DOCUMENT RESUME

ED 286 884 TM 870 153

AUTHOR	Braun, Henry; And Others
TITLE	The Predictive Validity of the Scholastic Aptitude Test for Disabled Students. Studies of Admissions Testing and Handicapped People, Report No. 8.
INSTITUTION	College Entrance Examination Board, Princeton, N.J.; Educational Testing Service, Princeton, N.J.; Graduate Record Examinations Board, Princeton, N.J.
REPORT NO	ETS-RR-86-38
PUB DATE	Oct 86
NOTE	73p.; For other reports in this series, see ED 251 485, ED 251 487, ED 268 154, ED 269 418, ED 275 697, and ED 281 863.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	*College Entrance Examinations; *Disabilities; Educational Testing; Grade Point Average; *Grade Prediction; Hearing Impairments; Higher Education; High Schools; Learning Disabilities; *Physical Disabilities; *Predictive Validity; Regression (Statistics); Scores; Visual Impairments
IDENTIFIERS	*Scholastic Aptitude Test

ABSTRACT

Scholastic Aptitude Test (SAT) validity data on disabled students were obtained from 145 institutions with validity data on nonhandicapped students. First year grade point averages (FYAs) were obtained for almost 1,000 disabled students who had taken special test administrations of the SAT with extra time and for more than 650 disabled students who had taken standard test administrations. Empirical Bayes procedures were used in conjunction with the sample of nonhandicapped students to develop separate regression equations for each of the 145 institutions. This study examined whether regression equations based on data from nonhandicapped students predicted the performance of handicapped students as well as the nonhandicapped. The SAT performance of visually impaired and physically handicapped people was not very different from that of the nonhandicapped students. The SAT scores of learning-disabled students were considerable lower and those of hearing-impaired students even lower. Results also showed patterns of overprediction and underprediction in FYA predictions based on high school grades alone. In addition, SAT scores from special test administrations overpredicted the college performance of students with disabilities, especially learning disabilities. This was not true, however, for hearing-impaired students. (Author/GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED286884

RR-86-38

THE PREDICTIVE VALIDITY OF THE SCHOLASTIC APTITUDE TEST FOR DISABLED STUDENTS

Henry Braun
Marjorie Ragosta
and
Bruce Kaplan

October 1986

U S DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it
☐ Minor changes have been made to improve reproduction quality

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Educational Testing Service
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Report No. 8
Studies of Admissions Testing and Handicapped People
A Project Sponsored by

College Entrance Examination Board
Educational Testing Service
Graduate Record Examinations Board

TM 870 153

BEST COPY AVAILABLE

Studies of Admissions Testing and Handicapped People

Most admissions testing programs have long made accommodations for handicapped examinees, though practices have varied across programs and limited research has been undertaken to evaluate such test modifications. Regulations under Section 504 of the Rehabilitation Act of 1973 impose new requirements on institutional users, and indirectly on admissions test sponsors and developers, in order to protect the rights of handicapped persons. The Regulations have not been strictly enforced since many have argued that they conflict with present technical capabilities of test developers. In 1982, a Panel appointed by the National Research Council released a detailed report and recommendations calling for research on the validity and comparability of scores for handicapped persons.

Due to a shared concern for these issues, College Board, Educational Testing Service, and Graduate Record Examinations Board initiated a series of studies in June 1983. The primary objectives are:

To develop an improved base of information concerning the testing of handicapped populations.

To evaluate and improve wherever possible the accuracy of assessment for handicapped persons, especially test scaling and predictive validity.

To evaluate and enhance wherever possible the fairness and comparability of tests for handicapped and nonhandicapped examinees.

This is one of a series of reports on the project, which will continue through 1986. Opinions expressed are those of the authors. See Appendix for an annotated bibliography of earlier reports of the series.

ETS Research Report RR-86-38

The Predictive Validity of the Scholastic Aptitude Test for Disabled Students

Henry Braun

Marjorie Ragosta

Bruce Kaplan

October 1986

ERIC

Copyright © 1986 by Educational Testing Service. All rights reserved.

The ETS logo, Educational Testing Service, GRE, and Graduate Record Examinations Board. are registered trademarks of Educational Testing Service.

College Board and the acorn logo are registered trademarks of the College Entrance Examination Board.

Abstract

In two separate rounds of data collection, validity data on disabled students were obtained from 145 institutions with validity data on nonhandicapped students. First year grade point averages (FYAs) were obtained for almost 1,000 disabled students who had taken special test administrations of the SAT with extra time and for more than 650 disabled students who had taken standard test administrations. Empirical Bayes procedures were used in conjunction with the sample of nonhandicapped students to develop separate regression equations for each of the 145 institutions. The focus of this study was whether regression equations based on data from nonhandicapped students predict the performance of handicapped students as well as performance of the nonhandicapped.

Consistent with findings from other reports in this series the SAT performance of visually impaired and physically handicapped people was not very different from that of the nonhandicapped students. The SAT scores of learning-disabled students were considerably lower and those of hearing-impaired students even lower.

A pattern of over- and underprediction was evident in FYA predictions based on high school grades alone. Disabled students earning the lowest high school grade point averages tended to be underpredicted--i.e. predicted to earn FYAs lower than their actual FYAs--while disabled students earning the highest high school grades tended to be overpredicted.

A second pattern emerged from predictions based on SAT scores alone. Except for hearing-impaired students, SAT scores from special test administrations have a strong tendency to overpredict the college performance of students with disabilities. The effect is strongest for those with learning disabilities.

Using both high school grades and SAT scores to predict the college performance of students from special test administrations results in good overall predictions, but only because overprediction in some areas is offset by underprediction in others. The overprediction of the strongest third of the candidates is balanced by the underprediction of the weakest third.

Acknowledgements

We acknowledge with gratitude the work of those disabled student service providers and college admissions personnel without whose cooperation this project could not have been completed. Their time and effort in responding to our request for data are greatly appreciated.

Within ETS we wish to thank all those whose work contributed to the project, and especially:

Ka-Ling Chan and Peter Smith for their help in analyzing the data;

Linda DeLauro for coordinating and proofreading and Shirley Perry for typing the many drafts and revisions;

Randy Bennett, Faye Frieson, and Len Ramist for reviewing the final draft; and

Warren Willingham for his many helpful suggestions.

Appreciation is also due those people outside of ETS who reviewed the final report: Robert Linn from the University of Illinois, David Taggart from the University of Rhode Island, and Bernice Wong from Simon Fraser University.

Finally, we express our appreciation to the College Board Joint Staff Research and Development Committee and the Graduate Record Examinations Board Research Committee for their financial assistance.

Table of Contents

Introduction

In response to a call by the Panel on Testing of Handicapped People (Sherman & Robinson, 1982) for a program of research, the College Board (CB), Educational Testing Service (ETS), and the Graduate Record Examinations Board (GREB) jointly funded a project, "Studies of Admissions Testing and Handicapped People." As part of that research effort, the present study supplies data on the validity of the Scholastic Aptitude Test (SAT) as a predictor of college performance for people in four disability classifications: hearing impairment, learning disability, physical handicap, or visual impairment. These validity data address the question of whether the SAT predicts the college performance of people with disabilities as well as it predicts the performance of college students in general.

1. Research Design & Implementation

Although validity studies of the SAT have been routinely performed for the general population and for some special populations (e.g. minority examinees), very few validity studies have involved specific handicapped groups (Bennett, Ragosta, & Stricker, 1984; Harrison & Ragosta, 1985; Jones & Ragosta, 1982). A major focus of the federal regulations implementing Section 504 of the Rehabilitation Act of 1973 was the validity of admissions tests for disabled test takers.

The Panel on Testing of Handicapped People (Sherman & Robinson, 1982) also emphasized validity in its program of research. If it could be shown that all of the modifications made for handicapped people in a given test produced scores that predicted future performance as well as scores on the regular version, then an important source of doubt about the appropriateness of the test would disappear. The panel noted the paucity of data and suggested studies of the effects of modifying tests and testing procedures. Recognizing the difficulty of finding enough disabled students within any single institution to provide data for a standard validity study, the panel recommended developing a validation technique which would facilitate the pooling of information across many institutions. With that charge in mind, a research design was developed incorporating empirical Bayes methodology as the basis of the validation technique.

Empirical Bayes Methodology

Estimates of predictive validity are based on obtaining suitable estimates of the regression of some criterion on one or more proposed predictors. In practice, small sample sizes and the effects of self-selection impede the estimation process. Empirical Bayes methods (Braun et al., 1983; Braun & Jones, 1985; Rubin, 1980) have been employed with good effect in improving the quality of the validity estimates in a number of different settings.

With empirical Bayes, a formal mathematical model is developed in which the sets of regression coefficients from different schools are related to one another. The complexity of the relationship varies from one application to another and the appropriate form may be determined from the data. The most important consequence of this formal model is that it facilitates the "sharing of information" across schools; that is, data from all schools contribute indirectly to the estimation of the regression equation in each school. This sharing of information leads to stable estimates which are superior to the usual least squares estimate based on a single school's data. In fact, the empirical Bayes estimate of a school's prediction equation represents a compromise between the usual least squares estimate and the global estimate based on pooling the data across all schools in the study.

In the analysis of the predictive validity of SAT scores, empirical Bayes facilitates the estimation of prediction equations for each school from relatively small numbers of nonhandicapped students. Details are provided in Appendix A. Given that a suitable methodology was available, other design considerations became paramount.

Other Design Considerations

The major focus of the validity studies was to be scores derived from special test administrations of the SAT. The adapted forms of the SAT used in ATP Services for Handicapped Students are produced in four editions--Regular, Braille, Large-type, or Cassette--and are given under special conditions that may include a separate location, extra time, a reader, an amanuensis, an interpreter, rest periods, special equipment, or other adaptations. During the four-year period from the fall of 1979 through the end of June 1983, about 15,000 people took special test administrations. At the time of data collection for this study, some proportion of those 15,000 could be assumed to be in postsecondary educational institutions, some as freshmen and some at more advanced levels. Others of the 15,000 perhaps never attended college, while still others may have attended and dropped out. Before we could collect data for the validity studies, we needed to locate those people who had been admitted to college after taking special test administrations of the SAT.

A second consideration was the need for control data from nonhandicapped people who had taken regular administrations of the SAT and attended the same institutions as the handicapped students in the study. The solution to this problem appeared to be immediately at hand. The College Board, through Educational Testing Service, provides a free validity study service (VSS) to educational institutions using the SAT in their admissions process (CEEB, 1982). These validity studies traditionally make use of a measure of high school performance together with SAT verbal and mathematical scores to predict first-year college performance. During the four-year period from September 1980 through the summer of 1984, more than 400 colleges and universities participated in more than 850 validity studies.

A third consideration was the interest in studying a second control group composed of disabled people who had taken regular test administrations of the SAT. Since there existed no question concerning the presence of a disability in the Student Descriptive Questionnaire, these individuals were not easily located. The immediate task was to identify these handic ped students in postsecondary institutions which had also admitted students with scores from special test administrations.

ERIC

Given the need for data on three kinds of students, an appropriate data-collection strategy was devised. Existing data files would be searched to determine those colleges or universities which had received the largest number of score reports from special test administrations--i.e., the largest number of disabled applicants. Those schools would be matched with schools having control data through participation in the VSS. Institutions which had both VSS data and relatively large numbers of disabled applicants would be asked to participate in the study.

Data Collection

Initial contact was made by letter with 40 institutions in April 1984, later increased to 61 institutions by June 1984. The letter requested help for a series of validity studies. Specifically the letter asked for

> ....a listing — very similar to the listing provided for the VSS studies -- of all students with disabilities who have attended your institution in the last 4 years. Special services personnel could provide the listing together with a small amount of additional data to help us classify these students. The critical information from the university would include each disabled student's high school grade-point-average and current grade-point-average.....

Follow-up phone calls were made shortly after the letters were sent and periodically thereafter. The requests for data were very well received. Only two schools immediately declined to participate—one because no records were available; another because the issue was of little concern. Once schools became aware of the scope of the task, however, many found they did not have the resources to devote to the task. Disabled student service personnel--asked to provide lists of handicapped students for the validity studies--universally recognized the need for the study but had difficulty on two accounts: time and the issues surrounding confidentiality. Even when the confidentiality issues were solved, the search of student files to provide data records was, in many cases, too formidable a task to work into a busy schedule. Ultimately 31 schools provided data on handicapped students, but only 28 schools supplied data on students with scores from special test administrations.

When it became apparent that the data collection strategy would not provide enough data from nonstandard test administrations, a second strategy was employed. Again, the existing files were searched and school-by-school printouts were obtained. For each school participating in the VSS, a four-year listing was obtained of disabled students whose scores from special test administrations had been sent to the Admissions Office, presumably as part of an admissions application. The listings frequently contained only one, two, or three names. Listings containing 5 or more names were sent to the corresponding 438 institutions, which were asked to return the forms with validity data for those students who may have attended. Many schools returned the listings indicating that none of the applicants had attended, but more than 100 schools provided validity information on their few students who had taken special test administrations of the SAT. From this second round of data collection, however, no information was obtained on handicapped students with SAT scores from standard administrations.

With the information obtained from the two rounds of data collection, a data base was built. The data base was composed of information on handicapped students from special test administrations (STAs), handicapped students from regular test administrations (REGs), and a random sample of at most 50 nonhandicapped students from each of the 145 institutions that provided data on handicapped students.

2. Description of the Sample

The data base assembled for this study contained some information on 1109 STAs (handicapped people with SAT scores from special test adminstrations), 866 REGs (handicapped people with scores from regular test administrations), and 6,418 controls (nonhandicapped people). From the full data base, relevant data were drawn for two kinds of validity studies: a comprehensive study using high school grade point average (HSGPA) and SAT verbal and mathematical scores to predict first-year college grade-point-average (FYA), and a second study using only SAT scores to predict college performance. Because of missing data for some students, the two data sets differed and will be described separately.

The Comprehensive Data Set (Sample 1)

The Comprehensive Validity Study required all students to have HSGPA, SAT-Verbal, SAT-Math, and FYA scores; it also required a disability classification for all handicapped participants. More than 6,000 control students and more than 1,200 handicapped students were included in the Comprehensive Validity Study. Of the handicapped students 214 were hearing-impaired, 536 were learning-disabled, 270 were physically handicapped, and 206 were visually impaired.

Before the analyses were begun, the criterion data, i.e. the FYAs, were standardized so that within each institution the mean FYA was zero with a standard deviation of one. The standardization was done to achieve a comparable FYA scale across all institutions. Mean HSGPAs, SAT scores, and standardized FYAs are presented in Table 2-1.

Insert Table 2-1 about here

Several features of the table are worth noting. First, the numbers in some of the cells are small. There were only 35 visually impaired REGs, only 72 physically handicapped STAs, and only 84 hearing-impaired STAs. Since the numbers of individuals in some cells are small and since the total group of STAs is small in proportion to the number tested, it is important to assess the representativeness of the group of disabled test takers involved in this study.

Second, the data for this sample consistently show that STAs within any disability group on average earn lower grades in high school and college than their counterparts taking standard test administrations.

Third, except for hearing-impaired students, STAs on average earn higher scores than REGs on at least one of the SAT scores, despite their lower grades.

Fourth, visually and physically handicapped students on average earn higher SAT scores than learning-disabled students who in turn earn higher scores than hearing-impaired students. Those findings are consistent with patterns found in other studies (Bennett, Ragosta, & Stricker, 1984; Bennett, Rock & Kaplan, 1985; Ragosta & Nemceff, 1982).

In Table 2-2 the correlations among the variables are presented. Correlations between SAT scores and standardized FYAs pooled over schools are really part correlations and should be interpreted with care (Gulliksen, 1950). In the lower half of each of the five tables are the correlations for scores derived from standard test administrations, while in the upper half are the correlations from special test administrations. The correlations, for example, between SAT-V and SAT-M ranged from .52 to .66 for handicapped candidates in regular test administrations, and from .47 to .63 in special administrations, compared to .25 to .61 for nonhandicapped test takers.

Insert Table 2-2 about here

In the nonhandicapped population the lowest correlations occur between the SAT-M and FYA (.25) and between the SAT-V and FYA (.26). If we look at those relationships for handicapped test takers we note several that are markedly lower. The SAT-M/FYA correlation for visually impaired REGs is only .14 with similarly low correlations for learning-disabled (.10), physically handicapped (.15), and visually impaired (.17) STAs. Note that for hearing-impaired STAs and REGs the correlations between SAT-M and FYA are the strongest in the table (.32 and .31).

With regard to the SAT-V/FYA correlations, lower values are found for both learning-disabled STAs (.09) and REGs (.14). An interesting result is evident for hearing-impaired test takers. Although the SAT-V/FYA correlation for hearing-impaired STAs is only .04, for hearing-impaired REGs it is .26. The difference between those correlations suggests different populations of students--a point which requires a closer look.

The comprehensive data set just described as Sample 1 was the basis for three sets of analyses: one using SAT test scores and high school performance to predict college grades, a second using only test scores for the prediction, and a third using only high school performance. A data set containing larger numbers of disabled students but less comprehensive data was available for comparison.

The Test Only Data Set (Sample 2)

Because the HSGPA was missing for some of the people in the original data base, a second data set was assembled requiring only SAT scores to predict first-year averages. This data set contained 3 percent more nonhandicapped people, 22 percent more hearing-impaired candidates, 31 percent more learning-disabled students, 48 percent more physically handicapped people, and 34 percent more visually impaired test takers.

Means and standard deviations of the SAT scores and standardized FYAs for handicapped and nonhandicapped test takers are presented in Table 2-3. Despite the increased numbers, the data still show handicapped people from special SAT administrations typically earning lower first-year averages than their counterparts taking regular administrations. The rank order of the SAT scores remains the same, with the mean for visually or physically handicapped people close to the mean for nonhandicapped people while LD students and hearing-impaired candidates earned lower scores on average. Except for hearing-impaired students, STAs typically earned higher scores than REGs on at least the mathematics components of their SATs, despite lower grades.

Insert Table 2-3 about here

The relationships among the three variables in this second sample of data are presented in Table 2-4. Again, the correlations are really part correlations. The pattern of relationships remains similar to that presented for sample 1 in Table 2-2. The poorest correlation, -.05, occurs between the SAT-V scores and FYAs of hearing-impaired STAs.

Insert Table 2-4 about here

A Closer Look at Data From Students With Hearing Impairments

More than other handicapped people, hearing-impaired students in this study tended to cluster at specific institutions. One group was located at an institution established especially for hearing-impaired students. Since there were no control students at that location, the data were not used in the empirical Bayes analysis.

Two other institutions in the current study had sufficiently large numbers of hearing-impaired students to warrant a closer look at their validity data. In the first of these institutions, all hearing-impaired students routinely attended classes with their hearing counterparts in a mainstream program. In the second institution--a separate 2-year school for hearing-impaired students within a much larger technical institute -- hearing-impaired freshmen routinely took most of their coursework in separate classes designed for hearing-impaired students. The remaining hearing-impaired students in this study were located in more than 50 institutions across the United States. Because the sample sizes are so small, interpretation of the data should be made cautiously.

Sample 1. For hearing-impaired students with complete data, mean scores for students in the mainstream college, the separate college and all other colleges are presented in Table 2-5. Means for the total group are repeated at the bottom of the table.

Insert Table 2-5 about here

Note that the SAT scores of students clustered in the mainstream and separate institutions are lower than the SAT scores of all other hearing-impaired students. In fact, the hearing-impaired Regulars in all other institutions have mean scores only 9-20 points lower on the average than nonhandicapped people. Their mean FYAs are one-third of a standard deviation higher than nonhandicapped people's FYAs, and their high school means are also higher. Clearly, this particular group of hearing-impaired students is not academically at risk. STAs in all other institutions earn lower scores and grades than their Regular counterparts but higher SAT scores and college grades than students clustered in the mainstream and separate institutions. The hearing-impaired students distributed in all other institutions appear to be quite a different group and, in their distribution at least, more closely parallel the students in the other handicapped categories.

The relationships among the four validity study variables are presented in Table 2-6, with correlations from regular administrations in the lower left of each section and STA's correlations in the upper right. The correlation which was most troubling in the total data presented earlier -- the correlation between SAT-V and college FYA -- remains a matter for concern. Except for a correlation of .47 for Regulars in the mainstream institution, the relationship between SAT-Verbal scores and college grades appears to be small (.17, .06, -.04, -.09, and -.13), varying from slightly positive to slightly negative.

Insert Table 2-6 about here

We note that the FYAs of REGs at the separate institution appear to be only slightly related to SAT scores or HSGPAs, and for STAs the SAT-V shows only slight correlations with HSGPAs.

Sample 2. When we require only SAT scores and FYAs for our model, we enlarge the data base by 5 percent in the mainstream institution, by 24 percent in the separate institution, and by 43 percent in all other institutions. Mean scores are presented in Table 2-7 and correlations in Table 2-8. The data are quite similar to the data in Sample 1. The larger sample of STAs in all other schools displays a correlation of -.23 between SAT-V and FYA.

Insert Tables 2-7 and 2-8 about here

Representativeness of the Samples

Table 2-9 presents a comparison of the SAT scores of participants in the current study with the SAT scores for handicapped people in the other studies in this series (Bennett, Rock, & Kaplan, 1985: Ragosta & Kaplan, 1986).

Insert Table 2-9 about here

Bennett, Rock, and Kaplan studied those handicapped people who took special administrations of the SAT over a 3-year period and had complete test data. The 9,286 handicapped students in their study represented about 56 percent of all handicapped SAT test takers during that period. The 836 handicapped people in the Ragosta and Kaplan survey represented about 40 percent of hearing-impaired, physically handicapped, and visually impaired test takers over a one-year period, but fewer than 10 percent of the learning-disabled test takers. Only slightly more than half of the respondents to the survey were attending college during the year following their special SAT administrations. In the current study, the 764 (Sample 1) or 985 (Sample 2) disabled students are probably only a small proportion of the population of students taking special test administrations and attending college, although we have no way of knowing the size of that population. The current sample contains all special test administration students who attended any of the 145 four-year colleges which provided information to the study.

The SAT-Verbal scores of the handicapped test takers appear to be relatively consistent across all studies. The lowest scores are for hearing-impaired candidates whose means from special test administrations are more than one hundred points below the mean of the national test-taker norms. Learning-disabled candidates earn slightly higher scores although their means tend to be well below the norm mean except in the current study. The scores of physically handicapped and visually impaired candiates are the highest. In the current study the mean scores of all handicapped groups are higher than the Bennett, Rock, and Kaplan means, as would be expected if the higher scoring candidates are accepted into 4-year institutions. The means in the current study are also higher than the means of college-attending students in the Ragosta & Kaplan study. The current study obtained data from 4-year colleges and universities while the earlier study also obtained data from students in a 2-year institution. Since the Sample 1 control group in the current study earned mean SAT scores about 40 points higher than the national norm, it is not surprising that disabled students in Sample 1 earned SAT-V scores from 17 (hearing impairment) to 43 (LD) points higher than means from the Bennett, Rock, & Kaplan study.

The SAT-Mathematical scores are also consistent across studies. The ordering of the disability groupings remains the same as it was for the SAT-Verbal means. The Math means for hearing impaired and learning disabled people, however, are not quite as divergent as were their verbal means. Again, the means for the current study are higher than the means for college-attending students in the Ragosta and Kaplan study, as might be expected. The control group in Sample 1 of the current study earned a mean SAT-M score 41 points higher than the national norm, while disabled students earned mean SAT-M scores from 39 (visual impairment) to 71 (LD) points higher than the total group used in the Bennett, Rock, and Kaplan study.

The high school grade-point-averages for college students from special test administrations are almost identical across the two studies.

To summarize, then, the current study located almost 1,000 students from special SAT administrations in the 145 colleges and universities that responded to our requests for data. The responding institutions were more competitive on the average than the institutions attended by respondents to the Ragosta and Kaplan survey. If one makes the assumption that students in competitive 4-year institutions should earn higher mean scores than those in less competitive two-year institutions and that students accepted into college should earn higher mean scores than those applying to college, the data appear to be relatively consistent across studies.

3. Analyses

Introduction

In this section we examine the data collected on a nonrandom sample of handicapped college students to ascertain how well their test scores and high school grades (singly and in combination) predict their performance as freshmen in college. In theory, separate regression equations for each handicapped group in each school could be estimated and such characteristics as the R^2 (proportion of variance explained) compared across groups as well as with the results for nonhandicapped students. This plan would be difficult to carry out because there are relatively few handicapped students in any one college, especially if they are disaggregated by type of disability. Even with empirical Bayes methods, the estimated regression equations would likely not be sufficiently stable. Moreover, college admissions officers would prefer, for a number of reasons, to employ a single prediction equation to evaluate the expected performance of applicants.

Both considerations of relevance and practical constraints, therefore, lead us to focus our attention on a somewhat narrower problem: Do regression equations based on data from nonhandicapped students predict the performance of handicapped students about as well as they do that of nonhandicapped students? If not, are there any particular patterns of under- or over-prediction that are worthy of note?

Before these questions are addressed, we need to point out that first-year grade point averages may have some deficiencies as a criterion measure. If for example, college students with disabilities are not given adequate testing time for final examinations, their FYAs might be lower than anticipated. On the other hand, grades may sometimes be inflated by professors who do not wish to fail students with disabilities. Noncomparability of FYAs might also result from differences in the number and kinds of courses taken by handicapped vs. nonhandicapped students. Therefore patterns of over- or underprediction may result in part from peculiarities in the criterion measures themselves, as well as the predictors. In this report we cannot speak to the adequacy of the criterion measures and will concentrate on the adequacy of the predictors: SAT scores and high school GPA. It is well to keep in mind, however, that the criterion itself is far from an ideal measure.

We begin by examining prediction equations incorporating both test scores and high school grades. This analysis is first augmented by one in which only test scores are used. Because we did not have high school grades for a fair number of handicapped students, the latter analysis is carried out for two samples: those for whom we had both test scores and high school grades (Sample 1) and those for whom we had only test scores (Sample 2). We also carry out an analysis, based on Sample 1, in which the only predictor is high school grades.

Validity of Test Scores and High School Grades (Sample 1)

To study the validity of test scores and high school grades for predicting college grades of handicapped students, it is essential to have an appropriate baseline. In the present analysis, this baseline is provided by samples of nonhandicapped (control) students attending the 145 institutions for which we have data on handicapped students. By design, each of the 145 institutions had participated in the College Board's Validity Study Service (VSS) at least once in the four years 1980 to 1984. (For those schools that had participated more than once, data from the most recent submission was used.) If the school had more than 50 nonhandicapped students on file, a random sample of size 50 was selected. If the school had fewer than 50 nonhandicapped students on file, then all were selected. This procedure resulted in 6,255 control students in the study.

For school i, a regression equation was proposed relating first-year average in college (FYA) to the high school grade point average (HSGPA) and the verbal and mathematical scores on the SAT (V,M). The distribution of FYAs is standardized separately in each school to have mean zero and unit standard deviation.

$$FYA = B_{0i} + B_{1i}\ (SAT\text{-}V) + B_{2i}\ (SAT\text{-}M) + B_{3i}\ (HSGPA) + ERROR \qquad (i = 1, 2, \ldots, 145)$$

The index i indicates that the coefficients in the regression may vary from school to school. Despite the relatively small number of control students in each school, the estimates of the coefficients derived from the empirical Bayes methodology should be quite accurate (See Appendix A).

The 145 estimated regression equations so obtained provide the requisite baseline of the performance of the nonhandicapped students. To answer the question of how well these equations predict the FYAs of handicapped students, we need to compare actual and predicted scores: A predicted FYA is obtained for each handicapped student by substituting his/her high school grades and test scores into the prediction equation for the college attended. The difference between the actual FYA earned by the student and the predicted FYA is called the residual (for that student):

Residual = Actual FYA - Predicted FYA.

We recall here that the year in which the "actual FYA" was obtained by the handicapped student may differ from the year in which the control data was collected from that school. The usefulness of the computed residuals depends, therefore, on the assumption that in each school the year-to-year variations in the true regression of FYA on the predictors are inconsequential. To the extent that this is not the case, the variance in the pooled distribution of residuals would be inflated, decreasing the apparent validity of the predictors for handicapped students.

If the control equations yield fair predictions of the performance of all handicapped students, then the distribution of residuals for each subgroup of handicapped students should be centered on zero, the mean residual for the control students. Moreover, ideally, the variances of these distributions of residuals should be comparable to that of the distribution for control students.

The means and standard deviations of the residuals for nonhandicapped and handicapped students are presented in Table 3-1 (rows 4 and 10). A positive mean residual indicates that students do better than expected, i.e., that the control prediction equation tends to underpredict performance. A negative mean residual indicates poorer performance than expected, i.e., overprediction. Except for the hearing-impaired group, the mean residuals for the different subgroups are reasonably close to zero, when measured against the size of the standard deviations of the distributions. There is a suggestion in the data that handicapped students from standard administrations tend to be underpredicted, while those from special test administrations tend to be overpredicted. The mean residuals for the hearing-impaired groups are strongly positive - indicating severe underprediction. This underprediction merits closer attention and will be discussed below.

Insert Table 3-1 about here

It is also of some interest to compare the correlations between actual and predicted FYAs across the different groups (Row 11 of Table 3-1). These correlations, obtained by pooling data over schools, tend to be lower for the handicapped groups than for the controls, indicating that test scores and high school grades do not predict the college performance of handicapped students as well as that of controls. This outcome is buttressed by the finding that the standard deviations of the distribution of residuals for the handicapped groups were about 10 to 15 percent higher than the standard deviation of the control residuals (Row 10 of Table 3-1).

The Structure of Validity (Sample 1)

While these gross comparisons have been informative, a more detailed analysis of the distributions of residuals is possible. We first note that there does not appear to be an association across groups between the mean predicted FYA and the mean residual (Compare rows 3 and 4 of Table 3-1). That is, the typical level of preparation of a particular group of handicapped students (as measured by predicted FYA) relative to the group of controls is not related to their being over- or underpredicted as a group. We next look for associations between predicted FYA and the residuals within groups.

To this end, the individuals in each handicap group were ranked on predicted FYA and divided into three roughly equal sections (low, medium, and high predicted FYA). The mean residual for each section was calculated and the data are displayed in rows 5, 6, and 7 of Table 3-1. There is a clear tendency in the special test administration groups for the mean residual to decrease as the mean predicted FYA increases. In fact, the higher the predicted FYA for a handicapped individual the more likely it is that his/her actual FYA will fall below the prediction (overprediction). For example, in all four special test administration handicap groups, the mean residuals in the sections with the highest predicted FYAs are negative.

It is noteworthy that these four groups of handicapped test takers differ in mean predicted FYA by nearly half a standard deviation (on the standardized FYA scale). So, again, the pattern does not depend on the relative standing of the group. For those who took a standard administration, this pattern is only weakly evident, if at all. There is no such pattern for the controls.

As one would expect from the pattern of mean residuals by predicted FYA section described above, full plots of residuals against predicted FYA for each of the special test administration groups also display a negative association. Residuals were also plotted against each of the predictor variables. Again, patterns of negative association with all three predictors were especially evident among special test administration takers. Figure 3-1 provides an illustration. These results are consistent with the hypothesis that these predictors are not as strongly related to the college performance of handicapped students taking special test administrations as they are for controls. That these statements appear to be as true of high school grades as of test scores is particularly interesting. One might suspect that the pattern in residuals obtained here may be due, in part, to the use of prediction equations estimated by empirical Bayes methods. Accordingly, a parallel set of analyses was executed employing standard least squares prediction equations. The residual patterns were very similar to those described above. In fact, the negative association between residuals and predictors was even more marked.

Insert Figure 3-1 about here

The described pattern in the residuals, which is observed in subsequent analyses as well, has at least one major implication; namely, that the prediction plane for students with a particular disability is very unlikely to be parallel to the prediction plane for nonhandicapped students in the same school. (While it is remotely conceivable that the two planes could in general be parallel, the observed pattern of over- and under- prediction would have had to result from a peculiar combination of circumstances: that the distance between the two planes varies considerably across schools and that the better prepared disabled students tend to congregate in schools where the prediction plane for disabled students is further below the control prediction plane. There is no empirical evidence to support this possibility.) Whatever the case, it appears to be impossible to make simple adjustments to either the predicted FYA or the predictors to obtain unbiased predictions for disabled students at all levels of achievement. It must be either theoretically impossible or practically impossible because of prohibitively small sample sizes.

The Validity of Test Scores Alone (Sample 1)

For this analysis, regression equations of the form

$$FYA = b_{0i} + b_{1i}\ SAT\text{-}V + b_{2i}\ SAT\text{-}M + ERROR \qquad (i = 1,2....,145)$$

were estimated from data collected on control students in 145 institutions. Again, empirical Bayes methods were employed, and the resulting prediction equations were used to generate new residuals for the handicapped students on the basis of using SAT scores alone.

The mean and standard deviations of the residuals by group are presented in rows 4 and 10 of Table 3-2. Except for the hearing-impaired groups, the mean residuals are all somewhat lower than the corresponding means in Table 3-1, using predictions based on both test scores and high school grades. The largest decrease occurs with learning-disabled students taking the special test administration. On the other hand, the standard deviations are little changed. Thus, we conclude that inclusion of high school grades in the prediction equation slightly reduces the chance of overprediction for handicapped students.

Insert Table 3-2 about here

As one would expect, the correlations between actual FYA and predicted FYA (row 11 of Table 3-2) are all lower, by about 25%, than the corresponding values in Table 3-1. This is true for the nonhandicapped group as well as the handicapped group. The correlations between the residuals and the two predictors are all negative for the special test administration groups, but mixed for the standard test group. Thus, at least for the former group, the two SAT scores are again not as strongly related to college performance as they are for controls.

Each handicapped group was divided into three approximately equal sections based on predicted FYA and the mean residual for each section computed. These are displayed in rows 5, 6, and 7 of Table 3-2. As in Table 3-1, a strong negative association is evident for the special test administration groups. In fact, the FYAs for most of the handicapped students in those groups (with the exception of the hearing impaired group) would be overpredicted by the control equation based on SAT scores alone.

The Validity of High School Grades Alone (Sample 1)

It is of some interest to compare the performance of high school grades as a sole predictor of college performance with that of test scores. Table 3-3 contains the relevant data, organized in the same format as Table 3-2. For the visually and physically handicapped as well as for the learning-disabled groups, neither predictor dominates the other with regard to the size of the mean residual or the variability of the residuals. Perhaps the most dramatic difference occurs for the learning disabled taking a special test

ERIC

administration, for whom the mean residual is only -0.10 using high school grades but -0.33 using test scores. Correlations between observed and predicted college grades (row 11) tend to be somewhat higher in Table 3-3 than in Table 3-2 though comparison with Table 3-1 indicates that the addition of test scores generally improves the correlations. (Note: The corrrelations in row 11 of Table 3-3 differ slightly from the correlations between HSGPA and FYA listed in Table 2-2. This outcome results from using different empirical Bayes-estimated regression lines for each school.)

Insert Table 3-3 about here

The pattern of negative association between residuals and predicted FYA for special test administrations is also evident here. Interestingly, though, mean residuals for the hearing-impaired groups are just about zero, in contrast to the situation that occurs when test scores are included. However, a more detailed analysis (see below) suggests that this apparently benign situation is somewhat misleading.

The Validity of Test Scores Alone (Sample 2)

Recall from Section 2 that Sample 2 contains all those in Sample 1 as well as those handicapped students for whom only SAT scores were available. Given the relative scarcity of data and the fact that the number of handicapped students in Sample 2 is about one-third larger than in Sample 1, it is of some interest to see whether the findings in the previous section for Sample 1 are replicated for Sample 2. A separate system of equations involving only the SAT as a predictor of FYA was set up and empirical Bayes methods were employed to obtain estimates of the parameters. Table 3-4 presents the relevant data. Comparison with Table 3-2 indicates little material change in the results.

Insert Table 3-4 about here

Thus our data suggest that, with the exception of the hearing-impaired students, handicapped college students' first-year college grades are fairly predicted on average by control regression equations employing both test scores and high school grades. However, control regression equations using only test scores tend to overpredict the performance of handicapped students taking special test administrations, especially in the case of learning-disabled students. Inclusion of high school grades, therefore, considerably enhances the quality of the predictions. For both sets of prediction equations, higher predicted FYAs, and higher scores on the predictors are more strongly associated with overprediction, indicating that the relationship between college performance and these predictors is not as strong for handicapped students as for controls.

ERIC

Supplementary Analyses

The differential correlation between test scores and FYAs across groups of students with different handicaps presented in line 11 of Table 3-4 has a number of consequences. An important one is related to a suggestion made by some observers that the over- or underprediction of grades by test scores can be remedied by rescaling the latter through their relation "with grades." Such a global rescaling would be analogous to an equating process. However, a result of Lord (1980, Chapter 13) precludes the existence of a single equating function under the conditions of differential correlation that obtain here.

We have also found that the residuals for handicapped students tend to decrease with increasing test scores. It may be, however, that the key factor is not the absolute level of the test scores but their difference, i.e. predictions for students with large positive values of M-V may tend to be of one sign while predictions for students with large negative values of M-V may tend to be of another sign. Were this true, and were the associations stronger here than they were with M and V separately, different inferences might be appropriate. Inspection of the appropriate cross-tabulations, however, revealed no such pattern.

Another possibility is that predictions for handicapped students whose test scores are relatively low in comparison to their high school grades might behave one way, while those for students with relatively higher test scores might behave another way. One approach to investigating this hypothesis is to define a new variable which measures the relative contribution of high school grades to the predicted FYA. Let the estimated prediction equation at school i be

$$\text{FYA} = \hat{b}_{0i} + \hat{b}_{1i}\,(\text{SAT-V}) + \hat{b}_{2i}\,(\text{SAT-M}) + \hat{b}_{3i}\,(\text{HSGPA})\ .$$

Suppose a student has SAT-V score = v, SAT-M score = m, and HSGPA = h. Then we define the value of the variable PROP for that student to be

$$\text{PROP} = \frac{\hat{b}_{3i}h}{\hat{b}_{1i}v + \hat{b}_{2i}m + \hat{b}_{3i}h}$$

Thus PROP measures the fraction of the variable portion of the predicted FYA that is due to high school grades.

One hypothesis of interest is that test scores obtained by some handicapped students from special test administrations may be inflated by virtue of the unlimited amount of time they have available to them. If this were the case, these students would tend to have low values of PROP, and one would expect to see substantial evidence of relative overprediction in comparison to students with higher values of PROP. To examine this hypothesis, we generated cross-tabulations of predicted FYA and PROP with the mean residual FYA = (actual FYA - predicted FYA) displayed for each cell. No particular patterns of overprediction were observed.

Conversely, one might speculate that test scores obtained in special test administrations might not adequately capture the academic potential of many students. Such students would, presumably, tend to have high values of PROP and would tend to be underpredicted. Again, examination of the cross-tabulations described above did not support this hypothesis. Analysis of similar cross-tabulations for handicapped students taking standard administrations yielded no interesting patterns.

The Effects of Timing Condition and Test Version

The analyses of this section have focused on the freshman-year performance of the members of the different disability groups. Each group was disaggregated on the basis of predicted performance, and for three of the groups, those taking special test administrations displayed a strong trend: the higher the predicted level, the more likely they were to be overpredicted. We now reexamine the special test takers, disaggregating each disability group on the basis of the amount of extra time employed. For the two largest groups, the learning disabled and the visually impaired, we further disaggregate the data by test version. The purpose of these analyses is to determine whether there is systematic over- or underprediction when the test is taken in different versions and whether the correlations between actual FYA and predicted FYA follow a consistent pattern.

The basic data are presented in Table 3-5. Special test takers were divided into three groups (denoted short, medium, and long) depending on the amount of extra time employed. Those using less than 216 minutes are labelled short, those using between 216 and 270 minutes are labelled medium, and those using more than 270 minutes are labelled long. (The standard testing time is 150 minutes for the sections of the test used for special test administrations.)

Insert Table 3-5 about here

Once again, the hearing-impaired students appear somewhat anomalous with large positive residuals from predicted FYAs based on SATs and HSGPA (line 7). In the other three disability groups, the residuals are zero or positive for those in the S-group. The residuals are negative in the M- and L-groups with the exception of the visually impaired M-group. We do not observe, however, the same strong trends noted in Table 3-1. Moreover, predictions based on SATs alone yielded residuals (line 4) which are typically much more negative than those derived from predictions based on HSGPA alone (line 10) or those based on both test scores and HSGPA. Indeed, the mean residuals for the M- and L- groups tend to be more negative than those for the S-group. This is particularly the case for the learning-disabled group.

ERIC

In this regard, it is instructive to examine Table 3-6, which displays the test scores and HSGPAs for each group. Note that the mean total test score for learning-disabled students in the M- and L- groups is nearly 100 points higher than for the S-group, although the mean HSGPA is only 0.1 point higher. These results are consistent with the hypothesis that learning-disabled students taking substantially more time are earning higher test scores that are not matched by better performance in college. Of course, this does not establish the hypothesis: The self-selection of students into the different timing groups and the subsequent selection of colleges preclude inferences of the sort possible from random samples.

Insert Table 3-6 about here

To carry the matter further, we disaggregate the learning-disabled students by the different testing modes (Regular, Large-Type, and Cassette) as well as by amount of time. The data are presented in Table 3-7. Most of these students employ the regular test and, consequently, the residuals for that mode closely parallel those for the group as a whole. Interestingly, for predictions based on both test scores and HSGPAs the mean residuals for the three timing groups using Large Type are all negative, while those for the three timing groups using the Cassette version are all positive. Note that both groups are still overpredicted where predictions are based on test scores alone. Unfortunately, the sample sizes are very small for these six groups. For the majority of learning disabled candidates--those using regular type--disaggregation does not result in substantially different inferences from those based on Table 3-5 alone: increased time results in increased overprediction. For those using Large-Type or Cassette however, analogous inferences cannot be drawn.

Insert Table 3-7 about here

A similar detailed analysis can be carried out for the visually impaired students. The relevant data are presented in Table 3-8. Those taking the Cassette or Braille versions, presumably the most severely impaired, tend to use more time and are generally underpredicted. The largest numbers taking the Regular or Large-Type versions do not display a consistent pattern. This is true even of the residuals from predictions based on test scores alone. Again, self-selection and the unavailability of im ortant factors complicate inferences. However, blind students who use Braille or Cassette versions appear to need considerably more time than visually impaired students who can use Regular Type or Large Type.

Insert Table 3-8 about here

ERIC

Physically handicapped candidates--all but four of whom took the Regular version with extra time--appear to be best predicted when using the least amount of extra time. Additional time beyond the lowest category appears to produce significant overprediction.

Validity for Hearing-Impaired Students

The underprediction of college FYA for hearing-impaired students (large positive residuals) referred to earlier can best be studied by dividing the students into three groups: the first comprises individuals attending an institution in which they are mainstreamed but are offered excellent support services; the second, those attending a separate school within a large technical institute; and the third those scattered among 50-odd mainstream institutions. Test scores of these three groups have already been discussed in Section 2. Because the number of students in these subgroups is so small, caution is necessary in interpreting the results.

The means and standard deviations of the residuals for both special and regular test administration students in the three groups are presented in rows 4 and 10 of Table 3-9. These are Sample 1 students and the predictions are based on a combination of test scores and high school grades. It is evident that the underprediction observed for hearing-impaired students as a whole in Table 3-1 is due to severe underprediction at the separate school (mean residual = 0.73) a circumstance only partially offset by moderate overprediction at the mainstream school. Students in the third group have an average residual that is essentially zero. Plots of residuals against predicted FYA (not shown) display the characteristic negative association we have repeatedly observed in this study. Interestingly, the correlation of 0.10 between observed and predicted college FYA (row 11 of Table 3-5) in the mainstream school, is unusually small, while the correlations in the other two groups are substantial.

Insert Table 3-9 about here

For individuals taking the standard administration, the results are rather different. While students at the mainstream institution are fairly predicted on average, students in the special institution and in the third, heterogeneous, group are quite strongly underpredicted, with mean residuals of 0.43 and 0.44 respectively. In the last two groups, a strong negative association in plots of residuals against predicted FYA (not shown) is still evident. Note also that this time the correlation of .13 between observed and predicted FYA for the separate school is quite weak, due in part to greater than ordinary variability in the distribution of residuals.

The results described for the hearing-impaired students at the mainstream and separate institutions are somewhat puzzling and invite further comment (see Section 4). The data for the third group of students is most nearly comparable (in distribution across many institutions) to the data available for the other handicap groups. In that context, comparing the third group of hearing-impaired students to the other three disability groups, the only anomaly occurs for the hearing-impaired regular test administration students who are strongly underpredicted.

If we reanalyze the data of Sample 1 using high school grades as the sole predictor, some puzzling results emerge (Table 3-10). While the mean residuals for the mainstream institution are substantially worse and the correlations lower than before, the mean residuals for the separate institution are substantially improved. For example, in the regular administration, the mean residual drops from 0.43 to 0.03. On the other hand, in the heterogeneous group, the mean residual improves slightly for regular test-takers (0.44 to 0.34) but deteriorates for those taking the special test administration (0.02 to -0.18). Again the sample sizes are so small that interpretation of the results becomes difficult.

Insert Table 3-10 about here

Table 3-11 presents the results of the validity study for all hearing-impaired students in Sample 2. (In the interest of economy, a parallel analysis for Sample 1 was omitted.) Only test scores are used to obtain predicted college FYA. The exclusion of the HSGPA as a predictor improves the situation for the mainstream institution, since the mean residual is now -0.17 rather than -0.33. On the other hand, there is virtually no change for the separate institution while there is some degradation for the heterogeneous group: the mean residual is 0.14 rather than -0.02. The variances of the distributions of residuals for the three groups are hardly affected, and the strong negative association between residuals and predicted FYA remains as well.

Insert Table 3-11 about here

For the standard administration students, the exclusion of HSGPA has a deleterious effect for both the mainstream and the separate groups: With test scores as the only predictors, the first has a mean residual of 0.13 (rather than 0.01), and the second has a mean residual of 0.66 (rather than 0.43). The heterogeneous group is unaffected. Again, the variances are practically unchanged.

Comparing Tables 3-9, 3-10, and 3-11 we observe that in some cases combining test scores and high school grades cancels out the biases in the individual predictors, producing essentially fair predictions. Examples are regular test takers at the mainstream institution and special test takers in the heterogeneous group. In other cases, one of the predictors, acting alone, does better than the combination. Examples are regular test takers at the mainstream institution (test scores) and special test takers at the separate institution (high school grades). These results suggest that there is a very complex relationship for hearing-impaired students between test scores, HSGPA, and college achievement. Interpretation and understanding are hampered by the recognition that small-sample fluctuations and selection may both be acting to obscure the true dynamics. Overall, though, test scores alone perform well at the mainstream institution while high school grades alone perform well at the separate institution. For the heterogeneous group, there is no clear choice, although the correlations between actual FYAs and predicted FYAs based on high school grades are much higher than when predictions are based only on test scores.

4. Discussion

A graphical summary of the performance of Sample-1 disabled students is presented in Figure 4-1. The scales for each of the 4 variables--SAT-V, SAT-M, HSGPA, and FYA--were developed independently using the relevant means and standard deviations of the nonhandicapped students in this study. These data visually demonstrate that:

- o STAs within any disability group earn lower mean grades in high school and college than REGs earn.
- o Except for hearing-impaired students, STAs earn higher SAT-M scores than REGs, and learning-disabled STAs earn higher SAT-V scores as well.
- o Visually and physically handicapped students earn higher mean SAT scores than learning-disabled students, who, in turn, earn higher SAT scores than hearing-impaired students.

Insert Figure 4-1 about here

In the remainder of this section we will first review results for each disability group and then discuss overall findings.

Results on Students with Hearing Impairments

Performance. The SAT performance of hearing-impaired people who took either regular or special test administrations is the lowest of the four handicapped groups studied. That finding agrees with data from other studies in this series (Bennett, Ragosta, & Stricker, 1984; Bennett, Rock, & Kaplan, 1985; Ragosta & Kaplan, 1986). The high school and college performance of these students, however, is closer to that of the visually impaired and physically handicapped students than that of the lower-performing LD students.

Within the hearing-impaired group, those who take the standard SAT administrations have higher test scores, higher high school grade point averages, and higher college grades than those students who take special test administrations. That result would be expected if people whose disability had most adversely affected their academic performance requested special test administrations.

Hearing-impaired students tend to cluster in colleges/universities designed to meet their special needs. In this data base there were 83 students in a state university which provides support to hearing-impaired students mainstreamed through the university. Another 93 hearing-impaired students were located in a separate 2-year institution on the campus of a larger institute of technology. An additional 86 hearing-impaired students were widely distributed in more than 50 colleges and universities. Although students of a similar ability level as measured by the SAT were attending the first two institutions, those students who were more broadly distributed had much higher SAT scores. That finding would be expected if large programs designed for deaf students were, in fact, attracting those hearing-impaired students whose disability had most adversely affected their academic performance. In the group of widely distributed hearing-impaired students, those who took regular administrations of the SAT earned mean scores above the national means for college-bound seniors, but below the means for the control group in this study, while those who took special test administrations earned a verbal mean well below the national norm but a math mean slightly above. Students who clustered at the two institutions with special programs for the deaf earned SAT scores that were well below the national norms.

Correlations among variables. For hearing-impaired students who took special administrations of the SAT, there was little or no correlation between SAT-V and SAT-M or between SAT-V and performance in college. When we look at subgroups of hearing-impaired students within 2 institutions or across 50 institutions, only those students who took regular SATs and who attended the mainstream university show moderate correlations between the SAT-Verbal scores and their college performance. All other hearing-impaired students who took regular SATs and all those who took special administrations show low positive or negative correlations between their SAT-Verbal scores and their college FYAs. Hearing-impaired students who took special administrations have college performance more strongly related to SAT-Mathematical scores than to HSGPA or SAT-Verbal.

Predictions. Unbiased prediction of college performance would be shown by mean residuals of zero, and when high school GPAs alone are used to predict the college performance of all hearing-impaired students, the overall prediction appears to be accurate. However, data from the three subgroups of hearing-impaired students show mixed results. Students at the mainstream institution and other mainstreamed students who took special test administrations of the SAT have their college performance overpredicted by high school grades (see also Jones & Ragosta, 1982) while other students have their performance underpredicted.

When high school performance and SAT scores were included in the prediction equation, the mean residuals for hearing-impaired students were .25 for scores from standard administrations and .27 for scores from special administrations. When SAT scores alone were used, the residuals were .37 and .34 respectively. Those figures seem to indicate that SATs and HSGPA underpredict the college performance of hearing-impaired students and that using the test scores alone increases the underprediction. However, a closer look within 2 institutions and across 50 others shows that for both special and regular test administrations one institution is largely responsible for those results. The separate college on the larger institute's campus shows strong underprediction of the performance of hearing-impaired students. That result would be expected if: (a) the grading practices between the college and the institute were different, and (b) the institute's grading practices were more stringent. In fact, most freshmen at the college attend classes specifically for hearing-impaired students, and the grading system is independent of the grading practices at the larger institute. The mean FYAs of hearing-impaired students at the college (Sample 2) are only slightly lower than the mean FYAs of the nonhandicapped students at the larger institute. It would appear likely that the grading systems may not be comparable. Consultation with a research analyst at the college supported that hypothesis.

Once the problem with the separate college's data is recognized, the residuals for the remaining groups of hearing-impaired people are still inconsistent. At the mainstream institution, high school grades and SAT scores from regular test administrations predict college performance quite accurately (see also Harrison & Ragosta, 1985) but across other institutions they underpredict. High school GPAs and SAT scores from special test administrations are accurate or overpredict. Using the SATs alone, there is a slight general tendency for underprediction except for students at the mainstream institution with scores from special test administrations. Small sample sizes may contribute to the inconsistencies.

There is much less ambiguity about the residuals when they are divided into low, medium, and high predictions. When HSGPA and SAT scores are included in the prediction equation (Sample 1 data), there is a strong tendency in data from special test administrations for low predicted performance to result in underprediction and high predicted performance to result in overprediction. At the separate college where the comparability problem exists, however, although there is no overprediction, low predicted performance results in much more of an underprediction than high predicted performance. When test scores alone are used for prediction, data from regular and special test administrations show that low predictions for all subgroups result in underprediction and most high predictions result in overpredictions.

The under- and overprediction identified in these data contrast with the lack of over- and underprediction in earlier analyses using the variable PROP (page 15). Even after averaging the over- and underpredictions of subgroups of disabled students, the PROP analyses failed to discover the existence of these prediction errors.

Hearing-impaired people with low English-language skills have been shown to score poorly on the SAT. Ragosta & Kaplan (1986) showed that hearing-impaired students who described themselves as most fluent in a manual language earned much lower SAT scores than students who described themselves as most fluent in English. It may not have been possible in a special test administration to compensate for the language deficiencies of the most severely impaired students, although it may have been possible in a special test administration to overcompensate for a lesser degree of impairment. The data are supportive of such an interpretation, especially since the general finding cuts across handicapped groups.

Overall, the data on the validity of the SAT for hearing-impaired people are mixed. There is some indication that high school performance alone may be the best overall predictor of college performance, although subgroup performance fails to support that finding completely. HSGPA does predict best for students at the separate school for which we have already reported problems with the control data. Of the remaining four subgroups the lowest residuals--i.e., the most accurate FYA predictions--occur once for HSGPA alone, once for SATs alone, and twice for the combination of HSGPA and SATs. If we look at the two subgroups from special test administrations, the SAT scores increased the accuracy of the prediction.

Results on Students with Learning Disabilities

Performance. The SAT performance of learning-disabled students in this study ranges from 7-39 points below the verbal mean for college-bound seniors nationwide (College Entrance Examination Board, 1984) and from 16 points lower to 15 points higher than the mathematical mean. Their mean SAT scores are lower than the mean scores of physically handicapped, visually impaired, and nonhandicapped people, but higher than the mean scores of hearing-impaired individuals. Their high school performance, however, is lower than all handicapped and nonhandicapped groups, averaging from one-half to a full standard deviation below the control group's performance. College performance averaged one-third to one-half of a standard deviation below that of the control group.

Within the LD group, the SAT scores of students who took special administrations are higher than the scores of students in regular administrations. Despite the fact that LD students in special administrations had HSGPAs about half a standard deviation lower than those in regular administrations, the special administration students earned SAT scores about 25 points higher on both verbal and mathematical subtests.

Correlations among variables. Like the data from people with hearing impairments, the data from LD students show the lowest correlations between verbal performance on the SAT and college grades. The SAT-V/FYA correlations for LD students from special test administrations are lower than the correlations from regular administrations, but both verbal scores are only slightly related to the criterion in the prediction equation. Results are similar for both Sample 1 and Sample 2 data bases.

Prediction. The residuals from the full model in Sample 1 are close to zero (.03 for regular administrations; -.07 for special test administrations), thus indicating rather good predictive power for the SATs in conjunction with high school grades. When SATs alone are used, the residuals become more negative (-.05 for regular administrations; -.33 for special test administrations), indicating some overprediction especially for scores from special test administrations. Using high school performance alone, we found residuals were -.15 (regular) and -.10 (special). Predictions were best when both high school performance and SAT scores were used.

ERIC

When we divided residuals into thirds for low, middle, and high predictions, in Sample 1 very slight underprediction is evident for low predictions (.12 regular; .14 special), and overprediction is evident in high predictions (-.07 Regular; -.31 Special) especially for scores from special test administrations. Using test scores alone (Sample 2) we found a marked tendency for overprediction, especially for students from special test administrations (-.18 low,. -.41 middle; -.56 high). Considering that LD students in special administrations earned higher scores than those in regular administrations despite lower high school grades, and considering that their high SAT scores caused overpredictions of their college performance, one must question whether the use of unlimited time for LD students in special administrations of the SAT is warranted.

The data on test versions and testing time appear to confirm the limited use of extra time. For the majority of learning-disabled students who use the Regular-Type version of the SAT with increased time, greater amounts of time are associated with increased overprediction. The best predictions based on the SAT are associated with the smallest amount of extra time.

Results on Physically Handicapped Students

Performance. The performance of physically handicapped students on the SAT is considerably above the mean performance of college-bound seniors for both the verbal and mathematical portions of the test. Physically handicapped students in Sample 2 (test only) earned slightly higher scores than the subset of students who were in Sample 1. Both groups had SAT scores closely approximating the scores for the control groups in this study.

The test performance of physically handicapped people was higher than that of hearing-impaired or learning-disabled test takers and very close to the performance of visually impaired and nonhandicapped students.

Within the group of physically handicapped test takers, students from special test administrations earned lower verbal scores but higher mathematical scores than their counterparts in standard test administrations. Students who took special administrations were a slightly less able group as measured by their high school and college grades.

Correlations. Generally the correlations among variables were moderate (.28 to .66 Regular; .24 to .47 Special) except for a low correlation of .15 between mathematical test scores and college performance for people in Sample 1 who had taken special test administrations. For physically handicapped people in Sample 2, correlations between SAT verbal and college performance were the highest among handicapped groups.

Residuals. For physically handicapped students in Sample 1, the smallest residuals (i.e., the best predictions) occur using only high school performance for students from special administrations of the SAT (-.09) and only SAT scores for students from regular test administrations (.01). When we used both HSGPAs and SAT scores, college performance was predicted quite well: .04 for regular administrations; -.11 for specials. The correlations between predicted and actual college grades are highest using both HSGPA and SATs for students in regular test administrations but only high school GPA for students from special test administrations.

When residuals were divided into thirds (composed of low, middle, and high predicted scores), and when test scores alone were used to predict college performance, all predictions from special administrations were overpredictions. Again, there is an indication that extra time may be producing an overcompensation.

The timing data lend support to that hypothesis. Increased amounts of testing time beyond the smallest extra amount appear to produce increased overprediction. The best prediction based on SAT scores is associated with the least amount of extra time.

Results for Visually Impaired Students

Performance. Visually impaired people in this study earned SAT verbal scores about 30-50 points higher and mathematical scores about 10-35 points higher than the means for college-bound seniors and only a bit lower--or higher--than the means for control students in this study. The high school performance of visually impaired students from regular examinations was about equivalent to the performance of control students. Visually impaired students from special examinations had high school grades only about .20 points (on a 0-4 scale) lower than controls.

Visually impaired students had SAT scores higher than hearing-impaired or learning-disabled students and close to the scores of physically handicapped and nonhandicapped students in this study.

Within the visually impaired group, students from special administrations generally earned lower mean scores than students from regular test administrations except that, in Sample 2 (test only), students with special accommodations obtained slightly higher means in mathematics.

Correlations among variables. For visually impaired test takers the lowest correlations among variables tend to occur between SAT-M and college performance. Students from regular test administrations had slightly lower correlations between SAT-M and college performance than did students from special administrations.

Residuals. For visually impaired students the lowest residuals (.14 Regular and .05 Special), and the highest correlations between predicted and actual scores (.37 Regular; .37 Special) occurred with the use of both high school performance and SAT scores for predicting college grades. Using HSGPA alone we found the residuals were .17 for Regular and .06 for Special test administrations. Using only test scores, we found that visually impaired students from regular administrations had their scores slightly underpredicted (residuals of .15 in Sample 1 and .18 in Sample 2) while students from special administrations tended to have their scores slightly overpredicted (residuals of -.05 in Sample 1 and -.12 in Sample 2). The findings are consistent with the hypothesis that some low-scoring students from regular administrations should probably have been tested in a special administration and that some high-scoring students should probably have been tested in regular administrations of the SAT.

When residuals were divided into thirds for low, middle, and high predictions of college performance, a now typical pattern for special accommodations emerged: low predicted scores tended to underpredict college performance while high predicted scores tended to overpredict. This phenomenon occurred with the use of HSGPA alone, SATs alone, or both. There was also some indication that the scores for low predicted students from regular test administrations might be underpredicted.

Data on test timing and special versions of the SAT show that blind students using Braille or Cassette versions tend to use large amounts of extra time and have the resulting scores underpredict their college performance. Modal testing time for Braille and Cassette versions is greater than four and one-half hours. Visually impaired students who used Regular or Large-type SATs tended in general to complete their testing in less time.

Overall Findings

The relative standing of the four groups of handicapped students in this study parallels the findings in other studies in this series. The test performance of visually impaired and physically handicapped people is not very different from that of nonhandicapped control students. Learning disabled students score considerably lower, as might be expected for students with a diagnosed learning disability. The test scores of hearing-impaired students are a standard deviation or more below the scores of nonhandicapped students and must surely indicate that these students are the most educationally disadvantaged group of those studied.

Across all disability groups there were several trends in the data. Test scores and high school grades did not predict the college performance of handicapped people as well as that of nonhandicapped controls. Correlations between actual and predicted college performance were lower for students from special test administrations, and the standard deviations of residuals tended to be higher. Near zero average residuals often mask important over- and underpredictions.

High school performance. One pattern of over- and underprediction is evident in the data based on high school performance alone. Students earning the lowest HSGPAs are more likely to be underpredicted while students earning the highest HSGPAs are more likely to be overpredicted. Although the trend is only slightly evident for those students who elect to take standard SAT examinations, the trend is much stronger for those students who earn lower grades and who elect to take special administrations of the SAT. The most severe over- and underprediction occur for hearing-impaired students from special test administrations: low grades underpredict college performance by more than half a standard deviation, and high grades overpredict by more than one-third of a standard deviation. Why this phenomenon occurs is beyond the scope of the current study. One hypothesis for the finding is that handicapped students in special schools with strong support services may earn higher grades than those (perhaps less handicapped) students who are mainstreamed in more competitive environments.

SAT scores. A second pattern emerges from the predictions based on SAT scores alone. Except for hearing impaired students, SAT scores from special test administrations have a strong tendency to overpredict the college performance of students with disabilities. This effect is strongest for relatively high-scoring learning-disabled students whose college grades are overpredicted by more than half a standard deviation. (For low-scoring hearing-impaired students from special administrations, SAT scores underpredict college performance by more than half a standard deviation.) One possible explanation for these results is based on the policy of extending almost unlimited time to persons taking special test administrations. An earlier study of handicapped people who had taken both standard and special administrations of the SAT (Centra, 1983) showed that greater score increases were associated with greater amounts of extra time. There is some indication that gain occurs for students whose disability necessitates the extra time; i.e. those who need the time most gain most by it. But there is also an indication that more capable students are taking greater amounts of time. That finding together with the current findings on overprediction lead one to the conclusion that, in general, special test administrations need to become more standardized. This conclusion is supported by the Standards for Educational and Psychological Testing (APA-AERA-NCME, 1985), which recommends that empirical procedures be used whenever possible to establish time limits for testing handicapped people.

The College Board has established a trial program to offer more standardized testing arrangements for learning disabled students (Student Bulletin, The College Board, 1985). Under the trial arrangements, LD students who need only 1 1/2 hours of extra time can be tested in small groups on certain national test administration dates. The program will be evaluated when the trial period of one year has been completed. Current data seem to indicate that use of this program should be expanded. Larger amounts of time are associated with increased overprediction not only for the majority of learning-disabled students but also for physically handicapped students. In addition, some visually impaired and hearing-impaired students can also be tested using regular type tests with a small amount of extra time. Increasing the use of these more standardized testing arrangements for all disabled students who can take advantage of them could help to increase the validity of the SAT and, in addition, provide more disabled students with access to the SAT Question and Answer Service (Registration Bulletin, 1985).

Not all disabled students could make use of the more standardized testing arrangements. Severely disabled candidates needing large amounts of extra time and those requiring special versions of the SAT will continue to need special arrangements, including extra time. One method of determining the appropriate amount of time for a special test administration is by empirical analysis. If for example, 80 percent of nonhandicapped students can finish the SAT in the standard time, how much time is needed for 80 percent of blind students taking the Braille SAT to finish? Empirically derived administration times could be established for most test takers using special test administrations.

Another method of systematically establishing reasonable time limits for SAT special administrations is to make use of the IEP Committee. AN IEP is an Individual Education Program which is established by law for all special education students. The committee which establishes or oversees the IEP is sometimes responsible for determining the conditions under which students are tested. For example, in states with minimum competency tests for high school graduation, the IEP Committee may decide whether or not an individual should be in the high school track leading toward a diploma or in the track leading to a lesser award (e.g. a Certificate of Attendance). The IEP Committee may also make recommendations about the conditions under which the minimum competency test will be administered. Although the IEP system is currently in place nationwide and might be used to establish testing guidelines, it might be more reasonable to use the IEP Committee only in cases where prescribed standards derived from empirical data are clearly not suitable for the individual being tested.

The comprehensive model. Using both high school grades and SAT scores to predict the college performance of students from special test administration results in good overall predictions, but only because overprediction in some areas is offset by underprediction in other areas. The overprediction of college performance resulting from high school grades of the strongest third of handicapped students is balanced by the underprediction of the weakest third.

The likely error arising from overprediction is that a student is admitted to an institution in which he or she does not succeed academically. Overprediction arising from the practice of allowing large amounts of extended time on standardized tests has the effect of reducing the validity of those test scores and decreasing the correlations between predicted and actual scores. Overcompensation is also unfair to those nonhandicappe students who do not have time enough to complete the test. By accommodating the needs of handicapped students in this way, we decrease the potential for obtaining special-test-administration data with validity as high as that from standard test administrations. To increase validity a more accurate match needs to be made between the extra time needed to compensate for the disability and the actual amount of time given.

Underprediction has more serious consequences for the student in that it might result in denial of admission to an institution in which the student could succeed. Further work should be done to investigate those groups of students for whom underprediction is most severe: hearing-impaired students with low grades and low test scores (whether or not they took special test administrations) and low-scoring handicapped students generally. If, for example, the strong underprediction consistently occurs for specific groups of students, e.g. deaf students whose primary mode of communication is sign language, one might recommend the SAT not be used for that population. At a minimum, admissions officers should be alerted to the fact that handicapped people who may appear to be poor risks for college tend to perform better than expected.

Further research might also help to explicate the conditions behind the over- and underprediction resulting from the use of high school grades. A clearer understanding for this phenomenon would be of practical importance to admissions officers and could lead to stronger demonstrations of the validity of the SAT for handicapped students taking special test administrations.

As was stated previously, the general finding of over- and underprediction implies that no simple rescaling of the predictors (test scores or high school grades) will achieve the goal of more accurate predictions for disabled students. Consequently, it appears that the issue of flagging test scores from special test administrations cannot be easily resolved by appealing to a statistical adjustment of the obtained scores. Moreover, the small sample sizes and the heterogeneity of handicaps, even within a particular class of disabled students, makes it unlikely that suitable transformation of test scores can be reliably estimated.

References

American Psychological Association. (1985). Standards for educational and psychological testing. Washington, D.C.: Author.

Bennett, R. E., Ragosta, M., & Stricker, L. (1984). The test performance of handicapped people (RR-84-32). Princeton, NJ: Educational Testing Service.

Bennett, R. E., Rock, D., & Kaplan, B. (1985). The psychometric characteristics of the SAT for nine handicapped groups (RR-85-49). Princeton, NJ: Educational Testing Service.

Braun, H. I., & Jones, D. H. (1985). Use of empirical Bayes methods in the study of the validity of academic predictors of graduate school performance (RR-84-34). Princeton, NJ: Educational Testing Service.

Braun, H. I., Jones, D. H., Rubin, D. B., & Thayer, D. T. (1983). Empirical Bayes estimation of coefficients in the general linear model from data of deficient rank. Psychometr ka, 48, 71-181.

Centra, J. (1983). Handicapped student performance on a timed vs. untimed SAT. Journal of Learning Disabilities, 19, No. 6, 324-328.

College Entrance Examination Board. (1982). Guide to the College Board Validity Study Service. New York, NY: Author.

College Entrance Examination Board (1984). College-bound seniors: Eleven years of national data from the College Board's Admissions Testing Program 1973-83. New York, NY: Author.

College Entrance Examination Board. (1985). Student Bulletin for the SAT and Achievement Tests. New York, NY: Author.

Gulliksen, H. (1950). Theory of Mental Tests. John Wiley & Sons, Inc., New York, NY.

Harrison, R. H., & Ragosta, M. (1985). Identifying factors that predict deaf students' academic success in college. Unpublished report, Educational Testing Service.

Jones, D. H., & Ragosta, M. (1982). Predictive validity of the SAT on two handicapped groups: The deaf and the learning disabled. (RR-82-9). Princeton, NJ: Educational Testing Service.

Lord, F.M. (1980). Applications of Item Response Theory to Practical Testing Problems. Lawrence Erlbaum Associates, Inc., Hillsdale, NJ.

Ragosta, M., & Kaplan, B. (1986). A Survey of handicapped students taking special test administrations of the SAT and GRE. Princeton, NJ: Educational Testing Service.

Rubin, D. B. (1980). Using empirical Bayes techniques in the law school validity studies. Journal of the American Statistical Association, 75, 801-816.

Sherman, S., & Robinson, N. (1982). Ability testing and handicapped people: Dilemma for government, science, and the public. Washington, D.C.: National Academy Press.

Snedecor, G. W., & Cochran, W. G. (1980). Statistical Methods. Ames, IA: Iowa State University Press.

TABLES AND FIGURES

Table 2-1

Means and Standard Deviations for Nonhandicapped and Handicapped Groups

Sample 1

	N	SAT-V $\overline{X}$ (SD)	SAT-M $\overline{X}$ (SD)	HSGPA $\overline{X}$ (SD)	FYA (Stand.) $\overline{X}$ (SD)
NONHANDICAPPED	6255	465 (99)	509 (107)	3.19 (0.55)	0.00 (1.00)
HEARING IMPAIRED					
Regular	130	356 (129)	438 (116)	3.10 (0.53)	-0.06 (1.08)
Special	84	315 (97)	429 (117)	2.87 (0.61)	-0.24 (0.96)
LEARNING DISABLED					
Regular	99	385 (88)	452 (98)	2.90 (0.53)	-0.38 (1.12)
Special	437	412 (87)	477 (116)	2.65 (0.54)	-0.49 (1.00)
PHYSICALLY HANDICAPPED					
Regular	198	473 (108)	494 (121)	3.21 (0.59)	0.00 (0.95)
Special	72	462 (94)	519 (111)	3.09 (0.50)	-0.19 (1.07)
VISUALLY IMPAIRED					
Regular	35	474 (117)	489 (87)	3.22 (0.47)	0.20 (1.00)
Special	171	452 (92)	502 (123)	3.00 (0.58)	-0.11 (1.06)

ERIC Full Text Provided by ERIC

Table 2-2

Correlations Among Variables for Nonhandicapped and Handicapped Groups[1]

Sample 1

Group			SAT-V	SAT-M	HSGPA	FYA	
NONHANDICAPPED							
		SAT-V	--				
		SAT-M	.61	--			
	Regular	HSGPA	.40	.45	--		
		FYA	.26	.25	.37	--	
HEARING							
		SAT-V	--	.63	.23	.04	
		SAT-M	.60	--	.38	.32	Special
	Regular	HSCPA	.31	.45	--	.21	
		FYA	.26	.31	.35	--	
LEARNING							
		SAT-V	--	.61	.22	.09	
		SAT-M	.58	--	.36	.10	Special
	Regular	HSGPA	.16	.37	--	.24	
		FYA	.14	.26	.22	--	
PHYSICAL							
		SAT-V	--	.47	.43	.24	
		SAT-M	.66	--	.38	.15	Special
	Regular	HSGPA	.38	.45	--	.40	
		FYA	.37	.28	.42	--	
VISUAL							
		SAT-V	--	.50	.49	.19	
		SAT-M	.52	-	.46	.17	Special
	Regular	HSGPA	.37	.48	--	.28	
		FYA	.29	.14	.44	--	

[1] Correlations from regular administrations are to the lower left of the diagonals; correlations from special administrations are to the upper right.

Table 2-3

Means and Standard Deviations for Nonhandicapped and Handicapped Groups

Sample 2

	N	SAT-V $\overline{X}$	SAT-V (SD)	SAT-M $\overline{X}$	SAT-M (SD)	FYA (Stand.) $\overline{X}$	FYA (Stand.) (SD)
NONHANDICAPPED	6448	465	(99)	509	(107)	0.00	(1.00)
HEARING IMPAIRED							
Regular	157	365	(135)	445	(122)	0.01	(1.08)
Special	105	307	(96)	418	(114)	-0.18	(0.93)
LEARNING DISABLED							
Regular	129	404	(100)	465	(104)	-0.36	(1.10)
Special	574	417	(88)	483	(118)	-0.53	(1.02)
PHYSICALLY HANDICAPPED							
Regular	311	481	(108)	504	(121)	0.00	(0.98)
Special	89	470	(98)	519	(120)	-0.17	(1.02)
VISUALLY IMPAIRED							
Regular	59	475	(111)	500	(110)	0.18	(1.04)
Special	217	455	(97)	504	(126)	-0.16	(1.08)

ERIC

Table 2-4

Correlations Among Variables for Nonhandicapped and Handicapped Groups[1]

Sample 2

			SAT-V	SAT-M	FYA	
NONHANDICAPPED						
		SAT-V	--			
		SAT-M	.60	--		
	Regular	FYA	.26	.24	--	
HEARING						
		SAT-V	--	.64	-.05	Special
		SAT-M	.68	--	.26	
	Regular	FYA	.24	.29	--	
LEARNING						
		SAT-V	--	.61	.12	Special
		SAT-M	.64	--	.12	
	Regular	FYA	.16	.22	--	
PHYSICAL						
		SAT-V	--	.51	.24	Special
		SAT-M	.65	--	.24	
	Regular	FYA	.28	.18	--	
VISUAL						
		SAT-V	--	.53	.20	Special
		SAT-M	.46	--	.20	
	Regular	FYA	.24	.13	--	

[1] Correlations from regular administrations are to the lower left of the diagonals; correlations from special administrations are to the upper right.

Table 2-5

Means and Standard Deviations for Hearing-Impaired Subgroups

Sample 1

	N	SAT-V $\overline{X}$	(SD)	SAT-M $\overline{X}$	(SD)	HSGPA $\overline{X}$	(SD)	FYA (Stand.) $\overline{X}$	(SD)
MAINSTREAM									
Regular	57	320	(120)	416	(109)	3.08	(0.46)	-0.20	(0.88)
Special	22	307	(85)	407	(102)	3.09	(0.39)	-0.59	(0.88)
SEPARATE									
Regular	34	315	(107)	389	(106)	2.89	(0.56)	-0.27	(1.37)
Special	41	291	(92)	413	(124)	2.63	(0.67)	-0.11	(0.97)
ALL OTHERS									
Regular	39	445	(112)	500	(107)	3.31	(0.52)	0.32	(0.95)
Special	21	369	(96)	484	(102)	3.12	(0.49)	-0.12	(0.95)
TOTAL (ABOVE)									
Regular	130	356	(129)	438	(116)	3.10	(0.53)	-0.06	(1.08)
Special	84	315	(97)	429	(117)	2.87	(0.61)	-0.24	(0.96)

Table 2-6

Relationships Among Variables for Hearing-Impaired Subgroups[1]

Sample 1

			SAT-V	SAT-M	HSGPA	FYA	
MAINSTREAM							
		SAT-V	-	.55	.12	-.13	
		SAT-M	.50	-	.35	.35	Special
	Regular	HSGPA	.13	.24	-	.03	
		FYA	.47	.27	.39	-	
SEPARATE							
		SAT-V	-	.61	.18	.17	
		SAT-M	.62	-	.42	.34	Special
	Regular	HSGPA	.29	.47	-	.38	
		FYA	-.04	.16	.13	-	
ALL OTHER							
		SAT-V	-	.61	.10	-.09	
		SAT-M	.50	-	.22	.26	Special
	Regular	HSGPA	.31	.51	-	.26	
		FYA	.06	.33	.49	-	
TOTAL (ABOVE)							
		SAT-V	-	.63	.23	.04	
		SAT-M	.60	-	.38	.32	Special
	Regular	HSGPA	.31	.45	-	.21	
		FYA	.26	.31	.35	-	

[1]Correlations from regular administrations are to the lower left of the diagonals; correlations from special administrations are to the upper right.

ERIC

Table 2-7

Means and Standard Deviations for Hearing-Impaired Subgroups

Sample 2

	N	SAT-V $\bar{X}$	SAT-V (SD)	SAT-M $\bar{X}$	SAT-M (SD)	FYA (Stand.) $\bar{X}$	FYA (Stand.) (SD)
MAINSTREAM							
Regular	59	317	(120)	422	(109)	-0.21	(0.89)
Special	24	299	(86)	398	(103)	-0.61	(0.88)
SEPARATE							
Regular	41	310	(111)	387	(107)	-0.05	(1.40)
Special	52	282	(86)	414	(113)	-0.02	(0.92)
ALL OTHERS							
Regular	57	453	(118)	511	(114)	0.27	(0.93)
Special	29	360	(102)	477	(109)	-0.13	(0.89)
TOTAL (ABOVE)							
Regular	157	365	(135)	445	(122)	0.01	(1.08)
Special	105	307	(96)	418	(114)	-0.18	(0.93)

Table 2-8

Relationships Among Variables for Hearing-Impaired Subgroups[1]

Sample 2

		SAT-V	SAT-M	FYA	
MAINSTREAM					
	SAT-V	-	.59	-.09	Special
Regular	SAT-M	.51	-	.32	
	FYA	.47	.27	-	
SEPARATED					
	SAT-V	-	.59	.08	Special
Regular	SAT-M	.66	-	.32	
	FYA	-.03	.18	-	
ALL OTHERS					
	SAT-V	-	.67	-.23	Special
Regular	SAT-M	.66	-	.07	
	FYA	.11	.32	-	
TOTAL (ABOVE)					
	SAT-V	-	.64	-.05	Special
Regular	SAT-M	.68	-	.26	
	FYA	.24	.29	-	

[1]Correlations from regular administrations are to the lower left of the diagonals; correlations from special administrations are to the upper right.

Table 2-9

Comparison of SAT Performance Across Three Studies of Disabled Candidates Taking Special Test Administrations of the SAT

	Bennett, Rock & Kaplan[1]	Ragosta & Kaplan		Current Study	
	Total	Total	In College	Sample 1	Sample 2
Number					
Hearing Impairment	456	123	72	84	105
Learning Disability	6435	275	194	437	574
Physical Handicap	644	131	54	72	89
Visual Impairment	1751	307	124	171	217
SAT-V					
Hearing Impairment	298**	292**	284**	315**	307**
Learning Disability	369*	380*	394*	412	417
Physical Handicap	427	420	430	462	470
Visual Impairment	418	440	442	452	455
SAT-M					
Hearing-Impairment	385*	383*	380*	429*	418*
Learning Disability	406*	428*	448	477	483
Physical Handicap	444	445	460	519	519
Visual Impairment	450	476	460	502	504
HSGPA					
Hearing Impairment	-	2.93	2.84	2.87	-
Learning Disability	-	2.65	2.68	2.65	-
Physical Handicap	-	3.00	3.07	3.09	-
Visual Impairment	-	3.04	3.08	3.00	-

National Norms: SAT-V 424; SAT-M 468

* 1/4-1 SD below the national norm
** >1 SD below the national norm

[1] Tabled values were calculated from data presented by Bennett et al.

Table 3-1

Sample 1: Using HSGPA & SATs to Predict FYA

Row		Nonhandi-capped Controls	Disabilities: Hearing Standard	Hearing Special	Learning Standard	Learning Special	Physical Standard	Physical Special	Visual Standard	Visual Special
1	Number	6255	130	84	99	437	198	72	35	171
	Means									
2	Actual FYA	0.00	-0.06	-0.24	-0.38	-0.49	0.00	-0.19	0.20	-0.11
3	Predicted FYA	0.00	-0.31	-0.51	-0.41	-0.42	-0.04	-0.08	0.06	-0.16
4	Residual	0.00	0.25	0.27	0.03	-0.07	0.04	-0.11	0.14	0.05
	Residuals									
5	Low Predicted	.03	.52	.74	.12	.14	.15	.21	.28	.31
6	Med. Predicted	-.07	.02	.25	.04	-.03	-.03	-.26	-.20	.02
7	High Predicted	.04	.21	-.20	-.07	-.31	.02	-.23	.27	-.18
	Standard Deviations									
8	Actual FYA	1.00	1.08	0.96	1.12	1.00	0.95	1.07	1.00	1.06
9	Predicted FYA	0.50	0.56	0.60	0.50	0.50	0.54	0.52	0.40	0.55
10	Residual	0.87	1.00	1.01	1.07	0.96	0.82	1.01	0.93	1.00
	Correlations									
11	Actual & Pred.	.49	.39	.23	.33	.34	.50	.35	.37	.37

Table 3-2

Sample 1: Using Only SATs to Predict FYA

Row		Nonhandi-capped Controls	Disabilities							
			Hearing		Learning		Physical		Visual	
			Standard	Special	Standard	Special	Standard	Special	Standard	Special
1	Number	6255	130	84	99	437	198	72	35	171
	Means									
2	Actual FYA	0.00	-0.06	-0.24	-0.38	-0.49	0.00	-0.19	0.20	-0.11
3	Predicted FYA	0.00	-0.37	-0.50	-0.32	-0.15	-0.01	0.00	0.05	-0.05
4	Residual	0.00	0.31	0.26	-0.05	-0.33	0.01	-0.18	0.15	-0.05
	Residuals									
5	Low Predicted	.03	.32	.60	.16	-.14	.09	-.11	.55	.07
6	Med. Predicted	-.04	.56	.08	-.39	-.30	-.03	-.24	-.27	-.09
7	High Predicted	.01	.05	.11	.07	-.53	-.02	-.19	.08	-.13
	Standard Deviations									
8	Actual FYA	1.00	1.08	0.96	1.12	1.00	0.95	1.07	1.00	1.06
9	Predicted FYA	0.37	0.46	0.43	0.39	0.39	0.39	0.39	0.34	0.39
10	Residual	0.93	1.03	0.98	1.08	0.98	0.89	1.04	0.99	1.01
	Correlations									
11	Actual & Pred.	.37	.31	.19	.29	.22	.36	.22	.22	.31

Table 3-3

Sample 1: Using Only HSGPA to Predict FYA

Row		Nonhandi-capped Controls	Disabilities							
			Hearing		Learning		Physical		Visual	
			Standard	Special	Standard	Special	Standard	Special	Standard	Special
1	Number	6255	130	84	99	437	198	72	35	171
	Means									
2	Actual FYA	0.0	-0.06	-0.24	-0.38	-0.49	0.00	-0.19	0.20	-0.11
3	Predicted FYA	0.0	-0.06	-0.23	-0.23	-0.38	-0.04	-0.10	0.03	-0.17
4	Residual	0.0	0.00	-0.01	-0.15	-0.10	0.04	-0.09	0.17	0.06
	Residuals									
5	Low Predicted	.03	.22	.56	-.02	.04	.08	.26	-.05	.33
6	Med. Predicted	-.04	-.06	-.24	-.10	-.05	.04	-.61	.64	-.01
7	High Predicted	.01	-.19	-.36	-.31	-.30	.01	.14	-.03	-.14
	Standard Deviations									
8	Actual FYA	1.00	1.08	.96	1.12	1.00	.95	1.07	1.00	1.06
9	Predicted FYA	.43	.47	.57	.45	.47	.50	.44	.37	.48
10	Residual	.91	1.03	1.02	1.10	.97	.85	1.00	.94	1.02
	Correlations									
11	Actual & Pred.	.41	.32	.19	.25	.30	.44	.36	.35	.32

Table 3-4

Sample 2: Using Only SATs to Predict FYA

Row		Nonhandicapped Controls	Disabilities							
			Hearing		Learning		Physical		Visual	
			Standard	Special	Standard	Special	Standard	Special	Standard	Special
1	Number	6448	157	105	129	574	311	89	59	217
	Means									
2	Actual FYA	0.00	0.01	-0.18	-0.36	-0.53	0.00	-0.17	0.18	-0.16
3	Predicted FYA	0.00	-0.36	-0.53	-0.28	-0.15	-0.02	0.00	0.00	-0.04
4	Residual	0.00	0.37	0.34	-0.08	-0.39	0.02	-0.17	0.18	-0.12
	Residuals									
5	Low Predicted	.03	.47	.72	.12	-.18	.17	-.07	.28	.05
6	Med. Predicted	-.04	.48	.28	-.34	-.41	-.02	-.26	.24	-.20
7	High Predicted	.00	.16	.04	-.03	-.56	-.09	-.17	.02	-.20
	Standard Deviations									
8	Actual FYA	1.00	1.08	0.93	1.10	1.02	0.98	1.02	1.04	1.08
9	Predicted FYA	0.37	0.47	0.42	0.41	0.39	0.39	0.38	0.38	0.41
10	Residual	0.93	1.05	0.99	1.06	1.01	0.95	0.99	0.99	1.04
	Correlations									
11	Actual & Pred.	.37	.28	.09	.28	.23	.30	.26	.30	.29

ERIC

Table 3-5

Predicted Performance of Handicapped Students
Disaggregated by Disability and Timing Condition

	HEARING			LEARNING			PHYSICAL			VISUAL		
	S	M	L	S	M	L	S	M	L	S	M	L
1. Number	31	23	11	167	137	121	24	23	21	57	52	56
2. Actual FYA	-.11	-.20	.00	-.50	-.43	-.47	.09	-.40	-.46	-.18	.00	-.08
3. Predicted FYA (SAT Only)	-.68	-.43	-.27	-.28	-.07	-.07	.02	.08	-.12	-.05	-.13	.05
4. Residual	.57	.23	.27	-.22	-.36	-.40	.07	-.48	-.32	-.13	.13	-.13
5. Correlation, Actual and Predicted FYA	.24	.40	-.04	.26	.19	.31	.05	.38	.14	.04	.27	.54
6. Predicted FYA (SAT & HSGPA)	-.73	-.45	-.45	-.54	-.34	-.35	-.08	-.10	-.16	-.19	-.31	-.01
7. Residual	.62	.26	.45	.03	-.09	-.12	.17	-.31	-.28	.00	.31	-.07
8. Correlation, Actual and Predicted FYA	.43	.39	-.01	.38	.24	.42	.23	.46	.46	.22	.35	.55
9. Predicted FYA (HSGPA Only)	-.36	-.22	-.42	-.42	-.35	-.38	-.12	-.17	-.10	-.19	-.28	-.06
10. Residual	.25	.03	.41	-.08	-.09	-.09	.21	-.23	-.34	.01	.28	-.02
11. Correlation, Actual and Predicted FYA	.46	.30	-.06	.32	.19	.32	.38	.45	.44	.30	.31	.42

Table 3-6

Test Scores and High School Grades for
Disabled Students, Disaggregated by Disability
and Timing Condition

	Timing Condition	Hearing	Learning	Physical	Visual
SAT-V	S	288	387	470	455
	M	332	433	470	439
	L	376	423	441	467
SAT-M	S	390	438	555	496
	M	477	497	525	483
	L	494	505	474	526
HSGPA	S	2.74	2.59	3.15	2.95
	M	2.96	2.71	2.99	2.89
	L	2.65	2.68	3.03	3.14

Table 3-7

Predicted Performance of Learning Disabled Students
Disaggregated by Test Version and Timing Condition

Learning Disabled

	Regular			Large Type			Cassette		
	S	M	L	S	M	L	S	M	L
1. Number	143	121	103	7	3	3	13	11	13
2. Actual FYA	-.48	-.45	-.46	-.89	-.45	-1.80	-.41	-.28	-.24
3. Predicted FYA (SAT Only)	-.26	-.05	-.08	-.29	.10	-.10	-.32	-.26	.09
4. Residual	-.22	-.40	-.38	-.60	-.55	-1.70	-.09	-.02	-.33
5. Predicted FYA (SAT & HSGPA)	-.54	-.32	-.36	-.53	-.37	-.31	-.50	-.45	-.34
6. Residual	.06	-.13	-.10	-.36	-.08	-1.49	.09	.17	.10
7. Predicted FYA (HSGA Only)	-.43	-.34	-.37	-.44	-.60	-.31	-.34	-.32	-.49
8. Residual	-.05	-.11	-.09	-.45	.15	-1.49	-.07	.04	.25

Table 3-8

Predicted Performance of Visually Impaired Student
Disaggregated by Test Version and Timing Condition

Visually Impaired

	Regular			Large Type			Cassette			Braille		
	S	M	L	S	M	L	S	M	L	S	M	L
1. Number	19	13	14	36	35	25	0	3	4	1	1	13
2. Actual FYA	-.35	-.51	-.09	-.14	.19	-.31	-	-.15	.33	1.06	.39	.24
3. Predicted FYA (SAT Only)	-.17	-.13	-.02	-.01	-.12	.04	-	-.28	.20	.12	-.01	.10
4. Residual	-.18	-.38	-.07	-.13	.31	-.35	-	.13	.13	.94	.40	.14
5. Predicted FYA (SAT & HSGPA)	-.35	-.49	-.09	-.12	-.23	-.04	-	-.58	.03	.30	.07	.11
6. Residual	.00	-.02	.00	-.02	.42	-.27	-	.43	.30	.76	.32	.13
7. Predicted FYA (HSGA Only)	-.28	-.48	-.08	-.15	-.20	-.09	-	-.54	-.15	.30	.04	.05
8. Residual	-.07	-.03	-.01	.01	.39	-.22	-	.39	.48	.76	.35	.19

Table 3-9

Hearing-Impaired Subgroups

Sample 1: Using HSGPA & SATs to Predict FYA

		Mainstream		Separate		All Others	
Row		Regular	Special	Regular	Special	Regular	Special
1	Number	57	22	34	41	39	21
	Means						
2	Actual FYA	-0.20	-0.59	-0.27	-0.11	0.32	-0.12
3	Predicted FYA	-0.21	-0.26	-0.70	-0.85	-0.12	-0.11
4	Residual	0.01	-0.33	0.43	0.73	0.44	-0.02
	Residuals						
5	Low Predicted	.09	.35	.71	.78	.54	.31
6	Med. Predicted	-.09	-.60	.30	.83	.00	.00
7	High Predicted	.11	-.71	.21	.28	.41	-.18
	Standard Deviations						
8	Actual FYA	0.88	0.88	1.37	0.97	0.95	0.95
9	Predicted FYA	0.49	0.42	0.54	0.59	0.50	0.36
10	Residual	0.74	0.93	1.41	0.90	0.82	0.87
	Correlations						
11	Actual & Predicted FYA	.53	.10	.13	.42	.50	.41

ERIC

Table 3-10

Hearing-Impaired Subgroups

Sample 1: Using HSGPA Only

Row		Mainstream		Separate		All Others	
		Regular	Special	Regular	Special	Regular	Special
1	Number	57	22	34	41	39	21
	Means						
2	Actual FYA	-0.20	-0.59	-0.27	-0.11	0.32	-0.12
3	Predicted FYA	0.06	0.06	-0.29	-0.53	-0.02	0.06
4	Residual	-0.26	-0.65	0.03	0.42	0.34	-0.18
	Residuals						
5	Low Predicted	-.20	-.35	.46	.65	.34	-.10
6	Med. Predicted	-.02	-.79	.20	.54	-	-
7	High Predicted	-.44	-.79	.77	.07	.34	.21
	Standard Deviations						
8	Actual FYA	0.88	0.88	1.37	0.97	0.95	0.95
9	Predicted FYA	0.40	0.34	0.50	0.59	0.46	0.39
10	Residual	0.81	0.93	1.40	0.92	0.81	0.90
	Correlations						
11	Actual & Predicted FYA	.39	.03	.13	.38	.52	.34

Table 3-11

Hearing-Impaired Subgroups

Sample 2: Using Only SATs to Predict FYA

Row		Mainstream		Separate		All Others	
		Regular	Special	Regular	Special	Regular	Special
1	Number	59	24	41	52	57	29
	Means						
2	Actual FYA	-0.21	-0.61	-0.05	-0.02	0.27	-0.13
3	Predicted FYA	-0.35	-0.44	-0.70	-0.71	-0.13	-0.27
4	Residual	0.13	-0.17	0.66	0.69	0.41	0.14
	Residuals						
5	Low Predicted	.06	.08	.74	.53	.84	.66
6	Medium Pred.	.29	-.25	1.05	.80	.00	.05
7	High Predicted	-.08	-.52	-.39	-.11	.33	-.36
	Standard Deviations						
8	Actual FYA	0.89	0.88	1.40	0.92	0.93	0.89
9	Predicted FYA	0.45	0.38	0.40	0.36	0.39	0.39
10	Residual	0.80	0.91	1.43	0.90	0.89	0.96
	Correlations						
11	Actual & Predicted FYA	.43	.14	.08	.24	.30	.04

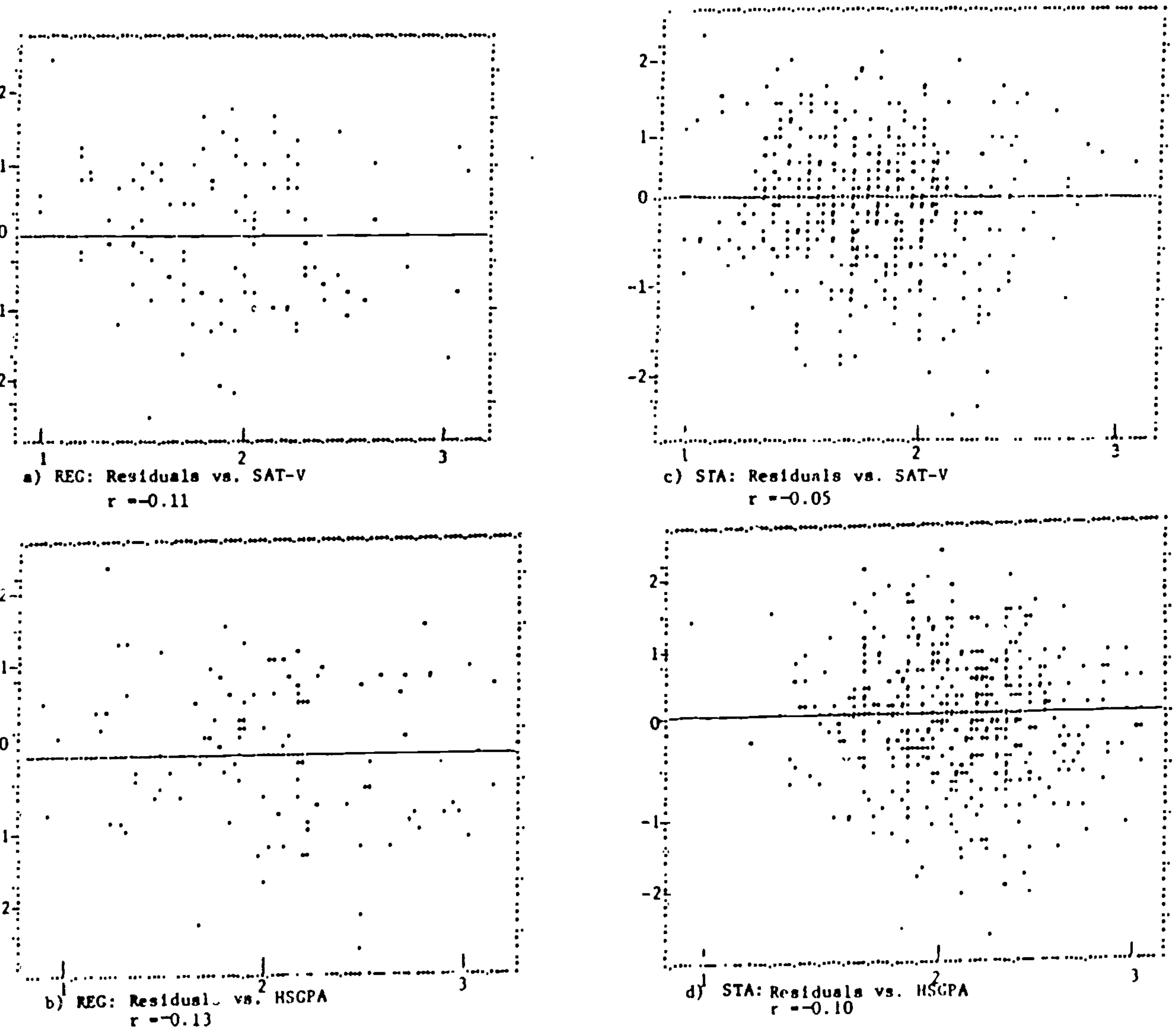

Figure 3.1 Plots of Residuals Against Predictors for Learning Disabled Students (REG and STA)

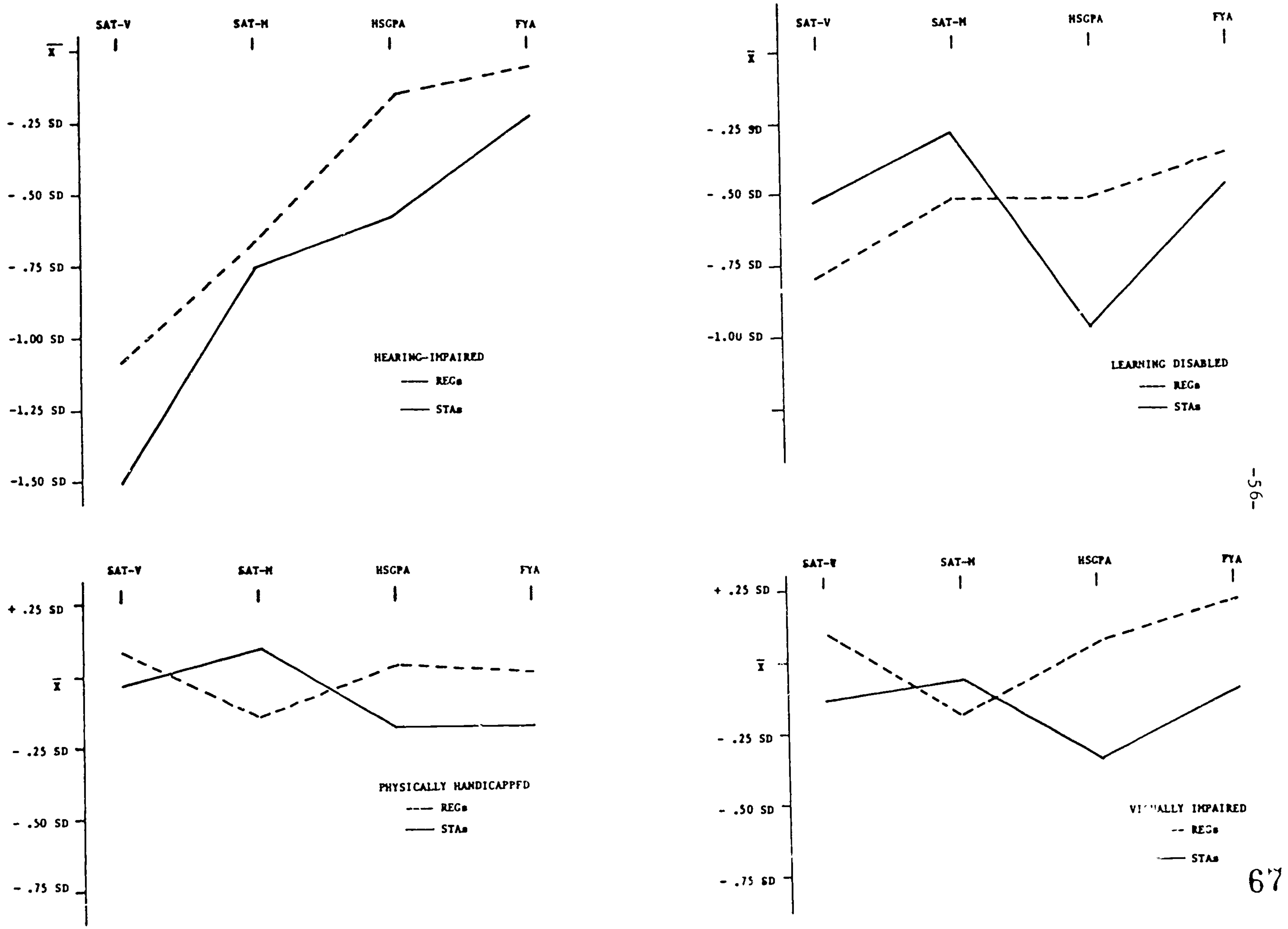

FIGURE 4-1. GRAPHICAL SUMMARY OF THE PERFORMANCE OF DISABLED STUDENTS: SAMPLE 1

(IN STANDARD DEVIATION UNITS OF THE NONHANDICAPPED POPULATION.)

APPENDICES

Appendix A

Standard regression methods (Snedecor and Cochran, 1980) are based on the least squares principle. It has been found, however, that the least squares regression line may be too greatly influenced by idiosyncrasies in the data and, consequently, does not generally perform well in cross-validation. That would be a severe flaw in the present setting, where the prediction equations based on data from nonhandicapped students are used to provide baseline predictions for handicapped students. If those predictions are poor (i.e. large bias and/or excessive variability), then the chance of making meaningful inferences from residual analyses becomes remote.

Fortunately, when many regressions in related problems can be estimated simultaneously, empirircal Bayes methods (Rubin, 1980; Braun et al., 1983; Braun & Jones, 1985) can be employed to good advantage. Empirical Bayes provides a practical and useful way of combining information across schools to improve the estimation of the regression line in each school.

To borrow a term from sociology, empirical Bayes facilitates a very general form of "contextual analysis." Essentially, the relation between the criterion and a constellation of predictors within a given department is examined in the setting of a large collection of departments. Of particular interest is any evidence that the nature of this relation varies in association with some measured characteristic(s) of the departments. An example might be the finding that the inclination of the regression plane increases as the department size increases. To the extent that such departmental findings are valid, the precision of the estimation carried out in any one department can be improved by drawing upon the information provided by the other departments.

Our aim is to estimate for each school in the sample an equation of the form:

$$Y_{ij} = B_{oi} + B_{1i} V_{ij} + B_{2i} M_{ij} + B_{3i} U_{ij} + e_{ij}, \qquad (1)$$

where i indexes schools and j indexes students within schools. The criterion, Y, is the first-year average (FYA) in college, standarized separately in each college to have zero mean and unit variance. V and M represent scores on the verbal and mathematical forms of SAT, rescaled by dividing by 200. Thus, the regression coefficients for these variables should be of comparable magnitude to that for undergraduate grade point average (UGPA), denoted by U in the equation, which is on a 0-4 scale. The errors e_{ij} are assumed to be independent and normally distributed with zero mean and variance $\sigma_i 2$.

Interest centers on estimation of the vector of parameters

$$B_i = (B_{oi}, B_{1i}, B_{2i}, B_{3i})'.$$

The empirical Bayes formulation takes the form of an hierarchical linear model by assuming in addition to (1) that

$$B_i' = Z_i'G + D_i'. \qquad (2)$$

where Z_i is a vector of school-level characteristics, G is a matrix of coefficients to be estimated, and D_i is a vector of random fluctuations:

$$D_i \sim N(o, \Sigma^*). \qquad (3)$$

The model encompassed by (1), (2), and (3) facilitates the sharing of information across schools since the empirical Bayes estimate of B_i, $\tilde{B}_i$ will depend not only on the data from school i (as would the least squares estimate, $\hat{B}_i$) but also on the value of $Z_i'\tilde{G}$, a point on the plane characterized by the matrix $\tilde{G}$. All the schools contribute to the estimation of G and hence will influence the value of $\tilde{B}_i$. For more details, see Braun and Jones (1985).

It can be shown that $\tilde{B}_i$ may be expressed in the form:

$$\tilde{B}_i = W_i \hat{B}_i + (1-W_i)Z_i'\tilde{G}.$$

That is, the empirical Bayes estimate for a school is a weighted combination of the least squares estimate for that school and a "pooled" estimate based on the apparent association between the school regression coefficients and various school characteristics. The weights are proportional to the relative (estimated) precisions of the two component estimates. Thus, if B_i has relatively low precision, perhaps because of a small sample size or the configuration of the sample, then $\tilde{B}_i$ will be "pulled" closer to the pooled estimate, $Z_i'\tilde{G}$. Note that for different schools, $\tilde{B}_i$ is pulled toward different points, depending on the value of Z_i. In this paper, we employ $Z_i = (1, V_i, M_i, U_i)$, where the last three components are the mean values on the three predictors for the students in school i.

Appendix B

The following previous reports from "Studies of Admissions Testing and Handicapped People" are available upon request from Educational Testing Service, Research Publications Unit-Room T143, Princeton NJ 08541:

#1 Bennett, R., and Ragosta, M. A Research Context for Studying Admissions Tests and Handicapped Populations, 1984. (ETS Research Report 84-31)

This is the first of a series of reports emanating from four year research effort to further knowledge of admissions testing and handicapped people. The authors describe the legal and educational issues that gave rise to this research and the major questions to be addressed. They discuss the distinguishing characteristics of different types of disability and the complex definitional problems that hamper any simple method of classifying examinees by type of handicap.

#2 Bennett, R., Ragosta, M., and Stricker, L. The Test Performance of Handicapped People, 1984 (ETS Research Report 84-32)

The purpose of this report was to summarize existing research information concerning the performance of handicapped people on admissions and other similar tests. As a group, handicapped examinees scored lower than did the nonhandicapped. Among the four major groups examined, physically handicapped and visually impaired examinees were most similar to the nondisabled population. Hearing disabled students performed least well. Available studies of the SAT and ACT generally supported the validity of those tests for handicapped people, but it was confirmed that research to date has been quite limited and has not addressed many important questions.

#3 Bennett, R., Rock, D., and Kaplan, B. The Psychometric Characteristics of the SAT for Nine Handicapped Groups, 1985. (ETS Research Report 85-49)

In this study the main finding was that with the exception of performance level, the characteristics of the Scholastic Aptitude Test (SAT) were generally comparable for handicapped and nonhandicapped students. The analyses focused on level of test performance, test reliability, speededness, and extent of unexpected differential item performance on the SAT. Visually impaired students and those with physical handicaps achieved mean scores similar to those of students taking the SAT in national administrations, while learning disabled and hearing impaired students scored lower than their nondisabled peers. Analysis of individual items revealed only a few instances of differential item performance localized to visually imparied students taking the Braille test.

ERIC

#4 Rock, D., Bennett, R., and Kaplan, B. The Internal Construct Validity of the SAT Across Handicapped and Nonhandicapped Populations, 1985. (ETS Research Report 85-50)

This study further investigated the comparability of SAT Verbal and Mathematical scores for handicapped and nonhandicapped populations. A two-factor model based on Verbal and Mathematical item parcels was posed and tested for invariance across populations. This model provided a reasonable fit in all groups, with the mathematical reasoning factor generally showing a better fit than the verbal factor. Compared with the nonhandicapped population, these factors tended to be less correlated in most of the handicapped groups. This greater specificity implies the increased likelihood of achievement growth in one area independent of the other and suggests that SAT Verbal and Mathematical scores be interpreted separately rather than as an SAT composite. Finally, there was evidence that the Mathematical scores for learning disabled students taking the cassette test may underestimate the reasoning ability of this group.

#5 Ragosta, M., and Kaplan, B. A Survey of Handicapped Students Taking Special Test Administrations of the SAT and GRE, 1986 (ETS Research Report 86-5).

Disabled people were surveyed to obtain their views on the appropriateness of special test accommodations available for the Scholastic Aptitude Test (SAT) and the Graduate Record Examinations (GRE). More than nine out of ten respondents reported satisfaction with special test accommodations. A minority experienced dissatisfaction with the level of test difficulty or about specific shortcomings associated with test administrations. In comparing SAT and GRE administrations with accommodations normally provided in college testing, respondents reported that the admissions tests were more frequently offered in special versions and with extra time than were college tests.

#6 Bennett, R., Rock, D., and Jirele, T. The Psychometric Characteristics of the GRE General Test for Three Handicapped Groups, 1986. (ETS Research Report 86-6).

This study investigated four psychometric characteristics of the GRE across handicapped and nonhandicapped groups: score level, reliability, speededness, and extent of unexpected differential item performance. Results showed the performance of visually handicapped students to closely approximate that of nonhandicapped examinees, while physically handicapped students performed substantially lower. Indications of speededness were suggested for those handicapped groups taking standard as opposed to special administrations. There was no evidence of higher or lower performance on any category of items on the GRE General Test than total score would indicate, suggesting that the different item categories operate similarly for handicapped and nonhandicapped groups.

#7 Rock, D., Bennett, R., and Jirele, T. The Internal Construct Validity of the GRE General Test Across Handicapped and Nonhandicapped Populations, 1986. (ETS Research Report 86-7).

The comparability of General Test scores for handicapped and nonhandicapped groups was investigated through confirmatory factor analysis. A three factor model was posed and tested for invariance across groups. The model provided a good fit in the nonhandicapped population, a moderately good fit for visually impaired students taking the General Test under standard conditions, and the least adequate fit for visually impaired students taking the large-type edition and physically handicapped students taking the standard test. For these latter two groups, differences in internal structure were traced to the Analytical scale, whose scores appeared to have a different meaning from those for nonhandicapped students.