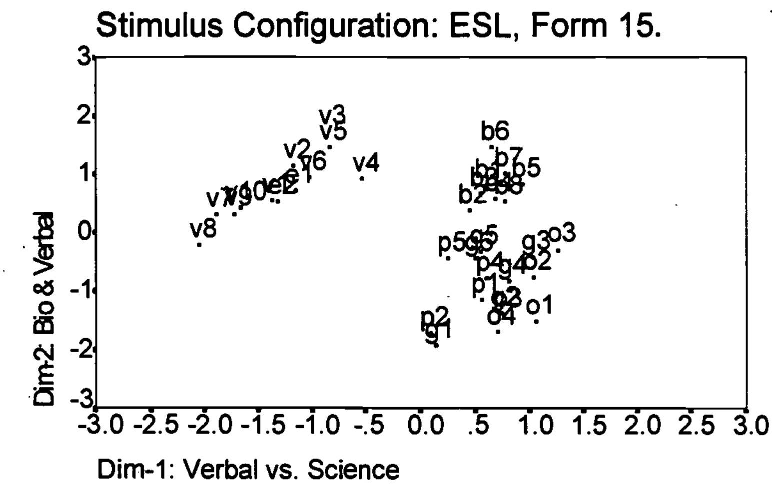



Figure 8b,

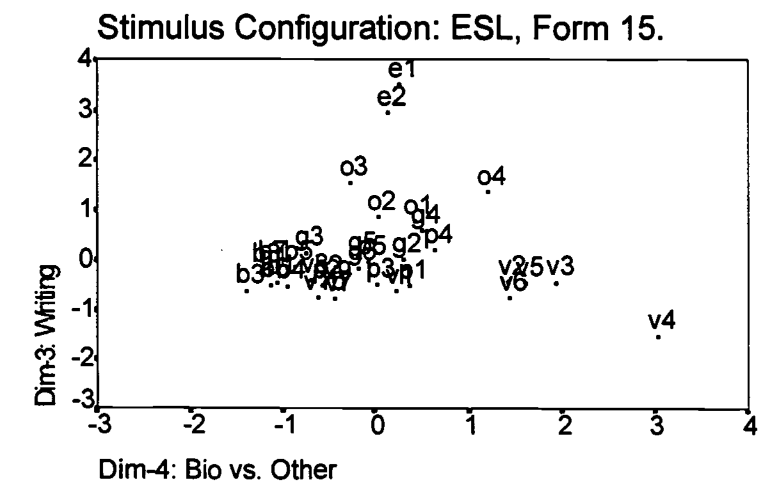
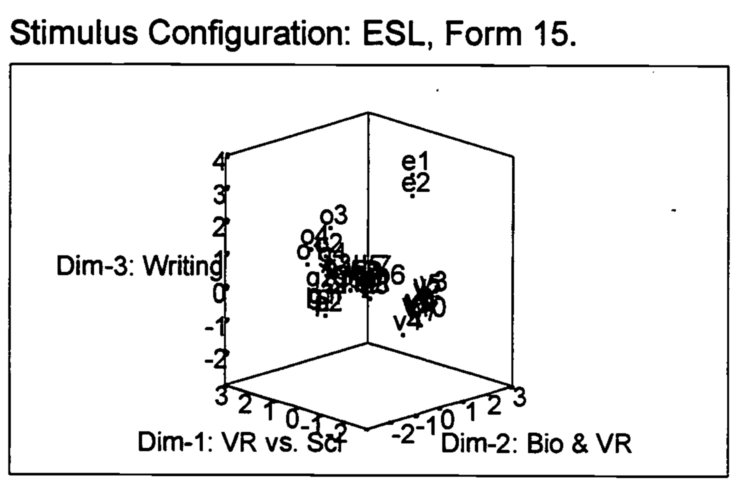



figure 8c,



**Multi-group WMDS Analysis:**
**ESL Form 23**

Table 22a, MDS Fit Statistics for ESL, Averaged Over Matrices.

| Form | Dimensional Solution | Stress | R-squared |
|---|---|---|---|
| 23 | 2 | .26 | .73 |
| | 3 | .21 | .76 |
| | 4 | .18 | .79 |
| | 5 | .16 | .81 |
| | 6 | .14 | .82 |

* shading indicates accepted solution

Table 22b, MDS Fit Statistics for Each ESL Group for the Four-Dimensional Solution.

| Form | Matrix | Stress | R-squared |
|---|---|---|---|
| 23 | 1. English first, Male | .13 | .89 |
| | 2. English first, Female | .15 | .87 |
| | 3. ESL Learned 6-10, Male | .21 | .71 |
| | 4. ESL Learned 6-10, Female | .21 | .72 |
| | 5. ESL Learned after 10, Male | .18 | .79 |
| | 6. ESL Learned after 10, Female | .20 | .75 |

Table 22c, MDS Subject Weights for Each ESL Group for the Four-Dimensional Solution.

| | | | Dimension | | | |
|---|---|---|---|---|---|---|
| **Form** | **Matrix** | **Weirdness** | **1** | **2** | **3** | **4** |
| 23 | 1. English first, Male | .44 | .72 | .23 | .55 | .13 |
| | 2. English first, Female | .28 | .71 | .31 | .47 | .19 |
| | 3. ESL Learned 6-10, Male | .11 | .61 | .40 | .26 | .31 |
| | 4. ESL Learned 6-10, Female | .37 | .49 | .45 | .18 | .49 |
| | 5. ESL Learned after 10, Male | .08 | .69 | 39 | .28 | .29 |
| | 6. ESL Learned after 10, Female | .22 | .61 | .48 | .21 | .34 |
| | Importance of each dimension: | | .42 | .15 | .12 | .10 |

## Multi-group WMDS Analysis:
## ESL Form 23

Figure 9a,

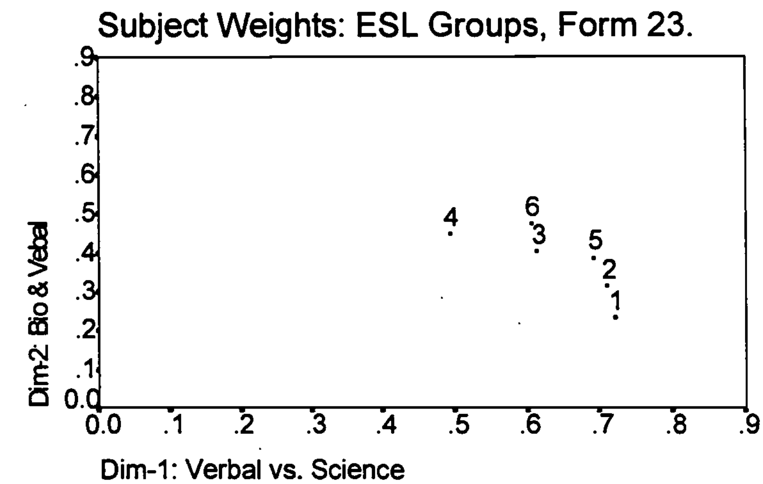



Figure 9b,

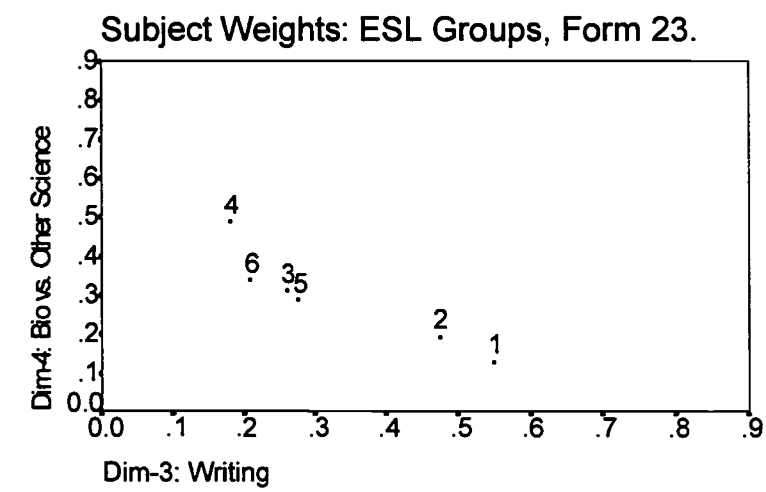
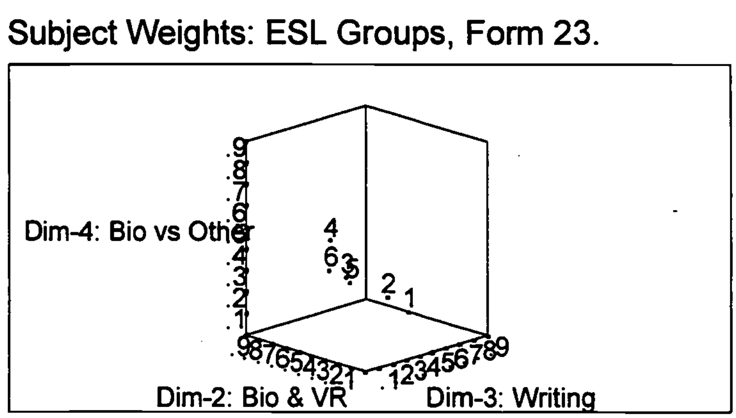



Figure 9c,




1) E 1st, Male 3) L 6-10, Male 5) L >10, Male
2) E 1st, Female 4) L 6-10, Female 6) L >10, Female

## Multi-group WMDS Analysis: ESL Form 23

Figure 10a,

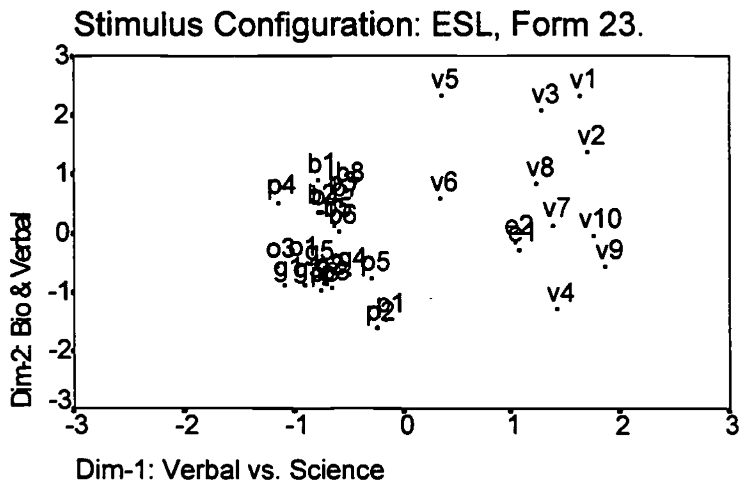



Figure 10b,

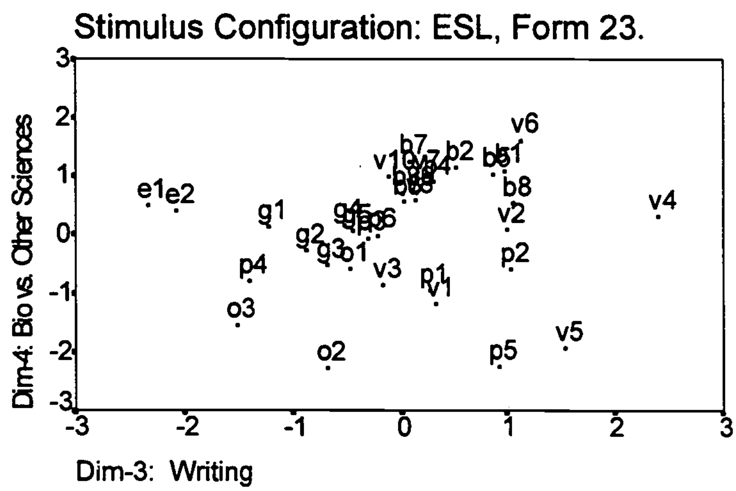
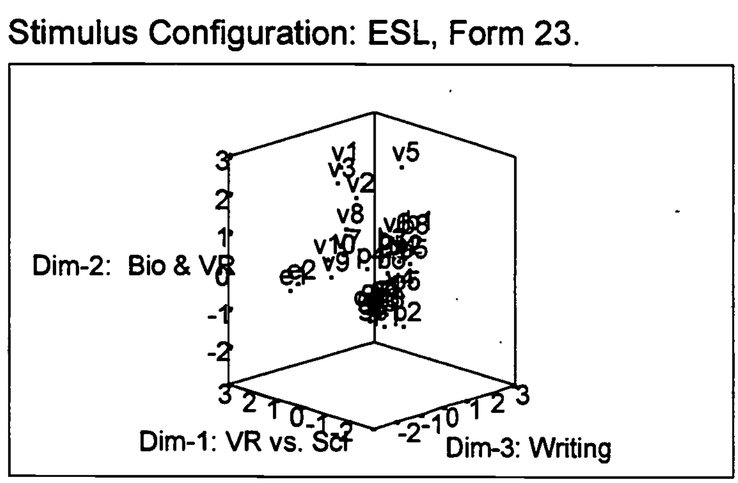



Figure 10c,



Race/ethnicity. Ten matrices were involved in the race/ethnicity analysis. A five-dimensional solution was accepted for both test forms. The STRESS values were .19 for both forms, and the values of $R^2$ were .63 (Form 15) and .60 (Form 23). Two of the dimensions were the familiar verbal and writing skills dimensions. The other three dimensions roughly distinguished the biology, chemistry, and physics items. Thus, the dimensions corresponded to the known content structure of the MCAT. For Form 15, the percentages of variance accounted for by the first through fifth dimensions were 36%, 9%, 7%, 5% and 5% respectively. For Form 23, the percentages of variance in the data accounted for by the dimensions were similar (29%, 9%, 8%, 8% and 7%). Fit statistics and subject weights for both forms can be found in Tables 23a-24c. Some notable differences were observed among the group weights. The weights for both male and female Mexican Americans on the essay dimension were relatively lower in comparison to the other groups (including the Asian and Other Hispanic groups). For both test forms, these two groups had higher weights on the "chemistry" dimension relative to the other groups. The subject weights for the "essay" and "chemistry" dimensions are presented in Figures 13b for From 23. The subject weights for the other three dimensions are portrayed in Figures 13a and 13c. These figures illustrate the relative similarity between Asians and Caucasians in the weighting of the dimensions, and the differences noted for the Mexican Americans. These weight differences suggest that the writing samples account for less variation, and the chemistry parcels account for more variation, in the data for Mexican Americans relative to the other racial/ethnic groups. Figures 11a through 12c show the subject weights and stimulus configurations for the five-dimensional solution for Forms 15.

An important observation noted across all the weighted MDS analyses was that there appeared to be very little variation in dimension weights across the sexes for any of the groups studied. Thus, the structure of the MCAT appears very similar for male and female test takers within each studied group.

## Multi-group WMDS Analysis: Race Form 15

Table 23a, MDS Fit Statistics for Race, Averaged Over Matrices.

| Form | Dimensional Solution | Stress | R-squared |
|---|---|---|---|
| 15 | 2 | .32 | .55 |
| | 3 | .26 | .59 |
| | 4 | .23 | .60 |
| | 5 | .19 | .63 |
| | 6 | .17 | .64 |

* shading indicates accepted solution

Table 23b, MDS Fit Statistics for Each Racial Group for the Five-Dimensional Solution.

| Form | Matrix | Stress | R-squared |
|---|---|---|---|
| 15 | 1. Asian American, Male | .15 | .84 |
| | 2. Asian American, Female | .15 | .85 |
| | 3. African American, Male | .21 | .55 |
| | 4. African American, Female | .21 | .50 |
| | 5. Other Hispanic, (both) | .21 | .57 |
| | 6. Native American, (both) | .23 | .41 |
| | 7. Mexican Am./P.R., Male | .21 | .48 |
| | 8. Mexican Am./P.R., Female | .24 | .38 |
| | 9. Caucasian, Male | .14 | .87 |
| | 10. Caucasian, Female | .15 | .82 |

Table 23c, MDS Subject Weights for Each Racial Group for the Five-Dimensional Solution.

| | | | Dimension | | | | |
|---|---|---|---|---|---|---|---|
| Form | Matrix | Weirdness | 1 | 2 | 3 | 4 | 5 |
| 15 | 1. Asian American, Male | .35 | .79 | .32 | .29 | .10 | .16 |
| | 2. Asian American, Female | .30 | .82 | .25 | .27 | .11 | .15 |
| | 3. African American, Male | .15 | .48 | .29 | .29 | .28 | .28 |
| | 4. African American, Female | .21 | .42 | .32 | .22 | .29 | .30 |
| | 5. Other Hispanic, (both) | .05 | .58 | .29 | .24 | .23 | .19 |
| | 6. Native American, (both) | .16 | .38 | .30 | .24 | .26 | .21 |
| | 7. Mexican Am./P.R., Male | .32 | .41 | .30 | .13 | .31 | .32 |
| | 8. Mexican Am./P.R., Female | .37 | .36 | .29 | .08 | .25 | .31 |
| | 9. Caucasian, Male | .32 | .78 | .31 | .35 | .15 | .10 |
| | 10. Caucasian, Female | .20 | .71 | .32 | .37 | .18 | .18 |
| | Importance of each dimension: | | .36 | .09 | .07 | .05 | .05 |

**Multi-group WMDS Analysis:**
**Race Form 15**

Figure 11a,

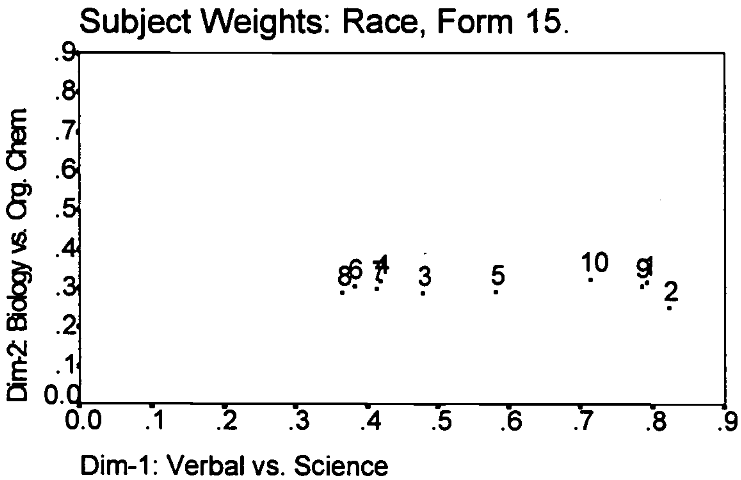



Figure 11b,

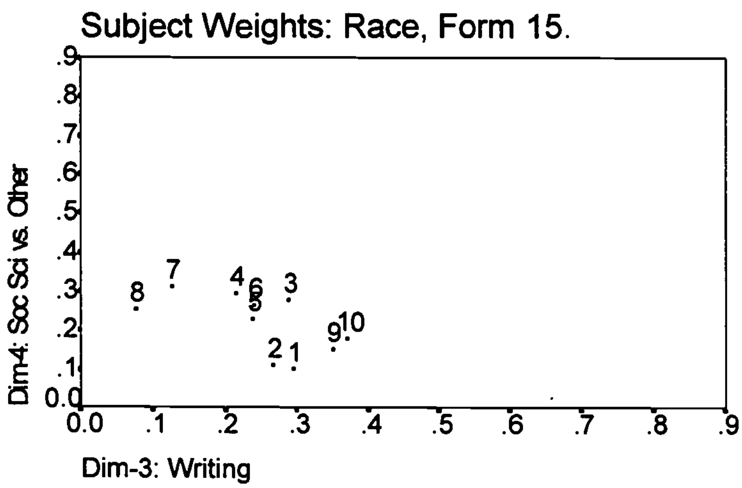
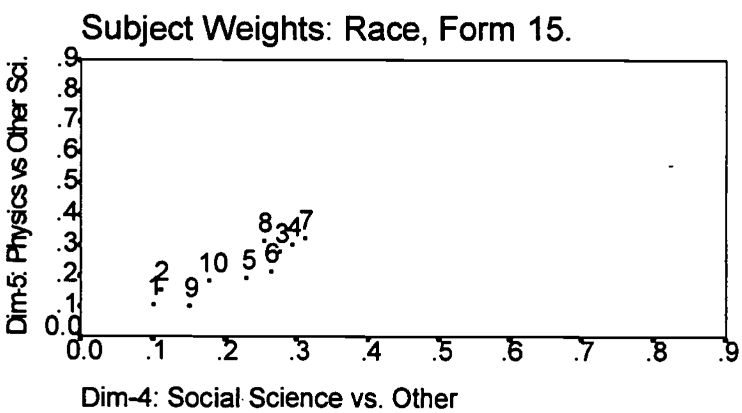



Figure 11c,




1) Asian, M 3) Afric-Am, M 5) Spanish, B 7) Mex-Am,
2) Asian, F 4) Afric-Am, F 6) Native-Am, B 8) Mex-Am,

**Multi-group WMDS Analysis:**
**Race Form 15**

Figure 12a,

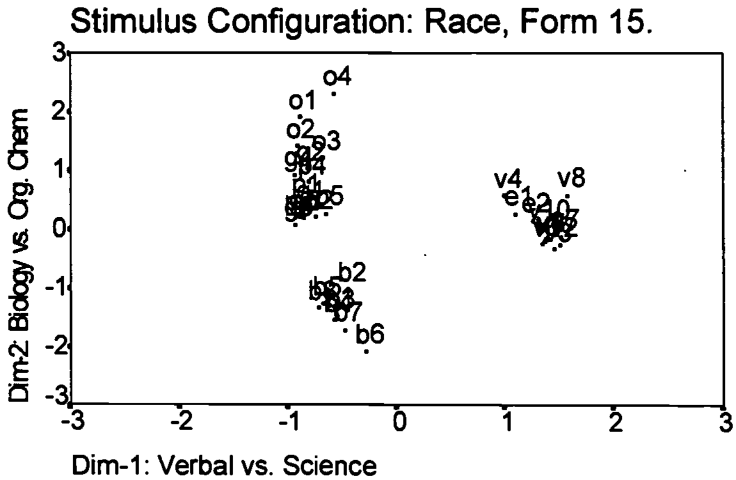



Figure 12b,

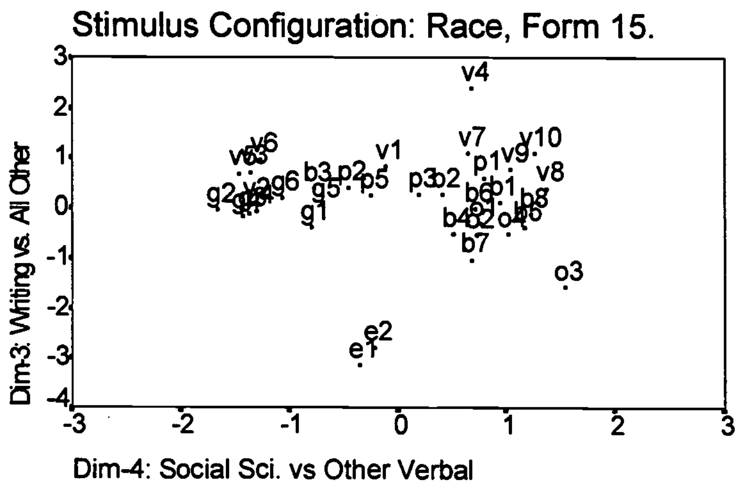
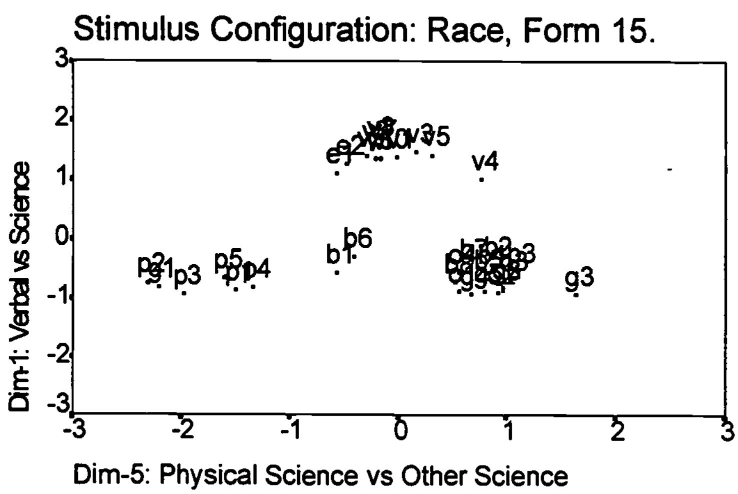



Figure 12c,



**Multi-group WMDS Analysis:**
**Race Form 23**

Table 24a, MDS Fit Statistics for Race, Averaged Over Matrices.

| Form | Dimensional Solution | Stress | R-squared |
|---|---|---|---|
| 23 | 2 | .32 | .51 |
| | 3 | .26 | .56 |
| | 4 | .22 | .57 |
| | 5 | .19 | .60 |
| | 6 | .18 | .61 |

* shading indicates accepted solution

Table 24b, MDS Fit Statistics for Each Racial Group for the Five-Dimensional Solution.

| Form | Matrix | Stress | R-squared |
|---|---|---|---|
| 23 | 1. Asian American, Male | .15 | .82 |
| | 2. Asian American, Female | .15 | .81 |
| | 3. African American, Male | .21 | .52 |
| | 4. African American, Female | .21 | .52 |
| | 5. Other Hispanic, (both) | .21 | .51 |
| | 6. Native American, (both) | .25 | .33 |
| | 7. Mexican Am./P.R., Male | .23 | .43 |
| | 8. Mexican Am./P.R., Female | .22 | .45 |
| | 9. Caucasian, Male | .14 | .82 |
| | 10. Caucasian, Female | .14 | .84 |

Table 24c, MDS Subject Weights for Each Racial Group for the Five-Dimensional Solution.

| | | | Dimension | | | | |
|---|---|---|---|---|---|---|---|
| Form | Matrix | Weirdness | 1 | 2 | 3 | 4 | 5 |
| 23 | 1. Asian American, Male | .28 | .75 | .35 | .21 | .15 | .26 |
| | 2. Asian American, Female | .21 | .73 | .33 | .22 | .20 | .28 |
| | 3. African American, Male | .16 | .38 | .30 | .29 | .35 | .28 |
| | 4. African American, Female | .17 | .40 | .29 | .32 | .35 | .23 |
| | 5. Other Hispanic, (both) | .11 | .44 | .25 | .32 | .25 | .29 |
| | 6. Native American, (both) | .16 | .33 | .23 | .25 | .27 | .17 |
| | 7. Mexican Am./P.R., Male | .29 | .35 | .14 | .30 | .35 | .26 |
| | 8. Mexican Am./P.R., Female | .34 | .34 | .14 | .32 | .40 | .23 |
| | 9. Caucasian, Male | .27 | .68 | .43 | .26 | .14 | .30 |
| | 10. Caucasian, Female | .25 | .72 | .39 | .27 | .15 | .26 |
| | Importance of each dimension: | | .29 | .09 | .08 | .08 | .07 |

**Multi-group WMDS Analysis:**
**Race Form 23**

Figure 13a,

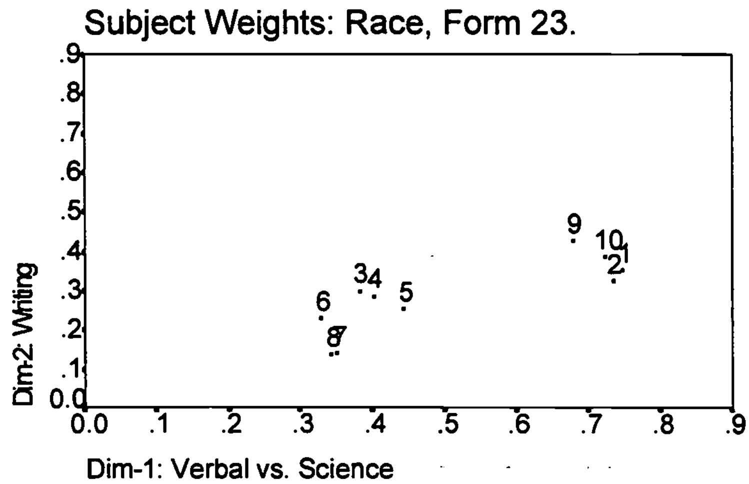



Figure 13b,

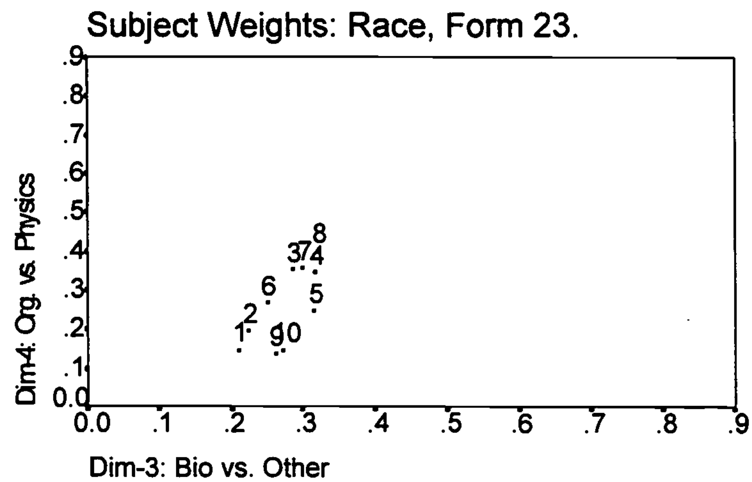
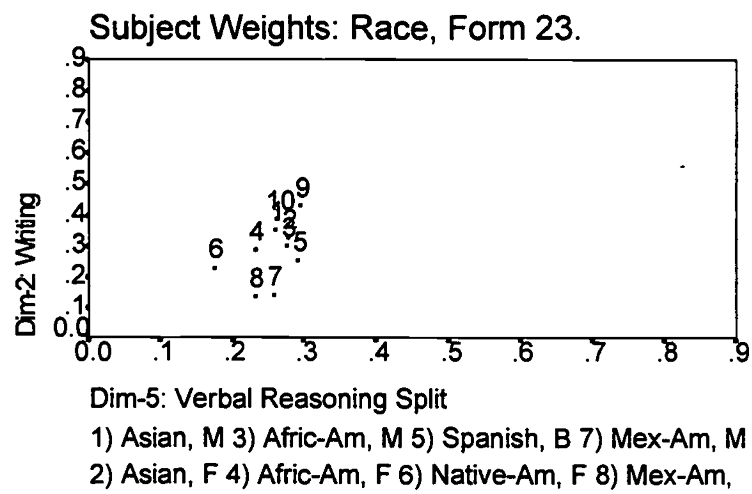



Figure 13c,



Group CFA Results

CFA analyses were carried out using both LISREL 7.2 and 8.0 (Joreskog & Sorbom, 1996). This produced two different sets of results, and thus, two different sets of information about the relationships among groups. Version 8.0 yielded *overall* GFI and RMSR statistics for two hypothesis, B & C. Hypothesis B, tested whether all groups had the same pattern and starting values (*equivalent structure*). Hypothesis C which was slightly more restrictive; tested whether all parameter matrices had the same pattern of fixed and free elements, and that all elements, which were defined as free, were equal across groups (*invariant structure*). Tables 25 and 26 (beginning on the next page) contain the results of the overall CFA hypothesis tests for test forms 15 and 23. LISREL 7.2 yielded a GFI and RMSR for each group (i.e., no overall fit statistics). Tables 27a-27d summarize the individual group CFA results for the following groups for form 15: sex, repeaters, ESL, and race. For all models, the RMSR were below .10. In addition, the CFA results for form 23 are nearly identical to those for form 15. Tables 28a through 28d summarize the individual group CFA results for form 23.

**Multi-group CFA Analysis: Form 15**

Table 25, Overall Group GFI and RMSR Statistics for Hypothesis B&C, for Form 15.

| | | Hypothesis B. | | Hypothesis C. | |
|---|---|---|---|---|---|
| | | **GFI** | **RMSR** | **GFI** | **RMSR** |
| **SEX** | **1** | .82 | .057 | .82 | .058 |
| | **2** | .91 | .038 | .91 | .039 |
| | **3** | .93 | .033 | .93 | .034 |
| | **4** | .95 | .028 | .95 | .028 |
| | **6** | .97 | .022 | .97 | .022 |
| **REPEATER** | **1** | .82 | .061 | .82 | .062 |
| | **2** | .92 | .039 | .92 | .040 |
| | **3** | .94 | .033 | .94 | .034 |
| | **4** | .95 | .029 | .95 | .030 |
| | **6** | .97 | .023 | .97 | .024 |
| **ESL** | **1** | .82 | .055 | .82 | .059 |
| | **2** | .89 | .039 | .89 | .045 |
| | **3** | .91 | .034 | .91 | .039 |
| | **4** | .93 | .031 | .93 | .036 |
| | **6** | .95 | .026 | .94 | .033 |
| **RACE** | **1** | .83 | .059 | .83 | .059 |
| | **2** | .92 | .039 | .92 | .038 |
| | **3** | .94 | .032 | .94 | .032 |
| | **4** | .95 | .028 | .95 | .028 |
| | **6** | .97 | .022 | .97 | .022 |

**Multi-group CFA Analysis: Form 23**

Table 26, Overall Group GFI and RMSR Statistics for Hypothesis B&C, for Form 23.

| | | Hypothesis B. | | Hypothesis C. | |
|---|---|---|---|---|---|
| | | **GFI** | **RMSR** | **GFI** | **RMSR** |
| **SEX** | **1** | .83 | .055 | .82 | .057 |
| | **2** | .92 | .040 | .92 | .042 |
| | **3** | .94 | .034 | .94 | .036 |
| | **4** | .95 | .031 | .95 | .031 |
| | **6** | .97 | .023 | .97 | .024 |
| **REPEATER** | **1** | .83 | .060 | .83 | .060 |
| | **2** | .91 | .044 | .91 | .045 |
| | **3** | .94 | .038 | .94 | .038 |
| | **4** | .95 | .034 | .95 | .034 |
| | **6** | .98 | .024 | .98 | .024 |
| **ESL** | **1** | .80 | .058 | .79 | .066 |
| | **2** | .88 | .044 | .88 | .057 |
| | **3** | .90 | .040 | .90 | .047 |
| | **4** | .92 | .038 | .92 | .044 |
| | **6** | .95 | .029 | .95 | .036 |
| **RACE** | **1** | .85 | .057 | .84 | .058 |
| | **2** | .92 | .043 | .92 | .042 |
| | **3** | .95 | .034 | .95 | .034 |
| | **4** | .96 | .031 | .96 | .032 |
| | **6** | .98 | .024 | .98 | .025 |

Sexes. Given the results of the WMDS analyses, we did not expect to find differences in the factor structures for females and males. Inspection of the overall fit indices for both hypotheses indicate *equivalent* and *invariant* structures across sexes for both exams, for 2-factor and higher models (2-factor GFI/RMSR for exam 15=.91/.039 & exam 23=.92/.041). For exam 15, the fit indices for individual groups, shown in Table 27a, confirm similar fit for both males and females for all models. These results were consistent with the parcel-level 15A and 15B CFA results. A two-factor model fit well (GFI=.94 and RMSR=.033) and all higher models fit slightly better. The two-factor model specified constructs measuring a) verbal reasoning and essay items on factor 1, and b) all science parcels on factor 2. For test form 23, again, there were no differences between males and females. The two-factor model fit well (GFI=.93 and RMSR=.038) and all higher models fit better.

## Multi-group CFA Analysis: Form 15

Table 27a, CFA Fit Indices for Females and Males, Exam 15.

| | # of | Female | | Male | | | |
|---|---|---|---|---|---|---|---|
| **Form** | **Factors** | **GFI** | **RMSR** | **GFI** | **RMSR** | **CHI-SQR** | **DF** |
| **15** | **1** | .85 | .052 | .84 | .053 | 32,778 | 1154 |
| | **2** | .94 | .033 | .94 | .033 | 16,912 | 1151 |
| | **3** | .96 | .028 | .96 | .027 | 10,128 | 1146 |
| | **4** | .98 | .019 | .98 | .022 | 6,481 | 1139 |
| | **6** | .98 | .017 | .98 | .018 | 4,752 | 1119 |
| | **N** | *7,651* | | *8,714* | | | |

## Multi-group CFA Analysis: Form 23

Table 28a, CFA Fit Indices for Females and Males, Exam 23.

| | # of | Female | | Male | | | |
|---|---|---|---|---|---|---|---|
| **Form** | **Factors** | **GFI** | **RMSR** | **GFI** | **RMSR** | **CHI-SQR** | **DF** |
| **23** | **1** | .86 | .050 | .86 | .052 | 23,848 | 1087 |
| | **2** | .92 | .038 | .93 | .038 | 14,749 | 1084 |
| | **3** | .95 | .033 | .95 | .033 | 8,449 | 1079 |
| | **4** | .97 | .028 | .98 | .024 | 5,130 | 1072 |
| | **6** | .98 | .022 | .99 | .020 | 3,371 | 1052 |
| | **N** | *5,952* | | *6,662* | | | |

First-time/repeaters. For first-timers/repeaters analyses, the overall fit indices for hypotheses B and C indicate *equivalent* and *invariant* structures exist across both groups for both exams, for 2-factor and higher models (2-factor GFI/RMSR for form 15 = .92/.040 & form 23 = .91/.045). For form 15, the individual CFA fit indices for repeaters and first-timers, shown in Table 27b, were nearly identical to each other. The two-factor model fit well (GFI=.94, RMSR=.033) and the indices were consistent with the parcel-level 15A and 15B CFA results. For form 23, again, there were no differences in individual CFA fit indices between first-timers and repeaters. Two-factor models fit the data well (GFI=.93, RMSR=.039) and all higher models fit slightly better.

Table 27b, CFA Fit Indices for First-timers and Repeaters, Exam 15.

| | # of | First-timers | | Repeaters | | | |
|---|---|---|---|---|---|---|---|
| **Form** | **Factors** | **GFI** | **RMSR** | **GFI** | **RMSR** | **CHI-SQR** | **DF** |
| **15** | **1** | .84 | .052 | .85 | .056 | 34,156 | 1154 |
| | **2** | .94 | .031 | .94 | .035 | 16,945 | 1151 |
| | **3** | .96 | .025 | .96 | .029 | 10,107 | 1146 |
| | **4** | .98 | .020 | .98 | .021 | 6,433 | 1139 |
| | **6** | .98 | .017 | .98 | .019 | 4,702 | 1119 |
| | **N** | *8,820* | | *7,661* | | | |

Table 28b, CFA Fit Indices for First-timers and Repeaters, Exam 23.

| Form | # of Factors | First-timers GFI | First-timers RMSR | Repeaters GFI | Repeaters RMSR | CHI-SQR | DF |
|---|---|---|---|---|---|---|---|
| 23 | 1 | .85 | .050 | .86 | .055 | 24,734 | 1087 |
| | 2 | .93 | .037 | .93 | .040 | 14,970 | 1084 |
| | 3 | .95 | .032 | .95 | .035 | 8,664 | 1079 |
| | 4 | .97 | .024 | .97 | .027 | 5,204 | 1072 |
| | 6 | .98 | .020 | .98 | .023 | 3,465 | 1052 |
| | N | *7,297* | | *5,320* | | | |

Native English/ESL. The overall fit indices for ESL were a bit surprising. Unlike the above analyses, the overall fit indices did not indicate *equivalent* and *invariant* structures across groups for the 2-factor model (2-factor GFI/RMSR for form 15 = .89/.042 & form 23 = .88/.051). Good fit was obtained for 3-factor and higher models (3-factor GFI/RMSR for form 15 = .91/.037 & form 23 = .90/.044). Looking at the individual CFA results, shown in Table 27c, we see that in all cases, fit is slightly worse for those who learned English between ages 6-10 (GFI/RMSR for 2-factor model =.92/.042) and after age 10 (GFI/RMSR for 2-factor model =.92/.039), compared to those for whom English is their first language (GFI/RMSR for 2-factor model =.94/.032). In contrast to the overall results, in all cases the GFI is greater than .90 for all three groups in two-factor and higher models. For form 23, results were similar. Individual data-model fit indices for ESL groups were poorer (GFI/RMSR for 2-factor model in each group = .90/.052) than model-data fit for native English speakers (GFI/RMSR for 2-factor model = .93/.037). Again, in all cases, model fit for those in the "learned English between ages 6-10" group is more similar to those who learned English after age10, than it is to those for whom English is a first language. This suggests that the factor structure is similar for these two ESL groups, but possibly different from the English as a first language group. Consequently, this may have implications for categorizing examinees as ESL or English first.

Table 27c, CFA Fit Indices for ESL Groups, Exam 15.

| Form | # of Factors | English First GFI | English First RMSR | Learned 6-10 GFI | Learned 6-10 RMSR | Learned after 10 GFI | Learned after 10 RMSR | CHI-SQR | DF |
|---|---|---|---|---|---|---|---|---|---|
| 15 | 1 | .88 | .047 | .86 | .056 | .85 | .053 | 29,623 | 1748 |
| | 2 | .94 | .032 | .92 | .042 | .92 | .039 | 17,139 | 1743 |
| | 3 | .96 | .024 | .94 | .031 | .94 | .032 | 10,310 | 1735 |
| | 4 | .98 | .019 | .96 | .028 | .96 | .029 | 7,110 | 1724 |
| | 6 | .99 | .017 | .96 | .025 | .96 | .027 | 5,413 | 1693 |
| | N | *13,337* | | *1,578* | | *1,477* | | | |

Table 28c, CFA Fit Indices for ESL Groups, Exam 23.

| Form | # of Factors | English First GFI | English First RMSR | Learned 6-10 GFI | Learned 6-10 RMSR | Learned after 10 GFI | Learned after 10 RMSR | CHI-SQR | DF |
|---|---|---|---|---|---|---|---|---|---|
| **23** | **1** | .88 | .048 | .84 | .065 | .84 | .057 | 22,508 | 1647 |
| | **2** | .93 | .037 | .90 | .054 | .90 | .049 | 14,878 | 1642 |
| | **3** | .96 | .030 | .93 | .044 | .93 | .042 | 8,545 | 1633 |
| | **4** | .98 | .023 | .95 | .039 | .95 | .037 | 5,545 | 1623 |
| | **6** | .99 | .019 | .96 | .036 | .96 | .035 | 3,942 | 1592 |
| | **N** | *10,245* | | *1,163* | | *1,088* | | | |

Race/ethnicity. The overall CFA fit statistics for racial/ethnic groups yielded *equivalent* and *invariant* structures across all groups for both test forms, for 2-factor and higher models (2-factor GFI/RMSR for form 15 = .92/.039 & form 23 = .92/.043). The individual CFA results for form 15 are shown in Table 27d. Due to the small sample size, the results for Native Americans were not stable and will not be discussed. For the two-factor model there was good fit for Asians, African Americans, and Caucasians (GFI=.93, .92, and .92, respectively). Across all models, the Asian Americans and Caucasians had the highest and most similar fit indices, with African Americans fitting nearly as well. Mexican Americans and Puerto Ricans had good model-data fit for the three-factor model (GFI=.92), although their fit statistics were consistently lower than those of Asians, Caucasians and African Americans. Spanish and South Americans (the smallest sample size) had the poorest fit. For this group, only the six-factor model fit the data well (GFI=.90). For form 23, again, Asians and Caucasians had the highest fit indices followed by African Americans, and then by Mexican Americans and Puerto Ricans. The model-data fit for Spanish and South Americans was generally poor (GFI for the 6-factor model = .89).

Table 27d, CFA Fit Indices for Racial Groups, Exam 15.

| Form | No. of Factors | Asian Am. GFI | African Am. GFI | Spanish/ South Am. GFI | Native Am. GFI | Mexican Am/P.R. GFI | White GFI | CHI-SQR | DF |
|---|---|---|---|---|---|---|---|---|---|
| **15** | **1** | .82 | .88 | .79 | .67 | .84 | .86 | 33,275 | 3530 |
| | **2** | .93 | .92 | .85 | .72 | .88 | .92 | 23,015 | 3519 |
| | **3** | .96 | .95 | .88 | .73 | .92 | .96 | 12,249 | 3502 |
| | **4** | .97 | .96 | .89 | .74 | .94 | .98 | 9,104 | 3479 |
| | **6** | .98 | .96 | .90 | .75 | .95 | .98 | 7,396 | 3415 |
| | **N** | *3,844* | *1,393* | *373* | *123* | *967* | *8,880* | | |

Table 28d, CFA Fit Indices for Racial Groups, Exam 23.

| Form | No. of Factors | Asian Am. GFI | African Am. GFI | Spanish/ South Am. GFI | Native Am. GFI | Mexican Am/P.R. GFI | White GFI | CHI-SQR | DF |
|---|---|---|---|---|---|---|---|---|---|
| 23 | 1 | .84 | .86 | .82 | .64 | .84 | .87 | 24,174 | 3327 |
| | 2 | .92 | .90 | .85 | .67 | .90 | .93 | 15,956 | 3316 |
| | 3 | .94 | .92 | .87 | .68 | .91 | .96 | 10,141 | 3299 |
| | 4 | .97 | .95 | .88 | .69 | .93 | .98 | 7,232 | 3276 |
| | 6 | .97 | .95 | .89 | .71 | .94 | .98 | 5,709 | 3212 |
| | N | *2,853* | *1,085* | *287* | *85* | *909* | *6,711* | | |

In summary, the multi-group CFA analyses supported the structural equivalence of the MCAT across the selected groups studied. Perhaps the only exceptions were in cases involving examinees for whom English is not their first language. For ESL groups, the 2-factor model did not appear adequate to account for their data. This is consistent with some of the WMDS findings. Certain groups tended to de-emphasize the writing dimensions and place more weight on science dimensions. For these groups, however, a 3-factor model which separated writing from verbal reasoning, appeared to adequately fit the data. A cross tabulation of ESL by Race reveals that approximately 40% of those in the Other Hispanic group and 60% of those in the Mexican American or Puerto Rican group learned English after age 5 (Table 29 and 30). These results, perhaps, have implications for how examinees are classified as ESL or Native English Speakers.

Table 29, Crosstab of ESL by Race for Exam 15.

| Race | English First | English 6 to 10 | English after 10 | Row Total |
|---|---|---|---|---|
| **Asian American** | 2,311 | 765 | 752 | 3,828 |
| | *60%* | *20%* | *20%* | *25* |
| **African American** | 1,255 | 57 | 68 | 1,380 |
| | *91%* | *4%* | *5%* | *9%* |
| **Other Hispanic** | 231 | 68 | 73 | 372 |
| | *62%* | *18%* | *20%* | *2%* |
| **Native American** | 118 | 1 | 0 | 119 |
| | *99%* | *1%* | *0%* | *1%* |
| **Mexican American** | 437 | 400 | 121 | 958 |
| | *45%* | *42%* | *13%* | *6%* |
| **Caucasian** | 8,254 | 215 | 382 | 8,851 |
| | *94%* | *2%* | *4%* | *57%* |
| **Column Total** | 12,606 | 1,506 | 1,396 | 15,508 |
| | *81%* | *10%* | 9% | *100%* |

Table 30, Crosstab of ESL by Race for Exam 23.

| Race | English First | English 6 to 10 | English after 10 | Row Total |
|---|---|---|---|---|
| **Asian American** | 1,869 | 460 | 493 | 2,822 |
| | *66%* | *16%* | *18%* | *24%* |
| **African American** | 966 | 34 | 70 | 1,070 |
| | *90%* | *3%* | *7%* | *9%* |
| **Other Hispanic** | 168 | 52 | 66 | 286 |
| | *59%* | *18%* | *23%* | *2%* |
| **Native American** | 83 | 1 | 1 | 85 |
| | *98%* | *1%* | *1%* | *1%* |
| **Mexican American** | 360 | 394 | 134 | 888 |
| | *41%* | *44%* | *15%* | *8%* |
| **Caucasian** | 6,242 | 157 | 285 | 6,684 |
| | *94%* | *2%* | *4%* | *56%* |
| **Column Total** | 9,688 | 1,098 | 1,049 | 11,835 |
| | *82%* | *9%* | 9% | *100%* |

## Discussion

This study reported the results of several different analyses conducted on data from recent administrations of the MCAT. The purposes of these analyses were to better understand the structure of these data, compare the observed structure to the content structure specified in the MCAT blueprints, and evaluate the similarity of the structure across selected groups of MCAT examinees.

Several important pieces of information were learned through these analyses. First, the results suggest that appraisals of the MCAT structure should be conducted at the parcel-level rather than at the item level. Consistent with the literature (e.g., Cattell, 1956; Dorans & Lawrence, 1987, 1991; Green, 1983), the item-level analyses uncovered numerous uninterpretable factors that were most likely due to random error. Furthermore, different results were obtained when replicating over the different item orderings of each test form, and over the different test forms. In contrast, the parcel-level analyses were readily interpretable and the results were consistent across replications.

The parcel-level results suggest a dominant factor underlies the MCAT. The first factor resulting from the PCAs accounted for over 30% of the variation among the item parcels. Given the diverse knowledge and skills measured, this dominant factor is probably a "general intelligence" factor. The results also suggest additional factors that represent the principle disciplines measured on the MCAT. Looking beyond the general factor, the next structural layer of the MCAT separates test material measuring science from test material measuring verbal reasoning and writing skills. The next structural level depicts three factors: science, verbal reasoning, and writing skill. These three factors were supported by all analyses (i.e., PCA, MDS, CFA) and were replicated across all test forms.

The results also support the distinctions among the science disciplines specified in the MCAT blueprint. The PCA results illustrated patterns of factor loadings that segregated one or more of the science disciplines, and the MDS results revealed clusters of items that distinguished among the biology, physics, and chemistry-related items. The specific pattern of factor loadings and clusters was not consistent across test forms, which suggests these disciplines are closely related. Thus, these disciplines are probably best thought of as separate facets of an unidimensional science proficiency construct. The CFA results supported factor models specifying unique factors for each of the science disciplines, but these models were less parsimonious than the three-factor model that displayed adequate fit to the data.

In general, the analyses support the current content structure of the MCAT reported in the test blueprint. However, from a *statistical* perspective, the results suggest it may be possible to scale the biological sciences and physical sciences items along a single continuum, rather than along two separate scales. These results are congruent with the item response theory (IRT) analyses conducted recently on these data (AIR, 1998), which showed that if separate IRT proficiency estimates were derived for each discipline, MCAT test takers would be rank-ordered similarly across disciplines. The results also support the viability of re-arranging the discipline areas across the two science test sections. In particular, the results suggest that the general chemistry and organic chemistry items could be included on the same test section. Item parcels from these two disciplines tended to have similar patterns of factor loadings and dimension coordinates.

The statistical similarity of items representing the science disciplines is an important finding to be borne in mind as the AAMC considers changes in the content structure of the MCAT and the possibility of computerized-adaptive testing. However, the most parsimonious structure of the MCAT from a statistical perspective may not be the "best" structure from other perspectives. For example, the specification of separate science disciplines may foster construct representation, assist students in preparing for the exam, and promote the development of higher quality items than if a general science proficiency construct were specified.

With respect to the consistency of the MCAT structure across selected groups of test takers, the results supported the hypothesis of structural invariance across groups. In general, the MDS analyses indicated that all dimensions were relevant for accounting for the variation in the data for each group. However, in two situations, notable differences among the dimensions that were most "important" for one or two groups were observed. Females who learned English between 6 and 10 years old, and male and female Mexican Americans, had relatively lower weights on the "essay" dimension in comparison to the other groups. For these groups, the dimensions related to one or more of the science disciplines accounted for relatively more variation in their data than did the verbal reasoning or essay dimensions. These results are interesting and should be followed up to see if these differences are related to differences in predictive validity for these groups. However, with respect to the general structure of the MCAT, these differences appear to be minor, and the overall impression provided by the MDS results is that the structure is roughly equivalent across groups.

The multi-group MDS analyses, which employed separate correlations matrices for each group, required more dimensions to fit the data than did the global MDS analyses. This was expected based on previous research (e.g., Sireci, 1998) because minor variation specific to any one matrix will increase the dimensionality of the solution. The interesting result of this study is that the increased dimensionality was directly related to the MCAT structure specified in the test blueprint. That is, the minor differences among the group weights reflected minor differences in the weightings of the discipline areas specified in the test blueprint. It is also interesting to note that when separate matrices were derived for males and females belonging to a specific racial/ethnic group, the dimension weights were very similar across men and women (with the exception noted above for females who learned English between 6 and 10 years old).

The multi-group CFA analyses also supported the structural invariance of the MCAT across the selected groups of test takers. In general, the three-factor solution that specified equivalent patterns of factor loadings across all groups displayed adequate fit to the data. These results confirm the interpretation that the differences among the group weights in the MDS solution were minor. Thus, the exploratory multi-group MDS results, and the confirmatory multi-group CFA results point to the same conclusion of factor structure equivalence across groups.

## References

Brennan, R. L. (1998). Misconceptions at the intersection of measurement theory and practice. Educational Measurement: Issues and Practice, 17(1), 5-9, 30.

Carroll, J. D., & Chang, J. J. (1970) Analysis of individual differences in multidimensional scaling via an N-way generalization of Eckart-Young decomposition. Psychometrika, 35, 283-319.

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

Cattell, R. B., & Burdsal, C. A., Jr. (1975). The radial parcel double factoring design: A solution to the item-vs.-parcel controversy. Multivariate Behavioral Research, 10, 165-179.

Davison, M. L., (1985). Multidimensional scaling versus components analysis of test intercorrelations. Psychological Bulletin, 97, 94-105.

Davison, M. L., & Sireci, S. G. (in press, 1999). Multidimensional scaling. In H.E.A. Tinsley & S. Brown (Eds.) Handbook of multivariate statistics and mathematical modeling. Washington, DC: American Psychological Association.

Day, S. X., & Rounds, J. (1998). Universality of vocational interest structure among racial and ethnic minorities. American Psychologist, 53, 728-736.

Day, S. X., Rounds, J., & Swaney, K. (1998). The structure of vocational interests for diverse racial-ethnic groups. Psychological Science, 9, 40-44.

Dorans, N. J., & Lawrence, I. M. (1987). The internal construct validity of the SAT (Research Report 87-35). Princeton, NJ: Educational Testing Service.

Green, S. B. (1983). Identifiability of spurious factors with linear factor analysis with binary items. Applied Psychological Measurement, 7, 3-13.

Joreskog & Sorbom (1989). LISREL-7.2 [computer program]. Chicago, IL: Scientific Software.

Joreskog & Sorbom (1989). LISREL 7: A guide to the program and applications. 2nd edition, Chicago, IL: Scientific Software.

Joreskog & Sorbom (1993). LISREL-8 [computer program]. Mooresville, IN: Scientific Software.

Joreskog & Sorbom (1996). LISREL 8: Structural equation modeling with the SIMPLIS command language. Mooresville, IN: Scientific Software.

Marsh, H. W. (1994). Confirmatory factor analysis models of factorial invariance: A multifaceted approach. Structural Equation Modeling, 1, 5-34.

Ramsay, J. O. (1982). Some statistical approaches to multidimensional scaling data. Journal of the Royal Statistical Society, 145, 285-312.

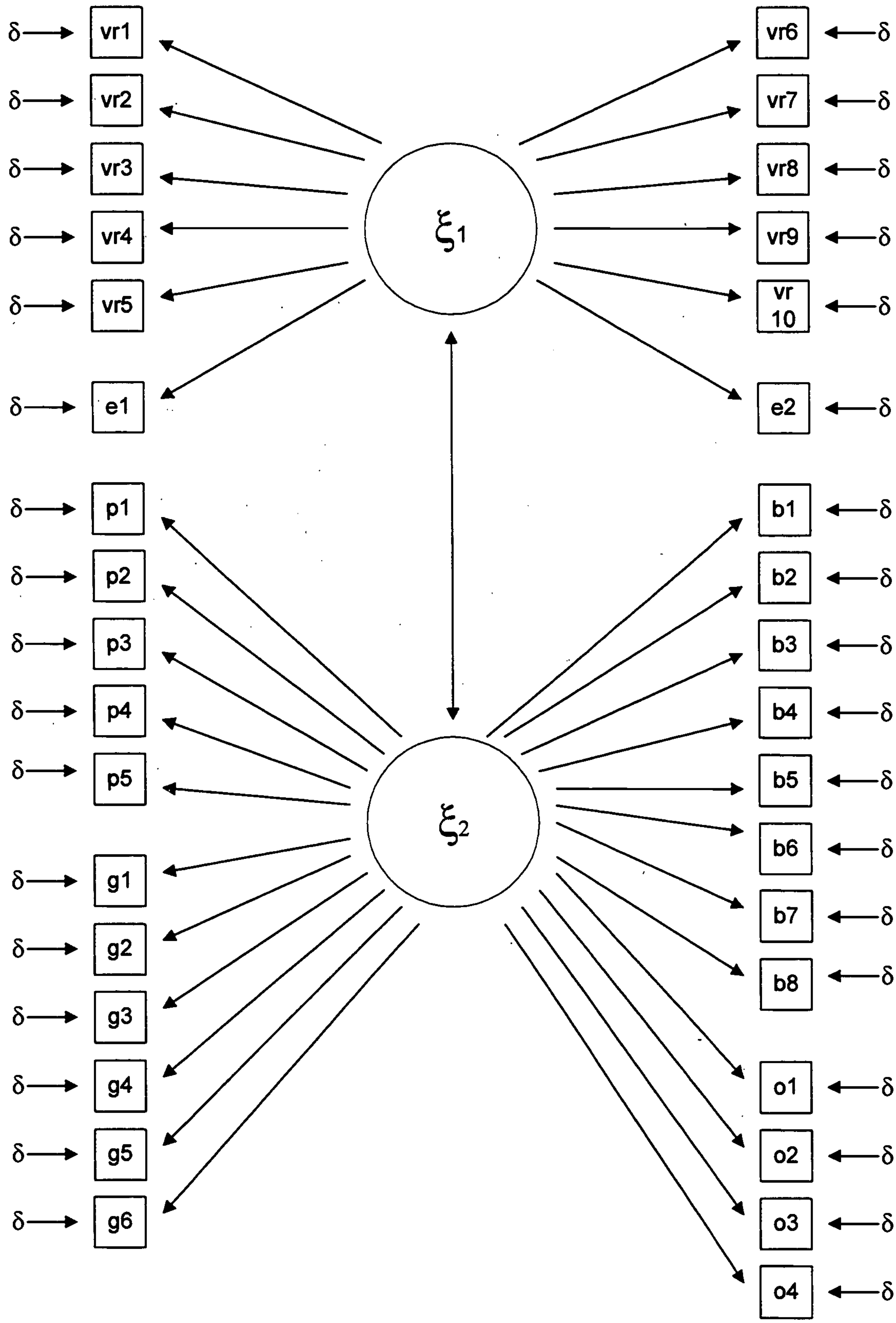
Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. Psychological Bulletin, 114, 552-566.

Robie, C. & Ryan, A. M. (1996). Structural equivalence of a measure of cross-cultural adjustment. Educational and Psychological Measurement, 56, 514-521.

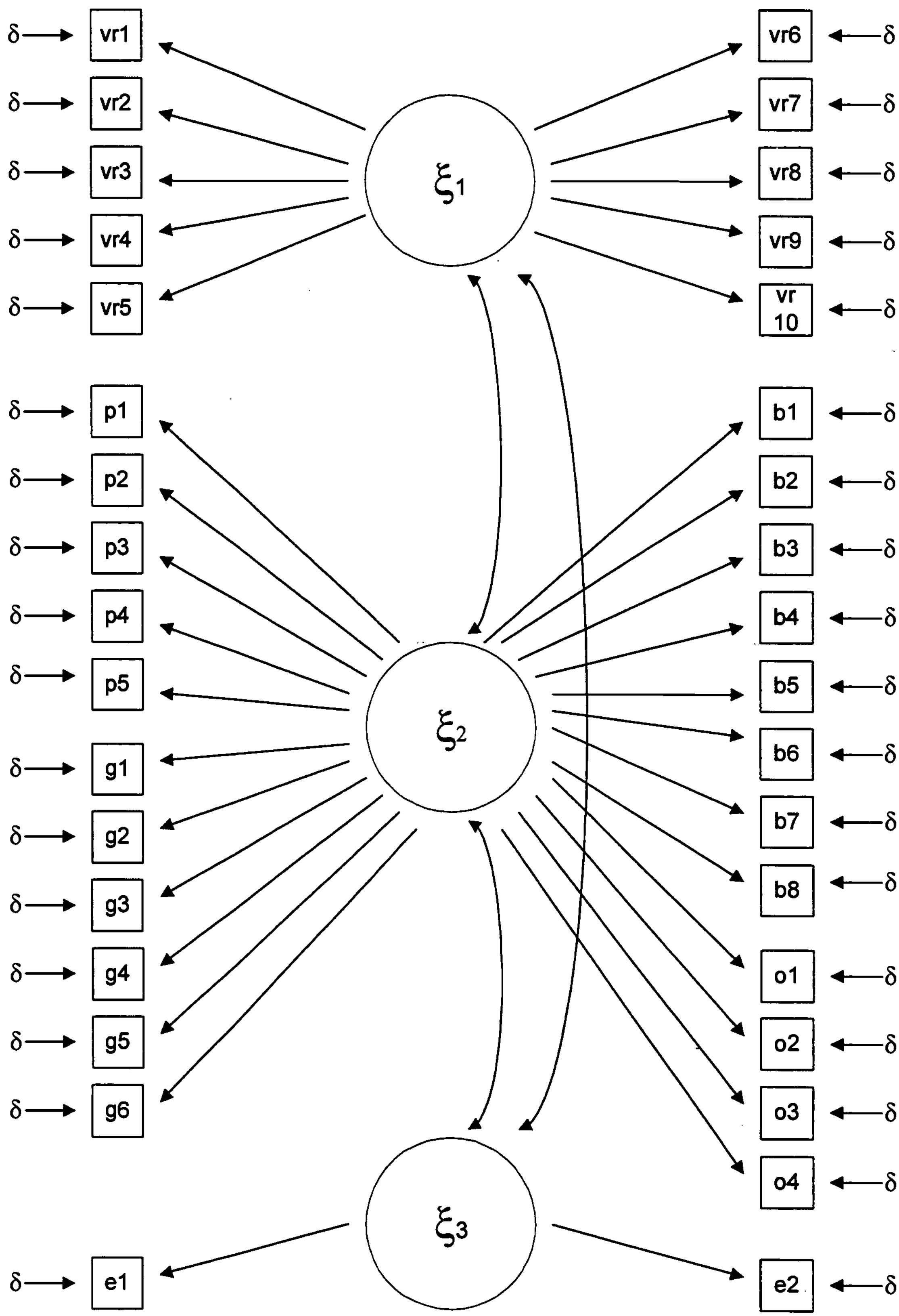
Sireci, S. G. (1998, April). Evaluating content validity using multidimensional scaling. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Sireci, S.G., Bastari, B., & Allalouf, A. (1998, August). Evaluating construct equivalence across adapted tests. Paper presented at the annual meeting of the American Psychological Association, San Francisco, CA.

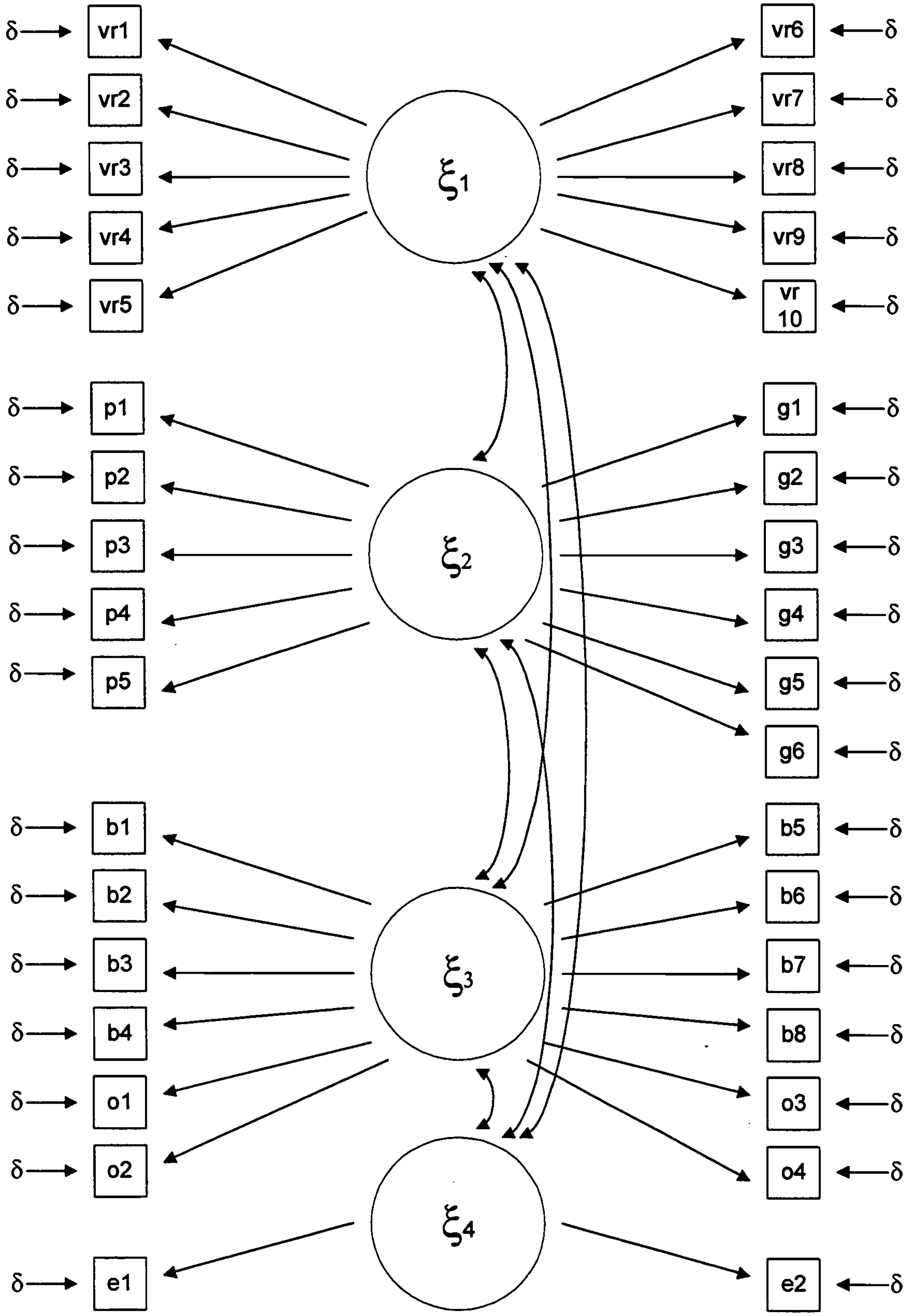
**Apendix A: 2-Factor CFA Model**



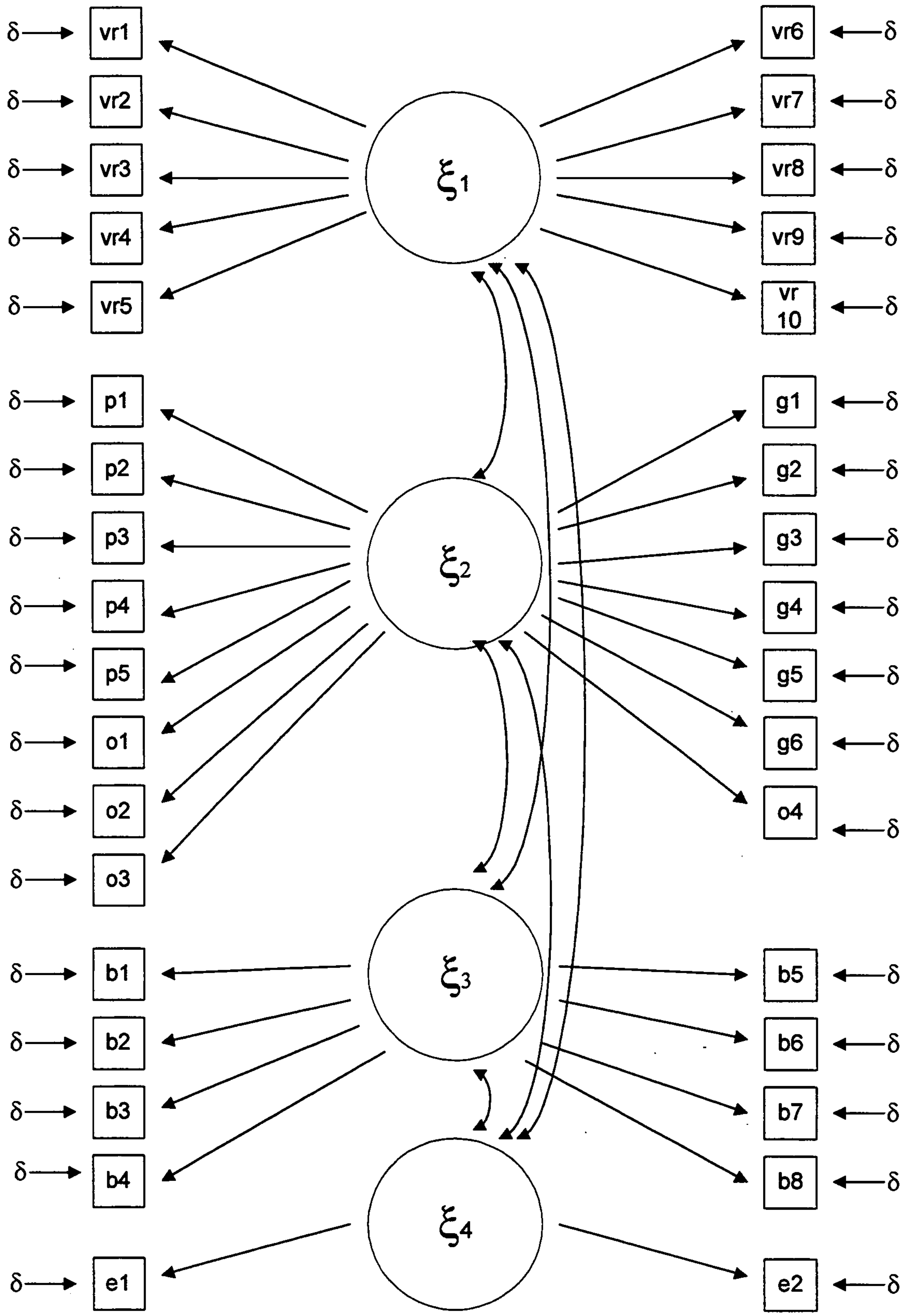
## Apendix A: 3-Factor CFA Model



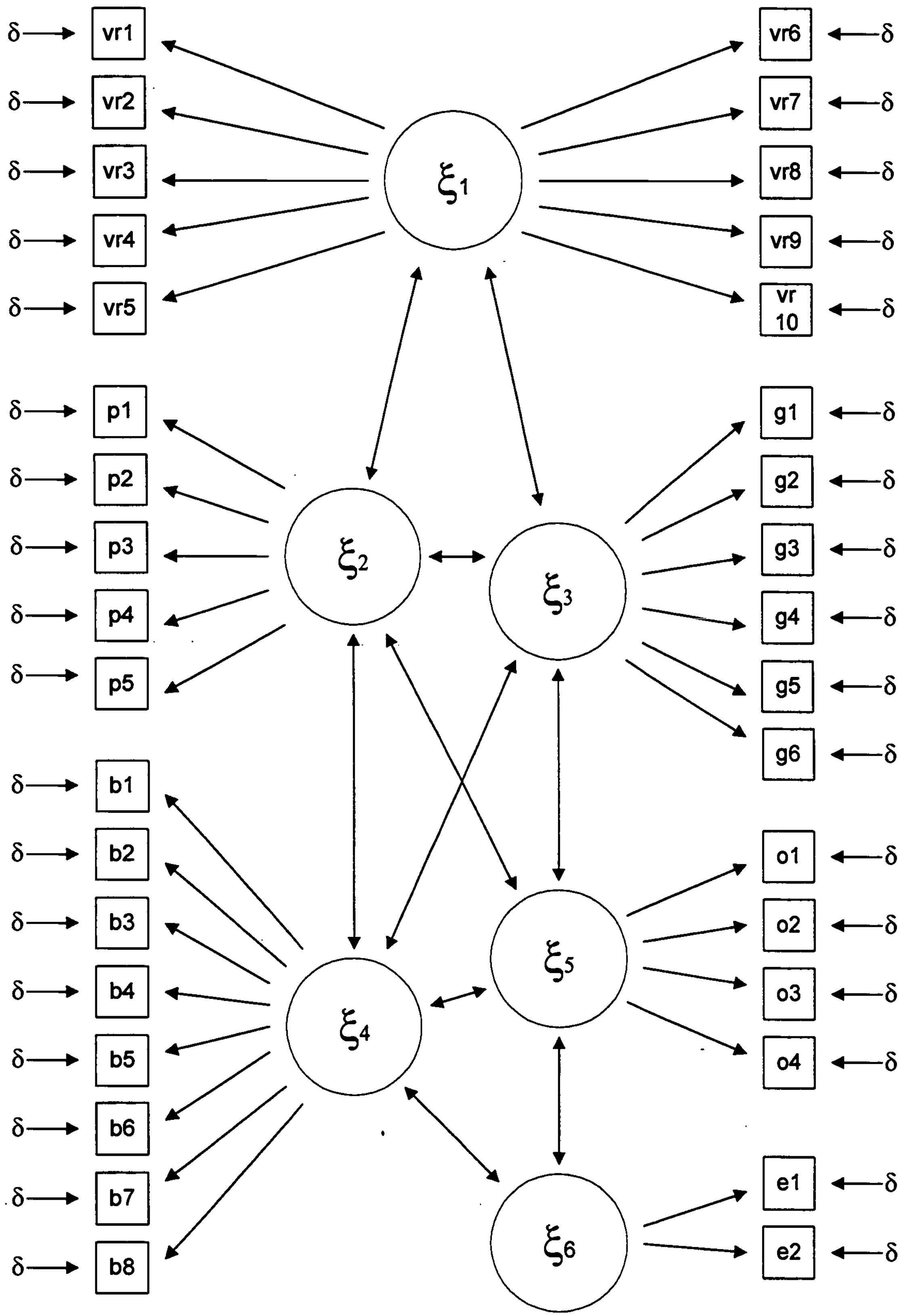
**Apendix A: 4a-Factor CFA Model**



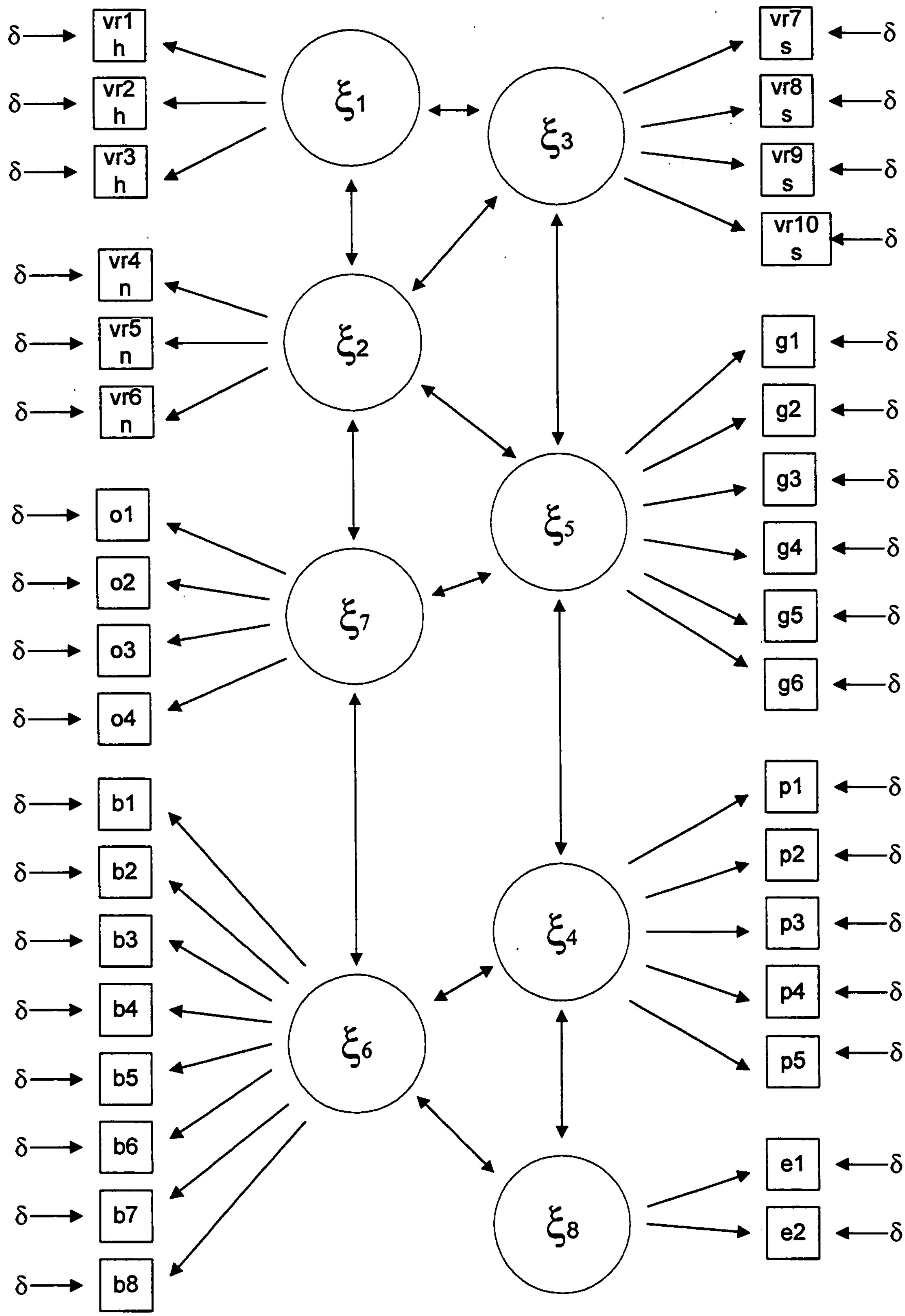
**Apendix A: $4_b$-Factor CFA Model**



**Apendix A: 6-Factor CFA Model**



## Apendix A: 8-Factor CFA Model



## Apendix A: Parceling Scheme by Passage/Discipline, Form 15

The following equations show which items were used to create the parcels.

**Verbal Reasoning:**

VR01ssc = sVRa01 + sVRa04 + sVRa05 + sVRa07 + sVRa10 .
VR02ssc = sVRa02 + sVRa03 + sVRa06 + sVRa08 + sVRa09 .
VR03hum = sVRa11 + sVRa12 + sVRa13 + sVRa14 + sVRa15 + sVRa16 .
VR04nst = sVRa17 + sVRa18 + sVRa19 + sVRa20 + sVRa21 + sVRa22 + sVRa23 .
VR05ssc = sVRa24 + sVRa25 + sVRa26 + sVRa27 + sVRa28 + sVRa29 .
VR06ssc = sVRa30 + sVRa31 + sVRa32 + sVRa33 + sVRa34 + sVRa35 .
VR07hum = sVRa36 + sVRa37 + sVRa38 + sVRa39 + sVRa40 + sVRa41 .
VR08hum = sVRa42 + sVRa43 + sVRa44 + sVRa45 + sVRa46 + sVRa47 .
VR09nst = sVRa48 + sVRa49 + sVRa50 + sVRa51 + sVRa52 + sVRa53 + sVRa54 + sVRa55 .

**Physics:**

phy01 = psa07 + psa08 + psa09 + psa10 + psa11 + psa12 + psa13 .
phy02 = psa14 + psa15 + psa16 + psa17 + psa18 + psa24 .
phy03 = psa26 + psa40 + psa41 + psa43 + psa59 + psa63 .
phy04 = psa47 + psa48 + psa49 + psa50 + psa51 + psa52 .
phy05 = psa53 + psa54 + psa55 + psa56 + psa57 .

**General Chemistry:**

gch01 = psa01 + psa02 + psa03 + psa04 + psa05 + psa06 .
gch02 = psa19 + psa20 + psa21 + psa22 + psa23 + psa25 .
gch03 = psa28 + psa29 + psa30 + psa31 + psa32 + psa33 + psa27 .
gch04 = psa34 + psa35 + psa36 + psa37 + psa38 + psa39 + psa62 .
gch05 = psa42 + psa44 + psa45 + psa46 + psa58 + psa60 + psa61 .

**Biology:**

blg01 = sbsa01 + sbsa02 + sbsa03 + sbsa04 + sbsa05 + sbsa06 .
blg02 = sbsa13 + sbsa14 + sbsa16 + sbsa17 + sbsa23 + sbsa24 .
blg03 = sbsa18 + sbsa19 + sbsa20 + sbsa21 + sbsa22 .
blg04 = sbsa27 + sbsa28 + sbsa29 + sbsa30 + sbsa31 + sbsa32 + sbsa33 .
blg05 = sbsa40 + sbsa41 + sbsa42 + sbsa43 + sbsa44 + sbsa45 + sbsa46 .
blg06 = sbsa52 + sbsa53 + sbsa54 + sbsa55 + sbsa56 + sbsa57 .
blg07 = sbsa58 + sbsa59 + sbsa60 + sbsa61 + sbsa62 + sbsa26 .

**Organic Chemistry:**

org01 = sbsa07 + sbsa08 + sbsa09 + sbsa10 + sbsa11 + sbsa12 + sbsa15 .
org02 = sbsa34 + sbsa35 + sbsa36 + sbsa37 + sbsa38 + sbsa25 + sbsa39 .
org03 = sbsa47 + sbsa48 + sbsa49 + sbsa50 + sbsa51 + sbsa63 .

## Apendix A: Parceling Scheme by Discipline/Difficulty, Form 15

The following equations show which items were used to create the parcels.

**Verbal Reasoning:**

VR01hum = sVRa45 + sVRa41 + sVRa39 + sVRa38 + sVRa36 + sVRa14 .
VR02hum = sVRa47 + sVRa13 + sVRa43 + sVRa15 + sVRa11 + sVRa42 .
VR03hum = sVRa37 + sVRa46 + sVRa40 + sVRa44 + sVRa12 + sVRa16 .
VR04nst = sVRa53 + sVRa22 + sVRa17 + sVRa23 + sVRa21 .
VR05nst = sVRa48 + sVRa54 + sVRa50 + sVRa49 + sVRa18 .
VR06nst = sVRa19 + sVRa51 + sVRa55 + sVRa52 + sVRa20 .
VR07ssc = sVRa35 + sVRa01 + sVRa04 + sVRa05 + sVRa09 .
VR08ssc = sVRa34 + sVRa31 + sVRa06 + sVRa10 + sVRa08 .
VR09ssc = sVRa26 + sVRa02 + sVRa03 + sVRa33 + sVRa24 + sVRa28 .
VR10ssc = sVRa29 + sVRa30 + sVRa32 + sVRa25 + sVRa07 + sVRa27 .

**Physics:**

phy01 = psa52 + psa18 + psa17 + psa12 + psa41 + psa08 .
phy02 = psa57 + psa55 + psa49 + psa13 + psa48 + psa07 .
phy03 = psa63 + psa40 + psa16 + psa15 + psa24 + psa11 .
phy04 = psa59 + psa51 + psa54 + psa43 + psa10 + psa47 .
phy05 = psa50 + psa56 + psa26 + psa53 + psa14 + psa09 .

**General Chemistry:**

gch01 = psa39 + psa32 + psa37 + psa61 + psa42 .
gch02 = psa33 + psa44 + psa60 + psa31 + psa02 .
gch03 = psa23 + psa22 + psa36 + psa20 + psa34 .
gch04 = psa46 + psa62 + psa58 + psa01 + psa30 + psa19 .
gch05 = psa45 + psa05 + psa04 + psa03 + psa27 + psa28 .
gch06 = psa06 + psa38 + psa35 + psa21 + psa29 + psa25 .

**Biology:**

blg01 = sbsa57 + sbsa61 + sbsa60 + sbsa14 + sbsa03 .
blg02 = sbsa22 + sbsa53 + sbsa42 + sbsa41 + sbsa01 .
blg03 = sbsa46 + sbsa31 + sbsa16 + sbsa04 + sbsa18 .
blg04 = sbsa56 + sbsa45 + sbsa21 + sbsa58 + sbsa24 .
blg05 = sbsa17 + sbsa44 + sbsa30 + sbsa26 + sbsa02 .
blg06 = sbsa62 + sbsa43 + sbsa52 + sbsa40 + sbsa19 + sbsa23 .
blg07 = sbsa33 + sbsa55 + sbsa05 + sbsa06 + sbsa13 + sbsa27 .
blg08 = sbsa54 + sbsa32 + sbsa20 + sbsa59 + sbsa29 + sbsa28 .

**Organic Chemistry:**

org01 = sbsa25 + sbsa09 + sbsa48 + sbsa39 + sbsa36 .
org02 = sbsa63 + sbsa50 + sbsa49 + sbsa47 + sbsa34 .
org03 = sbsa12 + sbsa38 + sbsa10 + sbsa15 + sbsa08 .
org04 = sbsa51 + sbsa11 + sbsa37 + sbsa35 + sbsa07 .

ASSOCIATION OF AMERICAN MEDICAL COLLEGES
2450 N STREET, NW, WASHINGTON, DC 20037
PHONE 202-828-0400 FAX 202-828-1125
HTTP://WWW.AAMC.ORG

TM033687

**U.S. Department of Education**
*Office of Educational Research and Improvement (OERI)*
*National Library of Education (NLE)*
*Educational Resources Information Center (ERIC)*

# REPRODUCTION RELEASE

(Specific Document)

## I. DOCUMENT IDENTIFICATION:

Title: Appraising the Dimensionality of the MCAT

Author(s): Kevin Meara, Stephen G. Sireci

Corporate Source: The Association of American Medical Colleges, MCAT Division

Publication Date: September, 98

## II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

| The sample sticker shown below will be affixed to all Level 1 documents | The sample sticker shown below will be affixed to all Level 2A documents | The sample sticker shown below will be affixed to all Level 2B documents |
|---|---|---|
| PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY ___ Sample ___ TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) | PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY ___ Sample ___ TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) | PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY ___ Sample ___ TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) |
| 1 | 2A | 2B |
| Level 1 ✓ | Level 2A | Level 2B |
| Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy. | Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only | Check here for Level 2B release, permitting reproduction and dissemination in microfiche only |

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

*I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.*

| | |
|---|---|
| Signature: [signature] | Printed Name/Position/Title: Patricia Etienne Ed.D. Director of MCAT Research |
| Organization/Address: 2450 N Street, NW Washington, D.C. 20037-1126 | Telephone: (202) 828-0693 FAX: (202) 828-4799 |
| | E-Mail Address: petienne@aamc.org Date: 2/8/02 |

## III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, or, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

| |
|---|
| Publisher/Distributor: |
| Address: |
| Price: |

## IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

| |
|---|
| Name: |
| Address: |

## V. WHERE TO SEND THIS FORM:

| |
|---|
| Send this form to the following ERIC Clearinghouse: |

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

**ERIC Processing and Reference Facility**
**4483-A Forbes Boulevard**
**Lanham, Maryland 20706**

**Telephone: 301-552-4200**
**Toll Free: 800-799-3742**
**FAX: 301-552-4700**
**e-mail: ericfac@inet.ed.gov**
**WWW: http://ericfacility.org**

EFF-088 (Rev. 2/2001)