THE PREDICTION OF TIME SCORES ON ACHIEVEMENT TESTS FROM ACADEMIC VARIABLES

Thesis for the Degree of Ph. D.
MICHIGAN STATE UNIVERSITY
BRUCE GLENN ROGERS
1968

This is to certify that the

thesis entitled

THE PREDICTION OF TIME SCORES
ON ACHIEVEMENT TESTS
FROM ACADEMIC VARIABLES

presented by

Bruce Glenn Rogers

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Counseling, Personnel Services and Educational Psychology.

Robert L Ebel

Date Feb 7,

ABSTRACT

THE PREDICTION OF TIME SCORES
ON ACHIEVEMENT TESTS
FROM ACADEMIC VARIABLES

by

Bruce Glenn Rogers

The purpose of this study was to determine the predictability of time scores on power tests from common measures of academic achievement. The study also sought further evidence on the stability of time scores, the existence of a time factor in items from power measures, and the comparability of internal consistency measures on timed portions of power measures.

In nine different university courses, time scores were recorded during the final examinations. In three of these courses, time scores were also taken on the midterm tests. Product-moment correlations calculated between these two measures produced coefficients in the neighborhood of .50, which were substantially equal to those obtained between the achievement scores on these same examinations.

The remaining six of the nine courses mentioned above were composed of a relatively broad sample of university freshmen and sophomores. For the students in these courses, scores from five entrance examinations were obtained,

covering the areas of English proficiency, reading, verbal ability, general information, and numerical ability. In addition, the number of credits earned, number of credits transferred (from another institution), sex of the student, and grade point average were recorded. When the first and second powers of these predictor variables were entered into a multiple regression equation, they provided a useful degree of prediction of the time scores. As the terms were stepwise deleted, two effects were noted. First, for some of the variables the quadratic component proved to be a significantly better predictor than the linear component. Second, verbal ability emerged as the strongest predictor, aided by supressor variables. A number of the differences between prediction equations in different courses could be logically explained, while others appeared to be the result of sampling errors.

On one of the tests, the matrix of item scores and time scores was subjected to factor analysis. No strong factors emerged, thus yielding no evidence of a time factor among the items.

When KR20 reliability coefficients were compared with odd-even coefficients for timed portions of the tests, the former were found usually to be smaller, but not by any large differences. Neither type of coefficient was substantially

inflated above coefficients calculated on the total test. The results were interpreted to be the consequence of using power tests, in which the students felt little or no time pressure.

It was concluded that the stability of time scores and their predictability from other academic variables is sufficient to warrant further investigation of time score properties.

THE PREDICTION OF TIME SCORES
ON ACHIEVEMENT TESTS
FROM ACADEMIC VARIABLES

By

Bruce Glenn Rogers

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY
College of Education

Department of Counseling, Personnel Services, and
Educational Psychology

1968

ACKNOWLEDGMENTS

Throughout the course of this study, many people lent a cooperative hand, when not to do so would have seriously hindered its progress. Though trite, it is true that it is not possible to acknowledge by name all of these contributors. But recognition must be given to those most closely associated with the project.

Dr. Robert L. Ebel, my major professor, is due special recognition. His generous help in discussing problems, his promptness and diligence in reviewing the various drafts, and his overall efforts in expediting progress of the thesis are most sincerely appreciated.

Dr. Willard G. Warrington, and the entire staff of the Office of Evaluation Services, provided the means for collecting the data. Extensive use of test scoring services and access to student records data were essential to the feasibility of this study.

Drs. Robert C. Craig, Charles F. Wrigley, and John Wagner provided helpful suggestions and assistance from the inception to the end of the work.

Mention should be made of generous contributions by Dr. Kenton Terry Schurr and Mr. Frederick Dyer, who assisted in the test proctoring, and Mr. David J. Wright and Mr. Stuart W. Thomas, Jr., who gave valuable suggestions in the computer analysis.

The writer expresses appreciation to the instructors and administrators responsible for granting permission to gather data during their examinations. To the students in those classes, who provided the basic data of the study. And to the numerous other unnamed individuals, who offered helpful suggestions, provided services when requested, or otherwise gave assistance where necessary.

Finally, much credit is due Zina G. Rogers, my wife, for the recording of data, typing of drafts, and steadfast encouragement throughout the entire venture.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table Page

Table Page

Table Page

LIST OF APPENDICES

CHAPTER I

INTRODUCTION

The Purpose and Significance of the Study

Since the earliest beginnings of psychological measurement, testers have observed with interest the differences in work rates of students. Investigators have speculated on the correlations between IQ and length of testing time, speed and comprehension in reading, speeded and power scores, etc., and although the evidence showed some relation, it also showed that a rate of work measure was a poor substitute for a power measure.

Consequently, the common practice of setting generous time limits (allowing at least ninety per cent of the examinees to finish) has continued with little opposition. In the great majority of testing situations no time measures of any type are collected. It is this latter omission which has prompted the present study. For the reliability of the time scores is quite unlikely to be less than the correlation between time scores and power scores. And if time scores have some useful degree of stability, they are likely to be found related to other aspects of achievement such as reading, general scholastic ability, etc. It will therefore be the purpose of this study to determine, within the confines of available data, some of the properties and predictors of time scores.

The significance of this study lies in the importance of determining and controlling the various factors which are predictive of time scores on power measures. Such information could be useful to both the student and the test constructor. To the student, it might prove helpful in suggesting efficient study and test taking procedures. To the test constructor, it could be useful in making decisions concerning test content. For example, if a measure of reading ability proves to be highly correlated with the time scores and total scores of a certain test, the test constructor may desire to modify the reading level so as to include more content and still remain within the time limits imposed by the practicalities of administration.

For the convenience of presentation, the study will be divided into four subproblems. The remainder of this chapter will be devoted to an introductory discussion of each.

The Problems Explored

Problem One: The Stability of Time Scores

Stability is certainly a necessary condition for establishing the usefulness of any type of score, and several studies have devoted some attention to the problem of time score reliability. In addition, the meaningfulness of a variable might be further increased by examining the extent to which it is related to other variables. Most commonly, investigators of time scores have sought to determine a relationship with the total score on the examination but the results are more suggestive of hypotheses than definitive conclusions.

While teaching a course in psychological testing, Freeman (1923) recorded the order in which his students returned their examination papers. In comparing the essay midterm and the multiple choice final, he found these rank orders to correlate about .50 while the total test scores correlated about .55. However, the correlation between order of finish and total score proved to be only about -.12 on both tests, and he concluded that little relation existed between these variables. Nevertheless, on the basis of the .50 correlation between the orders, he argued that there was something operating with sufficient reliability to merit further investigation.

In a monograph entitled "A Study of the Consistency of Rate of Work," Dowd (1926) attempted to disprove the belief that slowness in one aspect of a person's behavior (e.g. walking) was predictive of slowness in other aspects (e.g. performance on a job). She administered speeded tests in multiplication, writing alphabetic characters, etc., to 165 sixth graders and investigated the correlations between them. Since they were moderatley low (ranging from .15 to

.87, with the preponderance of coefficients at the low end), she concluded that there was no general speed factor. However, it is apparent from her study that she was convinced in advance that a speed factor would not be found and hence, may have been inclined to discount the correlations unduly.

Ebel (1947) measured the response time for each student to each item and found that with certain types of examinations this information could be profitably used in item selection and in setting test-time limits. Later (1954), while administering entrance placement examinations, he had the students record the number corresponding to the item on which they were working when one-half, three-fourths, and five-sixths of the time had elapsed. (Students were informed of this in advance and told to do the items in order without jumping back.) The half-period rate scores correlated with total accuracy scores in the neighborhood of .30, but with grade point average the correlations were near zero. He concluded that "it does not appear likely that the inclusion of rate scores would contribute much to the prediction of academic success" (p. 27).

Burak (1967) reported the rank-difference correlations between total score and time of completion on two tests in each of two psychology courses. The values were not significantly different from zero. He remarked that there were too many confounding variables but did not pursue the topic further.

In sum, time scores have been studied with respect to their stability and relationships with total scores, but the evidence does not lend itself to firm conclusions. It does, however, seem to suggest that there is a positive relationship between time scores taken on different occasions on the same group, and that this relationship is stronger than that between time scores and total test scores. The present study attempts to add to the available evidence by comparing the results from three different courses. The problem was formulated as follows:

> What degree of correlation exists, within a given university course, among time scores taken on the midterm and final examinations? How does it compare with the relationship between total scores on these same tests? If two time measures are taken relatively close together during the same test, how much variability will exist in their differences?

Problem Two: The Prediction of Time Scores

Heretofore, investigators have used the time score as an independent (or predictor) variable to account for variation in the dependent variable (usually an achievement measure). Even when the problem is cast in terminology other than that of dependent and independent variables, it is clear from context that the investigators were working and thinking along these same logical lines. Research on speeded tests is often of this type and has resulted in a sizable body of literature (e.g. see Morrison, 1960).

The design of the present study departs from those of previous investigations by identifying the time score with the dependent variable and the achievement measures with the independent variables. Essentially, the justification for this procedure rests on the assumption that a better understanding of the time score variable will accrue by attempting to maximize the proportion of its variability which can be accounted for by other measures. When it is related to other variables by using it in the role of a dependent variable as well as an independent one, the time score can be more firmly tied in the nomological net of test theory and thus increase both its empirical and theoretical import (Hempel, 1952, pp. 39-50).[1]

Although not addressing themselves directly to the topic of this study, several investigators have reported results relevant to the problem. As students finished their final examinations in elementary psychology, Briggs and Johnson (1942) had them place their papers in order on a pile. This pile was divided into thirds, and the results of an analysis

[1] While the intercorrelation matrix of the variables contains all the information that can be gained from a regression analysis, any one regression analysis does not exhaust this information. For example, Barch (referred to later in this section), using a time score as an independent variable, was able to show that it contributed only a small amount, over the other variables in his study, toward the prediction of grade point average. But one could not derive from those results alone the predictability of that time score from the other variables. Hence, there remained in the correlation matrix information which was not extracted.

of variance on the total scores proved statistically significant. When the means of the three groups were plotted against time, they formed a U-shaped distribution (with the early finishing group being the highest of the three). By performing an analysis of covariance, using IQ as the covariate, the investigators demonstrated that the higher IQ of the early group was sufficient to account for their higher total scores. The difference between the middle and late groups, they reasoned, was to be explained by the persistence of the latter.

Blumenfeld and Berry (1965) obtained time scores and total scores for a test given to 249 students in introductory psychology. After converting both sets to stanines, they divided the scales into thirds (each containing three stanines) and ran a 3x3 Chi-square analysis. In the authors' opinion, this data also supported the hypothesis that extreme time groups tend to get higher scores, although the results were not statistically significant. Both of the above studies, therefore, suggest that certain cognitive variables may bear a quadratic relationship to time scores.

Probably most closely related to the statistical methodology of the present investigation was the work of Barch (1957). He gathered time scores (referring to them as "departure times") from college students completing final examinations and sought to evaluate their importance in predicting academic achievement. Beginning with several entrance measures commonly used for prediction, he found that the accuracy of predicting grade point averages and final examination scores was slightly improved by the addition of the time scores.

The present study will employ the method of regression analysis[2] for the prediction of time scores and will seek to identify those independent variables which show significant relationships with the dependent variable. The specific problem investigated was:

> To what extent can time scores be predicted from measures commonly used in academic institutions? Can the predictions from linear components be improved by the use of quadratic terms? Can a reduced set of independent variables be found without serious loss in predictive power? How does the composition of such reduced sets vary across courses?

Problem Three: The Search for a Time Factor

Ascertaining the degree of speeding of a test is a common area of investigation in test theory. Gulliksen (1950) and Cronbach and Warrington (1951) represent only three of the many investigators who have studied this topic. It would seem reasonable that speeded tests would be intercorrelated as a result of measuring a common property and,

[2] Strictly speaking, the data were subjected to "correlational analysis" since the independent variables are random rather than fixed, although the statistics were calculated on a computer program written for regression analysis. Both terms are frequently used interchangeably in practice (Cooley and Lohnes, 1962, p. 31).

if so, a factor analysis might yield a factor which could be interpreted as a "time factor." Along these lines, Lord (1956) administered a number of speeded and unspeeded tests on vocabulary, spacial relations, and arithmetic reasoning to 649 freshmen at the United States Naval Academy at Annapolis, Maryland. He then combined these scores with course grades, factor analyzed the whole group, and divided the obtained oblique factors into three categories which he labeled as "level factors," "speed factors" and "grade factors." ("Level factors" included those which tended to relate to the level of attainment on unspeeded variables, "speed factors" were those related to rate of work, and "grade factors" were those related to grade point average.)

Since Lord claimed to find speed factors among his tests, the logic of his study might be extended to inquire if a time factor would emerge from a factor analysis performed on the individual items of a test combined with the rate score.[3] Thus, the set of items with high loadings on such a factor would tend to be predictive of the rate score. The major value of their identification would appear by investigating

[3] Because the values of reliability coefficients calculated for tests usually exceed those calculated for individual item scores, one cannot infer that the results obtained by factor analyzing tests will necessarily be obtained by factor analyzing individual item scores. We recognized this as a problem and realized that it decreased the probability of finding a time factor, but felt the question nevertheless warranted empirical evidence.

their distinguishing properties--for example, comparing their difficulty and discrimination indices with those of the remainder of the items. As the relations between these variables were determined, they would perhaps lead to a better understanding of the relations between time scores and measures of cognitive processes.

The third purpose of this study was to empirically investigate these relationships. The problem was stated as follows:

> Is there a "time factor" which can be identified in the responses to a set of items from a test? If so, can the items with high loadings be distinguished from the remainder of the test items on the basis of item discrimination and difficulty?

Problem Four: A Comparison of Two Measures of Consistency on Timed Portions of a Test

The procedure for collecting data for the second part of Problem One included plans to have the students indicate the items on which they were working at 40 and 45 minutes. It was expected (and later confirmed) that the students would, for the most part, proceed sequentially, item by item, and then check their work at the end. Accordingly, if the tests were treated as if all the items following the time mark were blank, the results should not depart far from those that would have been obtained had the papers been collected when the time signal was given. Since it was to be expected that internal consistency estimates of such a timed portion would be inflated, it was logical to compare the values of commonly used indices.

It might be inferred that the mathematically expected value for an odd-even coefficient would be equal to the value given by the well-known Formula 20 (KR20), developed by Kuder and Richardson (1938), since the latter is the average of all possible split-half coefficients for the test (Cronbach, 1951). However, the odd-even split is a special case, much more likely to yield two equivalent tests than is some split half-taken at random. Cronbach and Warrington (1951) obtained some data showing individual item times for a group of items completed by 36 high school students, which tended to support the hypothesis that the KR20 coefficients would be less than the split-half coefficients. However, as they pointed out, "it should be noted that our sample is small, so that our results are markedly influenced by sampling error" (p. 178). Later, Cronbach (1951), after examining several hypothetical cases, suggested that "for certain common types of tests, there is likely to be negligible variation among split half coefficients. Therefore, [alpha], the mean coefficient, represents such tests as well as any parallel split" (p. 319).

The fourth purpose of the present study was to investigate these two commonly used measures of internal consistency when applied to timed sections of a reasonably large sample of achievement scores. The problem was stated as follows:

> If KR20 coefficients are calculated for timed portions of a test, will they be inflated to the same degree as odd-even coefficients?

Limitations of the Study

The present study was conducted to relate the overall time for the completion of professionally made examinations (at the university undergraduate level) to selected academic variables. In addition, it sought to investigate measures of stability and consistency of several of the variables. It did not consider the ability of the instructor, the relative time spent on each item, nor the rate of work on different tasks by the same student. Because we desired to collect the data in actual academic settings, it was necessary to circumscribe the study to omit these and similar interesting problems.

CHAPTER II

DESIGN AND PROCEDURES

The Sample

At Michigan State University, all undergraduate students are required to complete four year-long courses in general education (offered by the University College), unless they have comparable transfer credits or pass special examinations in the areas. Because of the large numbers of students, each of the three parts in every course is offered every term, but we elected to take the sample from the "on-term" groups, e.g. those enrolled for the third part during the Spring (third) term of 1967. "Off-term" groups may have disproportionate numbers of transfer students (particularly those entering at mid-year), repeats (usually from failures), and waivers (those who are ahead of sequence by passing a waiver examination for an earlier course).

Arrangements were made to collect data from about 200 students in each course. The respective department chairmen recommended two examination groups of about 100 students each (only one group of about 200 in Freshman English), basing their choice mainly on adequate room conditions and the likelihood of cooperation of the proctors. All of the proctors readily agreed to participate. Similarly, arrangements were made to collect data in several courses outside University College. Table 2.1 lists the titles and abbreviations of the courses and Appendix A gives a brief description of each.

Instrumentation

The Examinations from which Time Scores were Obtained

Time scores were secured from the achievement examinations used in nine subject matter areas. The following is a brief description of the nature of these instruments.

Four of the measures used in the study were final examinations in the basic courses of University College. Each examination for a University College course begins as a series of multiple choice items written and assembled by an examiner holding a joint appointment with the Office of Evaluation Services and the department for which the examination is being written. After being reviewed and modified by an examining committee in the respective department, it becomes "Form A" and a scrambling of the items and alternatives produces "Form B." The University College examinations, like all other finals used in this study, purport to be power tests, as very few students fail to finish before the allotted time of two hours.

Table 2.1 Titles and Abbreviations of the Courses.

Abbreviation	Title	N[a]
ATL	American Thought and Language	225
NS	Natural Science	192
SS	Social Science	182
HUM	Humanities	146
ED200	Individual and the School	107
MATH	Foundations of Arithmetic	154
ED465	Introduction to Measurement and Evaluation in the Classroom	144
ED465SS	Same as ED465 but taught in Summer School	36
ED865	Psychological Measurement and Test Interpretation in Education	47
ED982	Seminar in Experimental Design	50
	TOTAL	1283

[a] Total number of test papers used in the study, although not all test papers were usable in every analysis. The results section will indicate the exact number of students used in each analysis.

A similar method is used to produce an ED200 examination. The midterm, as opposed to the final, is a set of 45 items administered in 50 minutes and thus generates a somewhat speeded atmosphere.

The two instructors in MATH wrote their own multiple choice examination. Since two of the items proved to have more than one correct answer, only 80 of the original 82 items were used in the reliability calculations.

The ED465, ED465SS, ED865, and ED982 classes also received instructor-made tests. In ED465, the Spring term class responded to true-false items, while the Summer School group (taught off-campus by two other instructors) was given multiple choice items on both midterm and final. The final examination in ED865 and the midterm in ED982 were likewise composed of multiple choice items but the ED982 final was a set of written problems.

That these examinations are typical of high quality measures is also attested to by the internal consistency coefficients shown in Table 2.2. Those for the University College courses were based on a random sample of about 1000 papers from each course and therefore contained some, but not all, of the papers from the sections in which time scores were recorded. The remaining indices were computed from the total group of test papers from each course, but in ED200 (both midterm and final) this included many others besides those with time response data.

Orientation Tests

Among the variables used to predict time scores were five measures of academic aptitude. These tests, usually given during the Freshman Orientation Week, are commonly known as Orientation Tests.

Table 2.2 Internal Consistency (KR20) Coefficients for the Examinations.[a]

Test[b]	N	No. of Items	KR20
ATL	1000	101	.786
NS	1149	100	.885
SS	1000	100	.858
HUM	1001	129	.872
ED200-(Midterm)	631	45	.586
ED200	680	80	.771
MATH	139	80	.873
ED465	136	120	.897
ED465SS	37	65	.779
ED865	42	100	.895

[a] Reliability coefficients were not available for the ED465SS midterm, the ED982 midterm nor the ED982 final.

[b] In ATL, NS, SS, and HUM, Form A of the test was used.

The MSU English Test (ENG) is composed of 38 objective test items and was designed to identify students deficient in English proficiency (who then must complete the Preparatory English Program before enrolling in the ATL sequence). Since its adoption in 1963, all new Freshmen and those transfer students who have not fulfilled the ATL requirements have been required to take it. One measure of its quality is shown by its reliability coefficient (KR20) of .79, as computed from 964 papers from the 1967 Summer orientation clinics.

The 1963 form of the MSU Reading Test (READ) presents the student with 50 objective items to measure his skill in interpreting reading passages representative of textbook materials in several areas. Its internal consistency (KR20) was estimated at .81 using 965 papers from the 1967 Summer orientation clinics.

Students are also required to take either the Mathematics Placement Test or the Arithmetic Test, depending on whether or not they plan to enroll in a course in the Department of Mathematics. Because of this option, each student had a blank on one or the other of these variables. But since the variables in this study were to be used in a regression analysis, which requires complete data on every individual, it was decided to delete both the Mathematics and Arithmetic scores rather than lose a large part of the sample.

General measures of scholastic aptitude were obtained from the College Qualification Tests--Form C (hereafter referred to as the CQT). The Verbal section (V), composed of 75 vocabulary items, is intended to predict success in courses emphasizing the language arts. Consisting of 50 items on conceptual skills in Algebra and Geometry, the Numerical test (N) was designed to predict success in scientific areas. Serving as a supplementary contributor to V and N is the Information (I) test, half of whose 75 items are on science and the other half on social studies. A general indication of the quality of these tests is given by the reliability coefficients reported by the authors. Corrected odd-even indices and alternate form indices are presented in Table 2.3.

Other Variables

In addition to the Orientation Tests, four other indices were secured from the student records to be used as possible predictors of the time scores.

1. Sex, coded Male = 1, Female = 2.

2. Transfer credits (TRANS), the number of credits accepted at MSU in term hours.

3. Credits earned (CRED), the number of credits earned at MSU.

4. Grade point average (GPA), based on the credits earned at MSU with A = 4.

Table 2.3 Reliability Coefficients of the CQT for College Freshmen.

Coefficient	Sex	N	Verbal	Numerical	Information
Corrected Odd-Even (form C)	M	416	.95	.89	.86
	F	363	.95	.89	.87
Alternate Froms (forms B & C)	M	227	.89	.86	.80
	F	194	.84	.85	.79

From Bennett, et al., 1961, p. 53.

Procedure

This section will describe how the various measures, which have been mentioned above, were collected and transformed to create the variables as used in the statistical analysis. It is a necessary technical part of the report, but the reader may pass to the next chapter, if he desires, without loss of continuity.

Time Measurements and Student Master Tape Records

In order to keep the testing situation as normal as possible, the timing proctor aided the regular testing proctor in passing out materials and making other preparations for the examination. The regular proctor then announced, in his own words, the instructions for the experiment, which usually consisted of rephrasing the ideas given in the "Instructions to Proctors" (see Appendix B). In most of the courses the timing proctor asked the students to circle, on the answer sheet, the item number on which they were working at the end of 40 minutes. At the end of 45 minutes they were asked to put an X through the item number.[1] Since the times had been chosen short enough that no one could be expected to finish the test, these indices were designed to given an indication of rate of response.

Several methods were used in obtaining the total time scores. For the ED200 midterm (the first data collected), two sets of cards, 5" x 8", were numbered consecutively 0 through 9. When stood upright side by side, the topmost card on the right set formed the units digit. Beginning at the end of 30 minutes the cards were flipped every 30 seconds, and the students were instructed to write the number showing when they were ready to hand in their test. No proctor was available to record times in ED982, and so the instructor was asked to place the papers on a pile in the order they came in, thus yielding a set of rank order scores.

[1] The exceptions were NS, one section of HUM, ED465SS, and ED982 (both midterm and final).

By the time of the final examination week, the timing apparatus had been improved. The numerals were six inches high and the lines one-half inch wide. They were cut from black poster paper, pasted on 5" x 8" sheets of cardboard, and mounted on a masonite frame with rings at the top, which allowed them to be flipped over. At the end of 50 minutes they were changed each minute and again the students were asked to write down the number when they completed their examination. This method was used in ATL, SS, ED200 and ED465.

The cooperation of the Natural Science Department was obtained under the condition that the students not be aware, during the test, that an experiment was being conducted. Therefore, the following method was used:

Eighty-one sheets of paper were labeled consecutively 0 through 80 and at 50 minutes were changed every minute. They were placed in front of the person timing the test, so that as a test paper was brought forward, he wrote the current number on the top of the test. (The numbered pile was only to help the timer keep his numbers straight.) This method proved satisfactory (i.e. the papers came in slow enough that the time scores were not appreciably affected by time standing in line) and was subsequently used in MATH, ED465SS (midterm and final), ED865 and ED982.

After the tests were scored, a card deck containing 1040 student numbers from the University College courses, ED200, and MATH was sent to the Registrar's Office. It was used to retrieve the orientation test scores, sex, transfer credits, credits earned, and grade point average from the student master data tape. Because the remaining classes consisted almost solely of graduate students, none of whom had orientation test scores, no retrieval cards were prepared for them.

Transformation of Orientation Test Scores

The orientation test scores for each entering student are recorded on the student master data tape as percentile scores, based on that year's entering class. Therefore, to obtain the raw scores for a student it is necessary to know his year of entrance and the conversion tables for that year. Since student numbers are assigned in blocks each term, the year of entrance was easily obtained. Only forty-one of the students were admitted prior to the Fall of 1964, and of these, most had incomplete orientation test scores. Finding that the available orientation test conversion tables were complete only back to 1964, we decided that we could begin there without appreciable loss of accuracy.

In the process of this conversion from percentiles to raw scores, two sources of error appeared. First, more than one raw score had sometimes been assigned the same percentile score, particularly at the extremes. In such cases, a conservative estimate was used, by assigning the score closest to the mean as the percentile equivalent. Second, a few percentiles appeared from Fall 1966, which were not used on the original raw score to percentile conversion. It was explained that they might have been erroneously converted using the previous year's transformation tables. Considering that the tables were quite similar from year to year, we elected to use the closest approximation.

A computer program was then written which produced the transformed values on punched output cards.

Transformation of Item Responses

After the test papers were scored, the computer produced as output a set of cards punched with 1 or 0 for correct or incorrect responses, respectively. As was mentioned in the Instrumentation section, the University College courses had two forms of the tests, Form B being a scrambled version of Form A. This posed a problem for the factor analysis in Problem Three, since in each course, there were less students who had taken any one form than there were items in the test. In order to increase the sample size, it was desirable to use the students from both forms. Accordingly, a correspondence was made between the items on the two forms and a computer program was written which punched the items scores from Form B into the order of Form A.[2]

In their final form, the data for each student consisted of a card containing the student master tape information, the time responses, and final score, and subsequent cards containing item response data. In ED465, ED465SS, ED865, and ED982 the master tape data were blank.

Summary

In nine different courses at Michigan State University the time of test completion was measured during the final examination. In most cases, the item numbers on which the students were working at 40 and 45 minutes were also procured. In six of the courses, composed of a relatively broad sample of university freshmen and sophomores, the Registrar's Office supplied several measures from the permanent records. These data included five aptitude measures, whose scores were then converted from percentiles to raw scores. Other data transformations included scoring

[2] It might be argued that the items on Form B would have been answered differently had they appeared in the order of Form A. In a study of the problem of item rearrangement using Verbal and Mathematics tests, Flaugher, et al. (1966), found some differences in the Verbal tests. They suggested that "A possible explanation for these results is that in some of the Verbal arrangements relatively easy items occurred last and were not reached by some students" (p. 20) (quoted by permission of the authors). Since the papers in the present study indicated that all the students reached the end of the test, it was inferred that the items could be rearranged with few adverse effects.

the items and rearranging their order so that both forms would have the same item sequence for use in the factor analysis. These three types of measures--time scores, records from the Registrar, and item responses--formed the basic data for the subsequent analyses.

CHAPTER III

RESULTS

Problem One

The Stability of Time Scores

> What degree of correlation exists, within a given university course, among time scores taken on the midterm and final examinations? How does it compare with the relationship between total scores on these same tests? If two time measures are taken relatively close together during the same test, how much variability will exist in their differences?

Table 3.1 presents the composite intercorrelation table for the variables midterm time, midterm score, final time, and final score. Within each block are shown the correlations in ED200, ED465SS and ED982 in that order. Three results stand out markedly:

1) The three courses give reasonably consistent results

2) The correlations between time variables and score variables are the smallest correlations in the table

3) Midterm and final times are correlated about the same degree as midterm and final scores.

One could attempt to explain the differences between courses in terms of differences in methods of time measurements[1]; however, considering the very small differences relative to the standard errors, it seems unwise to do so.

Table 3.1 Intercorrelations of Time and Accuracy Measures for Three Tests.

	Midterm Time	Midterm Score	Final Time
Midterm Score	.123 -.159 -.105		
Final Time	.780*** .554*** .416**	-.003 -.119 -.331*	
Final Score	.222 -.158 -.015	.567*** .676*** .503***	.173 -.013 -.184

Each block of three indices contains the Pearson r coefficients for ED200 (N = 44), ED465SS (N = 36), and ED982 (N = 50), in order. In ED982, the correlations involving final time are based on an N of 45.

*, **, and *** indicate correlations significantly different from zero at the .025, .005, and .0005 levels, respectively.

To obtain information relative to the third question, the 40 minute item numbers (40I) were compared with the 45 minute item numbers (45I). Table 3.2 contains several descriptive statistics concerning these measures. No item numbers were recorded in NS, ED465SS, or ED982.

Again we see reasonably consistent results between courses. Within each course, the students were fairly well spread out, as shown by the standard deviations of the two item number indices. Students seemed to progress at a fairly even rate as evidenced by the small standard deviation of the variable 45I-40I.[2] The stability of the two time measures is further accentuated by the high correlation between them.

[1] Since some of the data are rank ordered while other are considered equal interval, it might be expected that different types of correlations would be employed. However, as Guilford (1965) points out: "The rank difference correlation is rather closely equivalent to the Pearson r, numerically . . . on the average r is slightly greater than [rho] and . . . the maximum difference . . . is approximately .02, when both are near .50. We may therefore treat an obtained rho as an approximation to r" (p. 307). Since it is hardly conceivable that differences of these magnitudes could change the interpretation of the results, we elected to use the Pearson r on all the data.

[2] The reader may well wonder at this point, in inspecting Table 3.2, about the apparently discrepant results in the last two columns of ED200. A discussion of this appears later in Chapter IV (p. 59).

Table 3.2 Descriptive Statistics Comparing the 40 Minute Item Number (40I), the 45 Minute Item Number (45I), and Their Difference.

Group	N	40I		45I		45I-40I		$r_{40I,45I}$
		Mean	S.D.	Mean	S.D.	Mean	S.D.	
ATL	215	44.37	12.50	50.03	12.93	5.66	3.34	.966
SS	175	48.20	12.62	53.97	13.28	5.77	2.39	.984
HUM[a]	69	48.68	12.40	54.77	13.20	6.09	2.03	.989
ED200	102	47.77	12.66	52.79	11.74	5.02	9.24	.715
MATH	136	27.42	5.62	31.24	7.05	3.82	2.42	.952
ED465[b]	136	63.26	20.88	73.21	22.70	9.95	4.19	.985
ED865	40	60.95	11.39	66.45	12.39	5.50	2.72	.977

[a] Only one section recorded item numbers.

[b] Item numbers were called for at 35 and 40 minutes.

Problem Two

The Prediction of Time Scores

To what extent can time scores be predicted from measures commonly used in academic institutions? Can the predictions from linear components be improved by the use of quadratic terms? Can a reduced set of independent variables be found without serious loss in predictive power? How does the composition of such reduced sets vary across courses?

For this problem a least squares fit was sought for ten independent variables (listed with abbreviations in Table 3.3 and previously described in Section 2.2) using the time score as the dependent variable. From an inspection of the intercorrelation tables of each of the six groups (see Appendix C) it was clear that time scored did not show a strong linear relationship with any of the other variables. Judging from the results of several other studies, as previously mentioned, it seemed profitable to investigate the degree to which quadratic relations would improve the prediction.

For each group, the ten independent variables[3] and their squares[4] were entered into a multiple regression

[3] If we assume a multivariate normal model, random independent variables may be used in the regression equation (Smillie, 1966, p.41; Anderson, 1958, p.27). Ezekiel and Fox (1959, pp. 13-14) pointed out: "If random errors are associated with [all of the] variables simultaneously, their effects [tend] to reduce [the multiple correlation] below the true value." To test this effect, they introduced relatively large random errors (by dice throws) into a set of variables, but found relatively small changes in the multiple correlation. "It may be slightly reassuring to know," they concluded, "that observational errors even as large as those just considered still modify the regression results as little as these have been seen to do" (p. 316). Hence, the obtained values in this study should be considered as conservative estimates of the population parameters.

[4] Random errors have the same type of effect in curvilinear correlation that they do in linear regression" (Ezekiel and Fox, 1959, p. 316).

Table 3.3 Names and Abbreviations of the Variables Used in the Least Squares Analyses.

Name	Abbreviation[a]
MSU English Placement Test	ENG
MSU Reading Test	READ
College Qualification Test - Verbal	CQT-V
College Qualification Test - Information	CQT-I
College Qualification Test - Numerical	CQT-N
Sex	SEX
Transfer Credits	TRANS
Credits Earned at MSU	CRED
MSU Grade Point Average	GPA
Score on Test	SCORE
Time when Test was Turned in to Proctor	TIME

[a] Squared terms will be denoted with a "2", e.g. ENG^2.

equation[5] (SEX was represented only by its first power, since it is a dichotomous variable). Of the nine squared terms, that one whose omission would affect the multiple correlation coefficient (R) the least was deleted and a new regression analysis computed (Rafter and Ruble, 1967). This procedure was repeated until all the squared terms were deleted.[6] The beginning and final multiple correlation coefficients (R) are shown in Table 3.4.

Since we desired to keep quadratic terms if, and only if they substantially improved prediction, criteria had to be established. As each squared term was deleted, the remaining beta weights of the squared terms were examined and when all of them showed significance in the neighborhood of the .10 level, those variables were considered to have useful quadratic terms. A second, more stringent

[5] In the past, the method of orthogonal polynomials has frequently been used in such procedures to reduce the calculations to manageable proportions. But ". . . orthogonal polynomials . . . may be dispensed with if a suitable regression program for a high speed computer is available. The successive powers of the observations on the independent variable may be simply generated as the initial data are being read, and the normal equations may be set up and solved in the usual manner" (Smillie, 1966, p. 80).

[6] Dr. Charles F. Wrigley has pointed out that one should interpret with caution results based on the type of deletion procedure used in this study. Because of sampling errors, correlated measures, and squared terms, the resulting beta weights might fluctuate widely in successive replications. Dr. Wrigley's current research is expected to shed additional light on this problem.

Table 3.4 Multiple Correlation Coefficients (R) using TIME as the Dependent Variable, at the Beginning and End of the Deletion of Squared Terms.

Group	N	Beginning R[a]	Final R[b]
ATL	221	.388	.354
NS	178	.491	.442
SS	160	.457	.395
HUM	134	.456	.379
ED200	85	.702	.572
MATH	127	.485	.408

[a] Includes both linear and squared terms.

[b] Includes only linear terms.

criterion was set by continuing the deletion of squared terms until those remaining showed significance at the .05 level.[7]

[7] Since correlated variables result in unstable beta weights these criteria could be called sufficient but not necessary conditions for identifying good predictors of the dependent variable.

The above procedure will be referred to as Part I of the deletion process. In Part II the process was continued by lifting the restriction on the linear terms. The same two significance criteria were still used, thus producing two solutions, one in which all variables were significant in the neighborhood of the .10 level or less, and the other in which all variables were significant at the .05 level.[8]

ATL

At each iteration of Part I of the deletion process, the term whose beta weight showed the largest significance value was dropped.[9] The beta weight of the last squared term to be deleted (GPA^2) had a final significance level of .211. Since none of the squared terms were near the .10 significance level, the deletion procedure was continued on the linear terms. When only three variables remained, the significance level of each was in the neighborhood of .10. Table 3.5 contains the results.[10] When only two

[8] This method would allow the possibility that on a certain variable, the linear term could be deleted and leave the quadratic term. Since the purpose, however, was to ascertain how the quadratic terms added to the prediction over the linear terms, the procedure was modified to retain any linear term whose associated quadratic component had met the significance requirement described in the preceeding paragraph.

[9] This is an equivalent way of expressing the procedure described above, i.e. the process of deleting that variable which reduces R the least (Rafter and Ruble, 1967, p. 12).

[10] The Analysis of Variance for Overall Regression is equivalent to a significance test on R, under a fixed effects model (Ruble, *et al.*, 1967, pp. 33-34). However, it is also equivalent to the significance test on R under the multivariate normal model (Graybill, 1961, p. 216; Hays, 1963, p. 567).

Table 3.5 Results of the Least Squares Deletion Routine on ATL Data When All Remaining Variables Were Significant in the Neighborhood of the .10 Level.

Analysis of Variance for Overall Regression

Source	SS	df	MS	F	sig.
Regression	4300.56	3	1433.52	8.57	.0005
Error	36300.40	217	167.28		
Total	40600.96	220			

R = .326

Relative Contributions of the Variables

Variable	Beta	S.E.	F	sig.
CQT-V	-.249	.070	12.69	.0005
CQT-I	.107	.070	2.36	.126
TRANS	-.221	.064	11.79	.001

variables remained, the significance levels were all below .05. These results are shown in Table 3.6.

Table 3.6 Results of the Least Squares Deletion Routine on ATL Data When All Remaining Variables Were Significant at the .05 Level.

Analysis of Variance for Overall Regression

Source	SS	df	MS	F	sig.
Regression	3905.96	2	1952.98	11.60	.0005
Error	36695.00	218	168.33		
Total	40600.96	220			

R = .310

Relative Contributions of the Variables

Variable	Beta	S.E.	F	sig.
CQT-V	-.207	.064	10.29	.002
TRANS	-.223	.064	11.99	.001

NS

In Part I (deletion of squared terms), CQT-V^2 proved to be the only term satisfying either criterion, and had a final significance level of .061. In Part II, the linear terms were deleted until the remaining terms had significance levels in the vicinity of .10, as shown in Table 3.7. But the quadratic term, along with three linear terms, had to be deleted to obtain the set of variables which were all significant at the .05 level (see Table 3.8).

SS

Two potentially useful squared terms were generated in Part I, namely, CQT-I^2 ($p = .12$) and GPA^2 ($p = .08$). In Part II, the more lenient criteria produced the results shown in Table 3.9 while the more stringent criteria produced those shown in Table 3.10.

HUM

After seven of the squared terms had been deleted in Part I, TRANS^2 and SCORE^2 remained, with significance levels of .122 and .015, respectively. Continuing to Part II with the liberal criterion, the results in Table 3.11 were obtained. Using the conservative criterion, only TOTAL^2 was used as a quadratic term and the deletion of linear terms produced the results in Table 3.12.

Table 3.7 Results of the Least Squares Deletion Routine on NS Data When All Remaining Variables Were Significant in the Neighborhood of the .10 Level.

Analysis of Variance for Overall Regression

Source	SS	df	MS	F	sig.
Regression	8545.01	6	1424.17	7.35	.0005
Error	33115.37	171	193.66		
Total	41660.38	177			

R = .453

Relative Contributions of the Variables

Variable	Beta	S.E.	F	sig.
CQT-V^2	1.091	.592	3.40	.067
CQT-V	-1.375	.593	5.37	.022
CQT-N	- .306	.082	13.91	.0005
SEX	.131	.075	3.03	.083
TRANS	.117	.069	2.87	.092
SCORE	.423	.084	25.34	.0005

Table 3.8 Results of the Least Squares Deletion Routine on NS Data When All Remaining Variables Were Significant at the .05 Level.

Analysis of Variance for Overall Regression

Source	SS	df	MS	F	sig.
Regression	6855.29	3	2285.10	11.42	.0005
Error	34805.09	174	200.03		
Total	41660.38	177			

R = .406

Relative Contributions of the Variables

Variable	Beta	S.E.	F	sig.
CQT-V	-.259	.072	13.15	.0005
CQT-N	-.352	.081	18.92	.0005
SCORE	.396	.083	22.73	.0005

Table 3.9 Results of the Least Squares Deletion Routine on SS Data When All Remaining Variables Were Significant in the Neighborhood of the .10 Level.

Analysis of Variance for Overall Regression

Source	SS	df	MS	F	sig.
Regression	7715.22	5	1543.04	6.30	.0005
Error	37721.18	154	244.94		
Total	45436.40	159			

R = .412

Relative Contributions of the Variables

Variable	Beta	S.E.	F	sig.
CQT-V	- .230	.088	6.78	.010
CQT-I^2	-1.218	.703	3.01	.085
CQT-I	1.099	.702	2.45	.120
GPA^2	1.574	.775	4.12	.044
GPA	-1.662	.778	4.56	.034

Table 3.10 Results of the Least Squares Deletion Routine on SS Data When All Remaining Variables Were Significant at the .05 Level.

Analysis of Variance for Overall Regression

Source	SS	df	MS	F	sig.
Regression	6597.09	2	3298.54	13.33	.0005
Error	38839.31	157	247.38		
Total	45436.40	159			

R = .381

Relative Contributions of the Variables

Variable	Beta	S.E.	F	sig.
CQT-V	-.261	.083	9.95	.002
SCORE	-.184	.083	4.98	.027

Table 3.11 Results of the Least Squares Deletion Routine on HUM Data When All Remaining Variables Were Significant in the Neighborhood of the .10 Level.

Analysis of Variance for Overall Regression

Source	SS	df	MS	F	sig.
Regression	4107.79	7	586.83	4.18	.0005
Error	17679.70	126	140.32		
Total	21787.49	133			

R = .434

Relative Contributions of the Variables

Variable	Beta	S.E.	F	sig.
READ	- .158	.097	2.67	.105
CQT-I	- .298	.096	9.74	.002
$TRANS^2$	- .409	.255	2.58	.111
TRANS	.290	.254	1.30	.256[a]
GPA	.202	.101	3.98	.048
$SCORE^2$	-1.933	.747	6.69	.011
SCORE	1.930	.746	6.70	.011

[a] See note 8 above.

Table 3.12 Results of the Least Squares Deletion Routine on HUM Data When all Remaining Variables Were Significant at the .05 Level.

Analysis of Variance for Overall Regression

Source	SS	df	MS	F	sig.
Regression	3223.68	4	805.92	5.60	.0005
Error	18563.81	129	143.91		
Total	21787.49	133			

R = .385

Relative Contributions of the Variables

Variable	Beta	S.E.	F	sig.
CQT-I	- .343	.087	15.63	.0005
GPA	.212	.102	4.33	.039
SCORE^2	-1.616	.743	4.73	.031
SCORE	1.554	.736	4.46	.037

ED200

From Part I, the four variables, $READ^2$, $CQT\text{-}V^2$, $TRANS^2$ and $CRED^2$, appeared to have useful quadratic components since their final significance levels were .068, .007, .038, and .025 respectively. The results from Part II using the .10 criteria are shown in Table 3.13 and those using the .05 criteria in Table 3.14.[11]

MATH

In Part I, the squared terms were deleted to leave only $SCORE^2$ with a significance level of .015, the results being the same using either criterion. In Part II, as the linear terms were deleted, the liberal criterion yielded the results shown in Table 3.15 while the conservative criterion generated those in Table 3.16.

Summary

Thus, it can be seen that in all cases, the time scores were capable of being predicted at above chance levels by the independent variables. Multiple correlation coefficients ranged from .39 to .70 when both linear and squared terms were used for all the independent variables. In each

[11] The AOV in Table 3.14 shows six degrees of freedom for regression because READ and TRANS were both retained by the computer (see note 8 above) even though they were very insignificant. For simplicity of interpretation, they were deleted from the "Relative Contributions of the Variables."

Table 3.13 Results of the Least Squares Deletion Routine on ED200 Data When All Remaining Variables Were Significant in the Neighborhood of the .10 Level.

Analysis of Variance for Overall Regression

Source	SS	df	MS	F	sig.
Regression	14504.09	9	1611.57	6.25	.0005
Error	19327.13	75	257.70		
Total	33831.22	84			

R = .655

Relative Contributions of the Variables

Variable	Beta	S.E.	F	sig.
$READ^2$	1.566	.855	3.35	.071
READ	-1.634	.836	3.82	.054
$CQT\text{-}V^2$	-2.870	.984	8.50	.005
CQT-V	2.434	.973	6.26	.015
SEX	.223	.091	5.95	.017
$TRANS^2$	.703	.293	5.77	.019
TRANS	- .567	.303	3.49	.066
$CRED^2$	-1.134	.454	6.24	.015
CRED	1.120	.465	5.80	.018

Table 3.14 Results of the Least Squares Deletion Routine on ED200 Data When All Remaining Variables Were Significant at the .05 Level.

Analysis of Variance for Overall Regression

Source	SS	df	MS	F	sig.
Regression	11765.93	6	1960.99	6.93	.0005
Error	22065.30	78	282.89		
Total	33831.22	84			

R = .590

Relative Contributions of the Variables

Variable	Beta	S.E.	F	sig.
CQT-V^2	-2.088	.814	6.58	.012
CQT-V	1.721	.816	4.45	.038
$CRED^2$	-1.160	.475	5.96	.017
CRED	1.100	.487	5.11	.027

See footnote 11 above.

Table 3.15 Results of the Least Squares Deletion Routine on MATH Data When All Remaining Variables Were Significant in the Neighborhood of the .10 Level.

Analysis of Variance for Overall Regression

Source	SS	df	MS	F	sig.
Regression	5257.66	4	1314.41	7.07	.0005
Error	22681.15	122	185.91		
Total	27938.80	126			

R = .434

Relative Contributions of the Variables

Variable	Beta	S.E.	F	sig.
CQT-I	.187	.094	3.92	.050
CQT-N	- .177	.101	3.06	.083
SCORE^2	-1.476	.495	8.89	.003
SCORE	1.781	.490	13.19	.0005

Table 3.16 Results of the Least Squares Deletion Routine on MATH Data When All Remaining Variables Were Significant at the .05 Level.

Analysis of Variance for Overall Regression

Source	SS	df	MS	F	sig.
Regression	4338.19	2	2169.10	11.40	.0005
Error	23600.61	124	190.33		
Total	27938.80	126			

R = .394

Relative Contributions of the Variables

Variable	Beta	S.E.	F	sig.
SCORE^2	-1.508	.495	9.29	.003
SCORE	1.791	.495	13.09	.0005

analysis, a reduced set of independent variables was formed by the deletion of non-significant terms, and the resulting multiple correlation coefficients ranged from .33 to .66 (for .10 solutions). The presence of both squared and linear significant predictors varied from course to course. Each of these findings will be discussed further in the next chapter.

Problem Three

The Search for a Time Factor

> Is there a "time factor" which can be identified in the item responses from a test? If so, can the items with high loadings be distinguished from the remainder of the test items on the basis of item discrimination and difficulty?

A factor analysis (Williams, 1967) using 225 observations, was run on the ATL items and time score. Because of the size limitations of the program, only 89 items (of 101) were used (every tenth item was deleted, up through item 98), the time score then becoming the 90th variable (the maximum allowed. Communalities were set to unity. Following a principal axis solution, both the Quartimax and Varimax rotations were performed with the Kiel-Wrigley criterion set at nine[12], and in both cases the last solution

[12] "The procedure is that successively larger number of factors (as ordered by eigenvalues-largest first) will be rotated until the solution finds a factor with fewer variables with highest loadings on it than the number [specified] . . . The procedure starts with two and adds one factor

satisfying the criterion contained seven factors.

In setting the value of the K-W criterion, we had to be cognizant of the amount of computer time required to rotate 90 variables. Since it appeared that it would be difficult to generalize from statistical comparisons on small sets of items, and since we wished to avoid unnecessary rotations, the criterion was set at nine, which was the largest value allowed in the computer program.

Reproduced in Table 3.17 are the first seven eigenvalues of the principal axis solution. The highest accounts for only 4.96 per cent of the total variance and all seven for only 20.05 per cent. For the first and last Quartimax rotations, the proportions of variance accounted for by each factor and the time score loading are contained in Table 3.18. Similar results for the Varimax rotation are also shown. In both cases, the time score loading on the second factor of the last rotation was the largest time score loading reported in any of the rotations.

The interpretation of these results, in a manner consistent with an affirmative answer to the questions posed

for each rotational solution . . . The higher the number [specified] the smaller will be number of factors extracted. If, for example, 9 is [specified] only factors with at least 9 variables loaded most highly on them will satisfy the criterion" (Williams, 1967, p. 4).

Table 3.17 The First Seven Eigenvalues of the Principal Axis Solution on ATL Data.

1.	4.4653
2.	2.4592
3.	2.3410
4.	2.2820
5.	2.2505
6.	2.1403
7.	2.1074
Sum	18.0457

Table 3.18 Proportions of Variance and Time Score Loadings on the First and Last Quartimax and Varimax Rotations on ATL Data.

First Rotation

Factor	Proportion of Variance		Time Score Loadings	
	Quartimax	Varimax	Quartimax	Varimax
1	.0438	.0409	-.1722	-.1878
2	.0332	.0360	.1260	.1013

Last Rotation[a]

Factor	Proportion of Variance		Time Score Loadings	
	Quartimax	Varimax	Quartimax	Varimax
1	.0303	.0303	-.0282	-.0281
2	.0263	.0264	.2616	.2626
3	.0328	.0326	-.0134	-.0076
4	.0353	.0350	.1603	.1622
5	.0261	.0262	.0036	-.0018
6	.0240	.0240	-.1819	-.1782
7	.0257	.0260	.0299	.0352

[a] The K-W Criterion was set at nine.

at the beginning of this problem, is difficult for several reasons: (1) the proportions of variance are so low that there is little difference from that which would be expected by chance alone; (2) in a practical sense, one is hesitant to infer a strong association between a variable and a factor on which it loads less than .50, and the time score loadings were much less than that; and (3) the time score has several factor loadings of intermediate magnitude rather than a single loading of high magnitude and the others of low magnitude.

Consequently, it was decided not to pursue Problem Three, but conclude that little information about time variables in these tests could be obtained by an analysis of the individual items.

Problem Four

A Comparison of Two Measures of Consistency on Timed Portions of a Test

> If KR20 coefficients are calculated for timed portions of a test, will they be inflated to the same degree as odd-even coefficients?

For each student who indicated his item number at 45 minutes, it was possible to generate two vectors, namely one containing his scores on all the items and another containing his scores on only those items completed when the 45 minute time was called. For each of the two matrices thus formed by the vectors of all students in a particular course, odd-even and KR20 coefficients were calculated. The results from all seven courses in which time measurements were taken are shown in Table 3.19.

Table 3.19 Odd-Even and KR20 Reliability Coefficients.

Group	N	Items in Test	45 min.		Total		Large Sample
			Odd-Even	KR20	Odd-Even	KR20	
ATL	111	101	.881	.826	.715	.780	.786
SS	84	100	.923	.914	.889	.880	.858
HUM	33	129	.860	.875	.896	.924	.872
ED200	103	80	.938	.906	.803	.846	.771
MATH	139	80	.843	.801	.860	.873	.873
ED465	137	120	.966	.942	.897	.897	
ED865	42	100	.915	.904	.864	.895	

"45 minute" coefficients were calculated using only those items completed at the time the 45 minute signal was given.

Odd-even coefficients have been corrected by the Spearman-Brown Prophecy formula (so named because it was reported simultaneously by both Spearman (1910) and Brown (1910)).

Large Sample coefficients are from Table 2.2.

It can be seen that six out of the seven 45-minute odd-even coefficients are larger than the corresponding 45-minute KR20 coefficients, though perhaps not to a degree which would be considered serious in most practical applications. For the total test, only one of the odd-even coefficients exceeds the corresponding KR20 coefficient. These small differences seem consistent with Cronbach's (1951) suggestion that the KR20 coefficient is a close approximation to any split-half coefficient.

Table 3.19 also contains the KR20 coefficients from Table 2.2 which were calculated on large samples of students from each course. The last two columns were thus determined on separate samples from the population and the differences reflect the sampling errors inherent in obtaining such reliability coefficients.

In summary, when odd-even coefficients were compared with KR20 coefficients, six out of seven were larger, but the differences were small. Further, there was no evidence that either was seriously inflated.

CHAPTER IV

DISCUSSION

Problem One

The Stability of Time Scores

The findings reported herein are substantially in agreement with those of Freeman's (1923) study which compared the results from a midterm and final examination. Although his results, based on only one course, could conceivably arise by chance, it is not likely that chance alone would produce the consistent results obtained in the present study in three different courses.

In comparing the results obtained by Ebel (1954)[1] with those of the present study, we note that he used a rate of work measure, somewhat comparable to the 45 minute item score in this investigation. His conclusion that such a measure is not particularly promising in predicting grade point average is supported by the data in Table 4.1, which show the correlations between 45I, TIME, SCORE, and GPA for ATL data. It appears that time of finish measures something more than rate of responding to individual items. Perhaps between two students who work with approximately

[1] See Chapter I (p. 4).

Table 4.1 Correlations in ATL between 45I, TIME, SCORE and GPA.

TIME	-.499		
SCORE	.008	-.042	
GPA	.074	-.044	.502
	.45I	TIME	SCORE

N = 211

the same speed and accuracy, one will turn in his paper after completing the last item while the other will spend time looking over the items and thus may increase his score. This hypothesis could not be explored further because no direct evidence was available to bear upon it, however, the lack of linear relationships and the tendency toward quadratic relationships, as discussed in Problem Two, would tend to give indirect support.

Some students jump back and forth among the items of a test. If this tendency was widespread we would expect it to be reflected in the difference between the 45 minute item number and the 40 minute item number (see Table 3.2). It seems reasonable to interpret the standard deviation of 45I-40I as indicating that the students responded to the items in numerical sequence, with few exceptions.[2]

Problem Two

The Prediction of Time Scores

The results reported in Chapter III give affirmative answers to the questions posed in this problem. The final equations produced multiple correlation coefficients which were in the neighborhood of .40 and significant at the .0005 level.

In several cases, quadratic terms proved to be significant predictors, as suggested by the results reported by Briggs and Johnson (1942) wherein they found that total score and test time had a curvilinear relationship. Table 4.2 summarizes the signs of the beta weights for the .10 level solutions reported in Chapter III. As expected, there is an interaction effect between the courses and the independent variables. Thus CQT-V, TRANS, and SCORE show

[2] Such an exception might be ED200 (see Table 3.2, p. 30). Since that test consisted of only 80 items, it is possible that at 45 minutes some of the students had finished all the items and had begun rechecking their work. This should have increased the variance of 45I, but interestingly enough, the 45I variance is smaller than the 40I variance. Another possibility is that the items were harder near the end of the test (in the sense of requiring more time to answer). The students near the end of the test would then be doing fewer items in the same amount of time than those who were still on the easier items. This would increase the 45I-40I variance and suppress both the variance of 45I and the 40I,45I correlation, which accords with the data in the table. Unfortunately, there is no *a priori* reason to explain why behavior on the ED200 examination should be any different from that on the other examinations.

Table 4.2 Summary of Signs of Beta Weights Showing Variables Significant in the Neighborhood of the .10 Level.

	ATL	NS	SS	HUM	ED200	MATH
ENG						
READ				+	$+^2$	
CQT-V	-	$+^2$	-		$-^2$	
CQT-I	+		$-^2$	-		+
CQT-N		-				-
SEX		+			+	
TRANS	-	+		$-^2$	$+^2$	
CRED					$-^2$	
GPA			$+^2$			
SCORE		+		$-^2$		$-^2$

Condensed for convenient reference from Tables 3.5, 3.7, 3.9, 3.11, 3.13, and 3.15.

Squared signs refer to quadratic components.

significant squared terms, but only in two courses each. When compared with the summary of signs for .05 level solutions (Table 4.3), it can be seen that several of the squared terms are no longer significant, which may help to explain why Blumenfeld and Berry (1965) obtained suggestive but not statistically significant results when they sought curvilinear functions.

Table 4.4 shows the decreases in the values of R for the reduced sets of independent variables. Although the value of R necessarily dropped in each course in the process of deleting terms, it appears that the difference between the two values is not of serious practical import, considering the gain in simplicity. If we look for specific variables which remain, we see, returning to Table 4.2, that CQT-V, CQT-I and TRANS proved significant in four courses while SCORE proved significant in three. Furthermore, in the same table, we see that in every course at least two of these variables were significant predictors.

Even though certain variables tend to stand out more than others, there are differences between the courses. NS and MATH, as expected, differed from the other courses in showing CQT-N as a significant predictor (although it dropped out of MATH in the .05 solution). Since ENG, READ, and CQT-V have fairly high intercorrelations, the results in the first five courses can perhaps best be interpreted

Table 4.3 Summary of Signs of Beta Weights Showing Variables Significant at the .05 Level.

	ATL	NS	SS	HUM	ED200	MATH
ENG						
READ						
CQT-V	-	-	-		$-^2$	
CQT-I				-		
CQT-N		-				
SEX						
TRANS	-					
CRED					$-^2$	
GPA				+		
SCORE		+	-	$-^2$		$-^2$

Condensed for convenient reference from Tables 3.6, 3.8, 3.10, 3.12, 3.14, and 3.16.

Squared signs refer to quadratic components.

Table 4.4 Summary of Multiple Correlation Coefficients (R), Using "TIME" as the Dependent Variable, for the Beginning Solutions, the .10 Solutions, and the .05 Solutions.

Group	N	R		
		Beginning Solution	.10 Solution	.05 Solution
ATL	221	.388	.326	.310
NS	178	.491	.453	.406
SS	160	.457	.412	.381
HUM	134	.456	.434	.385
ED200	85	.702	.655	.590
MATH	127	.485	.434	.394

Summarized from Tables 3.4 through 3.16.

as showing that a measure of verbal ability is a useful predictor of the time score, while in MATH, such does not appear to be the case.

Some of the variables may be acting as suppressor variables[3] (see Table 4.2). For example, in ATL, CQT-I may

[3] A suppressor variable is one not correlated with the criterion, but rather with another variable which is, in turn, correlated with the criterion. Thus, it acts to subtract out that part of the second variable which is irrelevant to the criterion (see DuBois, 1965, p. 184).

thought of as being subtracted from CQT-V, leaving that portion of CQT-V which is independent of CQT-I, thus improving the prediction. In HUM, READ can be interpreted as a suppressor variable on CQT-I, while in MATH, CQT-I appears to act as the suppressor variables on CQT-N.

The interpretation of some of the quadratic variables is also made clearer by viewing them as suppressor variables. In NS, the quadratic variable CQT-V seems to act as a suppressor of CQT-N and in ED200, the quadratic variable READ similarly suppresses another quadratic variable, CQT-V.

It might be interesting to speculate on some of the other variables in the tables, e.g., why GPA appears to be a suppressor variable on CQT-V and CQT-I in SS but not in other courses; or why CRED shows up in ED200 but not elsewhere.[4] However, the investigation of these types of relations will need to await further study when replications can assure their stability.

Generalizing from the data in all six of the above courses, we see that the larger part of the variation in time scores is accounted for by variables other than those

[4] A comparison of Table 4.2 with Table 4.3 shows that some of these relations are not particularly stable. TRANS, for example, which gives seemingly contradictory results in Table 4.2, essentially drops out in Table 4.3.

used as predictor variables. Whether or not personality measures could account for a sizable proportion is not known, although it is improbable, considering the reliabilities and validities reported for personality scores.

Of the measures used, verbal ability and total score are the most frequently appearing predictors. In some courses, only one of these two appears, indicating considerable overlapping between them. This frequency of appearance might be interpreted as implying that the final examinations are "speeded tests," in the sense that those with the best knowledge of the field and best verbal ability finish first. But the term "speeded tests" is already well defined in the literature as tests where few students finish and where scores reflect how many items were completed. Using that definition, there is no evidence that these tests are speeded. It is reasonable, however, to assume that the possession of high verbal ability will enable a student to finish a test sooner than another possessing lower verbal ability, since it indicates a greater potential to quickly read and comprehend written passages. Likewise, the possession of subject matter knowledge will help a student to work faster, for it enables him to often answer questions without hesitation while the less able students ponder. On the other hand, the student with poor ability is also likely to finish early, especially when he recognizes that he knows little about the items and concludes that a quick guess will probably produce about the same results as prolonged reasoning. Thus, both the most capable and the least capable students will often be among the first who turn in their papers, resulting in a quadratic relation in the prediction equation.

It might be concluded, therefore, that slowness in an examination is often associated with lower verbal ability and moderate subject matter knowledge. It would be a mistake, however, to also conclude that students possessing these characteristics are thereby likely to receive a score reflecting less than their true ability. On each of the examinations in this study, ample time was allowed for almost all students to finish, and thus the relationships found probably did not adversely affect test scores. It is well for students and instructors to be aware of some of the factors that influence time scores, but to also understand that a well constructed examination does not differentially penalize among fast and slow students.

In summary, it appears that between cognitive variables and time scores there exist both linear and quadratic relations. And further, that for these relations and their variations between courses, there can often be found logical explanations.

Problem Three

The Search for a Time Factor

The existence of a time factor in the tests was not revealed by this study. The eigenvalue vector was not highly structured and neither the Varimax nor Quartimax rotations were capable of extracting any factor with a high loading from the TIME variable.

In contrast, Gulliksen (1950) was quite certain that he had found several time factors. There are a number of reasons that might account for this discrepancy. First, he used test scores while we used item scores and the difference in reliability between the two types of scores may be sufficient to explain the discrepancy. Second, he used several types of tests while we used items of only one type (i.e. all from the same ATL test), and it is possible that analyses in other areas might produce positive results. Third, he used an oblique rotation while we used orthogonal vectors, and thus a potential time factor in our data may have been broken up. However, if the latter be the case, one might be skeptical of the interpretation of an oblique time factor when it could not be shown to have some components independent of other measures.

It would be wrong to conclude that these results are necessarily in contradiction to those of Gulliksen. It is possible that with other types of measures in other cognitive areas an interpretable time factor might emerge. The present data only suggest that it is unlikely using variables similar to those used in this study.

Problem Four

A Comparison of Two Measures of Consistency on Timed Portions of a Test

Although the results of this study did reveal numerical differences between the KR20 and odd-even reliability coefficients, any interpretations drawn therefrom must be made with certain qualifications. First, the differences between the 45 minute odd-even coefficients and the 45 minute KR20 coefficients are not so large as to result in gross errors. Second, for all practical purposes the scores obtained in the two hour time limit can be considered as untimed measures, and when the reliability coefficients are obtained from these scores (using either the large or small samples), it is found that the coefficients based on times scores are inflated to relatively moderate degrees.

These findings are most likely the consequence of using tests which are largely power measures. Even for timed portions, the number of items that a student completes correctly seems to depend more on his knowledge than on his work speed. Similar results would probably not be found in clerical or secretarial tests, where rate of work varies widely and is considered a major criterion. There, of course, we would expect to find timed coefficients to be seriously inflated and the odd-even coefficients to be inflated more than the KR20 coefficients. But for those examinations which emphasize knowledge and reasoning ability, we would expect to find only small differences between odd-even and KR20 coefficients on a timed portion of a test, and further, only small differences between either of these coefficients from timed portions and coefficients calculated on untimed portions.

Thus, these results are not necessarily in conflict with those of Cronbach and Warrington (1951). The 36 high school students, from whom their data were collected, were given four mental tests and instructed to work for both speed and accuracy. It is, therefore, not surprising that a larger speed effect would be found in such data than in a final examination where speed (other than finishing within the time limits) receives little encouragement.

Practical Implications

In the first chapter of this thesis, it was mentioned that the results might be useful to both the test constructor and the student. While some conclusions about the stability and predictablity of time scores were derived from the study, the major conclusions themselves are not in the form of practical suggestions. However, some considerations relevant to effective test construction and test taking behavior can be inferred from them.

Even on professionally made tests, typical of those analyzed herein, a number of students finish within one-half of the allotted time, while others stay until the very end. If it were possible to reduce this wide variation, a larger sample of content could be included in the test. While it is not desirable to reduce the variation arising from different degrees of subject matter knowledge, it is desirable to reduce the variation arising from other, statistically independent sources.

As indicated by one of the major conclusions of this study, verbal ability is such an independent source. Therefore, attention should be paid to the control of its effects. For example, the vocabulary of the test questions should be examined to eliminate those unfamiliar words peripheral to the major ideas. Awkward grammatical expressions should be revised. The resulting reduction in the variability of the time scores would allow more test items to be included, which in turn, would make it possible to improve both the reliability and the validity of the examination.

For the student, there are several recommendations which can be inferred from the results of this investigation. They relate to the full use of the time available for reflective thought, to the acquisition of verbal ability, and to the acquisition of subject matter knowledge. With respect to the first, we have already mentioned that many students of lower ability hand in their tests quite early, thus depriving themselves of the insights which might come from further reflective thought. The student has more direct control over this variable than the other two, and should exercise it to his best advantage.

Some students complain that not enough time is allowed for them to complete their examination. While this may, upon occasion, be a legitimate complaint, the student should nevertheless consider whether or not the problem is due to his lack of verbal ability and subject matter knowledge. Both of these variables proved to be significant predictors of time scores. Therefore, an improvement in his competence in either or both of these areas, coupled with the application of the elementary rules of "test-wiseness," should help him to avoid wasted time, and provide him with a better opportunity to respond carefully to each of the test items.

CHAPTER V

SUMMARY

Time scores were obtained during the final examinations in nine different university courses. For six of these courses, composed of a relatively broad sample of university freshmen and sophomores, scores from five entrance examinations were obtained, covering the areas of English proficiency, reading, verbal ability, general information, and numerical ability. In addition, the number of credits earned, number of credits transferred (from another institution), grade point average, and sex were also recorded for each student in these same six courses. In the other three courses, time scores were also taken on the midterm examinations. Product-moment correlations indicated that the time scores had about the same stability from midterm to final as did the examination scores. But a factor analysis of item scores and time scores failed to substantiate the existence of a time factor in the items.

Multiple regression, using the time score as the dependent variable and the first and second powers of the other academic measures as the independent variables, produced evidence of a useful degree of prediction. There was evidence that for some of the variables, a quadratic component was a significantly better predictor than was a linear component. As the variables were stepwise deleted, verbal ability seemed to emerge as the strongest predictor, aided by supressor variables. A number of the differences between prediction equations in different courses could be logically explained, while others appeared to be the result of sampling errors producing unstable beta weights.

When KR20 reliability coefficients were compared with odd-even coefficients for timed portions of the tests, the former were found usually to be smaller, but not by any large difference. Nor were either of these coefficients found to be substantially inflated above coefficients calculated on the total test. These results were interpreted to be the consequence of using power tests, in which the student felt little or no time pressure.

The evidence for the stability of time scores and their predictability from other academic variables seemed to be sufficient to warrant further investigation of their properties.

Suggestions for Further Research

There are at least three lines of research suggested by this study that might prove fruitful in future investigations.

1. It may be that a comparison of extreme groups would reveal differences clouded by the analysis of the total population. Would a discriminant analysis, based on the earliest and latest twenty five per cent of the population, reveal interpretable factors?

2. The factor analysis results may have been peculiar to the ATL population and not indicative of the results to be expected from other groups. In addition to replication in other subject matter areas, future investigators might seek other time measures (e.g. the 45 minute item score) likely to produce a stronger time factor.

3. Although student opinions were not solicited in the present study, they might prove fruitful in suggesting other promising predictor variables. Such a questionnaire could also request the student to estimate his test score and time of finish (in terms of quartiles or deciles). The investigator would want to solicit the information early in the term to reduce the effect of a self-fulfilling prophecy. Were it possible to gather information for the same student in several different courses (which was not practically feasible in the present study), the consistencies and variations on these variables could be noted between courses.

BIBLIOGRAPHY

Anderson, Theodore W. An Introduction to Multivariate Statistical Analysis. New York: Wiley, 1958. 374 pp.

Barch, Abram M. "The Relation of Departure Time and Retention to Academic Achievement." Journal of Educational Psychology 48:352-58; October 1957.

Bennett, George K., Bennett, Marjorie G., Wallace, Wimburn L., and Wesman, Alexander G. College Qualification Tests, Manual. (Rev. ed.) New York: Psychological Corporation, 1961, 61 pp.

Blumenfeld, Warren S. and Berry, Richard N. "Rapidity of Test Completion and Level of Score Attained." Psychological Reports 16:327-30; February 1965.

Briggs, Arvella and Johnson, Donald M. "A Note on the Relation Between Persistence and Achievement on the Final Examination." Journal of Educational Psychology 33:623-27; November 1942.

Brown, William. "Some Experimental Results in the Correlation of Mental Abilities." British Journal of Psychology 3:296-322; October 1910.

Burak, Benjamin. "Relationship Between Course Examination Scores and Time Taken to Finish the Examination, Revisited." Psychological Reports 20:164; February 1967.

Cooley, William W. and Lohnes, Paul R. Multivariate Procedures for the Behavioral Sciences. New York: Wiley, 1962, 211 pp.

Cronbach, Lee J. "Coefficient Alpha and the Internal Structure of Tests." Psychometrica 16:297-334; September 1951.

Cronbach, Lee J. and Warrington, Willard G. "Time Limit Tests: Estimating Their Reliability and Degree of Speeding." Psychometrica 16:167-88; June 1951.

Dowd, Constance E. "A Study of the Consistency of Rate of Work." Archives of Psychology No. 84; 1926. 33 pp. (Psychological Abstracts Vol. 1, No. 705; 1927)

DuBois, Phillip H. An Introduction to Psychological Statistics. New York: Harper and Row, 1965. 513 pp.

Ebel, Robert L. "The Use of Item Response Times in Achievement Test Construction." Unpublished PhD Thesis. Iowa City: State University of Iowa. 1947.

Ebel, Robert L. "The Characteristics and Usefulness of Rate Scores on College Aptitude Tests." Educational and Psychological Measurement 14(1):20-28; 1954.

Ezekiel, Mordecai and Fox, Karl A. Methods of Correlation and Regression Analysis: Linear and Curvilinear. New York: Wiley, 1959. 535 pp.

Flaugher, Ronald L., Melton, Richard S., and Myers, Charles T. "A Study of the Effects of Item Rearrangement." (RB-66-39) Princeton, N.J.: Educational Testing Service, 1966. 56 pp.

Freeman, Frank N. "Note on the Relation Between Speed and Accuracy on Quality of Work." Journal of Educational Research 7:87-89; January 1923.

Graybill, Franklin A. An Introduction to Linear Statistical Models. Vol. 1. New York: McGraw Hill, 1961. 459 pp.

Guilford, Joy Paul. Fundamental Statistics in Psychology and Education. New York: McGraw Hill, 1965. 598 pp.

Gulliksen, Harold. "The Reliability of Speeded Tests." Psychometrica 15:259-60; September 1950.

Hays, William L. Statistics for Psychologists. New York: Holt, Rinehart, and Winston, 1963. 719 pp.

Hempel, Carl G. Fundamentals of Concept Formation in Empirical Science. Chicago: University of Chicago Press, 1952. 87 pp.

Kuder, G. Frederick and Richardson, Marion W. "The Theory of the Estimation of Test Reliability." Psychometrica 2:151-60; September 1937.

Lord, Frederick M. "A Study of Speed Factors in Tests and Academic Grades." Psychometrica 21:31-50; March 1956.

Morrison, Edward J. "On Test Variance and the Dimensions of the Measurement Situation." Educational and Psychological Measurement 20(2):231-50; 1960.

Rafter, Mary E. and Ruble, William L. "Stepwise Deletion of Variables from a Least Squares Equation (LSDEL Routine), Stat Series No. 8." East Lansing, Mich.: Michigan State University Agricultural Experiment Station, April, 1967. 20 pp.

Ruble, William L., Kiel, Donald F., and Rafter, Mary E. "Calculation of Least Squares (Regression) Problems on the LS Routine (Stat Series No. 7)." East Lansing, Mich.: Michigan State University Agricultural Experiment Station, May 1967. 61 pp.

Smillie, K. W. An Introduction to Regression and Correlation. New York: Academic Press, 1966. 161 pp.

Spearman, Charles. "Correlation Calculated from Faulty Data." British Journal of Psychology 3:271-95; October 1910.

Williams, Anthony. "Factor Analysis; Factor A: Principal Components and Orthogonal Rotations. Technical Report No. 34." East Lansing, Mich.: Michigan State University Computer Institute for Social Science Research, May 17, 1967. 14 pp.

APPENDIX A

Brief Description of Courses[1]

ATL 113 American Thought and Language

Training in reading and writing through the use of selected American documents; particular emphasis on problems of style. Library papers. Weekly writing assignments.

NS 183 Natural Science

The role played by theories in physical science in man's attempt to find a unified view of nature. The Copernican Revolution and Molecular and Atomic Theories related to man's concept of the universe and the nature of matter. Emphasis is placed on the social and philosophical preconditions necessary for the development and modification of scientific ideas.

SS 233 Social Science

Problems of change. Achieving national, political, economic and social objectives in the emerging nations. The Soviet Union and directed change. Problems of reconciling national self-interest with the needs for world peace.

HUM 243 Humanities

Considers aspects of modern Western culture since 1600. Topics include the impact of political and social revolutions, the intellectual and spiritual problems associated with the rise of modern science, and philosophical, religious, literary, and artistic interpretations of the contemporary human situation.

[1] Adapted from catalog descriptions.

ED 200 Individual and the School

Major psychological factors in the school learning teaching situation; concepts in human development related to problems in the school situation; teacher's role in motivation, conceptual learning, problem solving, and the development of emotional behavior, attitudes and values; learning of skills; retention and transfer; and measurement of student abilities and achievement.

MTH 201 Foundations of Arithmetic

Fundamental concepts and structure of arithmetic for prospective elementary school teachers.

ED 465 Introduction to Measurement and Evaluation in the Classroom

The construction, use, and evaluation of teacher-made classroom tests, objective and essay, in elementary schools, secondary schools and colleges. Statistical analysis of test scores and item responses. Grading problems and procedures.

ED 865 Psychological Measurement and Test Interpretation in Education

Measurement theory and analysis of test results. Survey of standardized tests of aptitude and intelligence; study of selection and use of such tests; an intensive evaluation of at least one measuring instrument. Concepts of reliability, validity, norms.

ED 982 Seminar in Experimental Design

Theory and practice in the design, analysis and interpretation of experimental and quasi-experimental research.

APPENDIX B

Notes to Proctors
Concerning the Collection of Rate Score
Data on Final Examinations

At the beginning of the examination, explain to the students in your own words, the following points:

1) At the end of 40 minutes, we will ask the students to circle the item on which they are working on the answer sheet.

2) At the end of 45 minutes we will ask them to mark an X through the item on which they are working. This will usually be a number following the circled number, but may precede it if for some reason they jumped back.

3) When they finish the test and are ready to bring it forward, they will record at top center the number which appears on the cards at the front of the room (which are changed every 60 seconds).

The test timer will change the numbers on the cards, and at the end of the test period will take the papers to the scoring office.

These data will in no way affect the students' grades (in fact it will not be analyzed for several weeks), but will be used solely to obtain information to improve test construction and student test taking behavior. A short summary of the statistical results should be available in about two months from Evaluation Services. A copy will be sent to each proctor.

APPENDIX C

Intercorrelations of the Variables

ATL Subjects, N = 221

READ	366									
CQT-V	326	478								
CQT-I	164	399	397							
CQT-N	335	374	105	415						
SEX	172	-021	063	-314	-391					
TRANS	076	090	040	-007	053	015				
CRED	-034	114	071	034	118	018	215			
GPA	349	407	232	236	267	105	194	326		
TIME	-159	-145	-216	010	-067	-076	-231	-059	-056	
SCORE	335	526	481	338	119	141	039	135	502	-056
	ENG	READ	CQT-V	CQT-I	CQT-N	SEX	TRANS	CRED	GPA	TIME

NS Subjects, N = 178

READ	488									
CQT-V	541	536								
CQT-I	288	369	397							
CQT-N	201	096	025	432						
SEX	263	099	136	-331	-338					
TRANS	-136	-082	055	-068	-113	-026				
CRED	-059	003	044	282	209	-260	-004			
GPA	367	243	230	251	224	059	-028	051		
TIME	-074	-085	-179	-117	-158	058	138	-027	144	
SCORE	303	255	226	510	507	-316	022	192	538	159
	ENG	READ	CQT-V	CQT-I	CQT-N	SEX	TRANS	CRED	GPA	TIME

SS Subjects, N = 160

READ	490									
CQT-V	452	665								
CQT-I	261	454	500							
CQT-N	365	401	288	448						
SEX	132	017	081	-336	-166					
TRANS	020	-097	030	-145	-080	107				
CRED	166	290	165	324	296	-139	-403			
GPA	247	396	395	378	319	101	002	335		
TIME	111	-237	-344	-263	-130	054	-056	-094	-229	
SCORE	143	433	450	585	173	-298	-106	277	511	-302
	ENG	READ	CQT-V	CQT-I	CQT-N	SEX	TRANS	CRED	GPA	TIME

HUM Subjects, N = 134

READ	472									
CQT-V	487	563								
CQT-I	195	476	480							
CQT-N	312	389	129	409						
SEX	338	102	142	-307	-098					
TRANS	-128	-041	-017	029	020	093				
CRED	-091	-024	-092	094	159	-135	-076			
GPA	236	298	220	329	262	058	-031	169		
TIME	-068	-204	-227	-302	-116	136	-105	023	049	
SCORE	135	365	332	261	091	051	033	-019	565	-021
	ENG	READ	CQT-V	CQT-I	CQT-N	SEX	TRANS	CRED	GPA	TIME

ED200 Subjects, N = 85

READ	669									
CQT-V	663	753								
CQT-I	467	572	539							
CQT-N	202	244	088	419						
SEX	167	078	118	-289	-273					
TRANS	216	236	206	174	057	-072				
CRED	071	-069	-036	062	-022	-037	-290			
GPA	415	395	320	452	143	103	-074	236		
TIME	-456	-402	-481	-305	-188	120	-137	-055	-156	
SCORE	138	266	139	198	078	091	-062	125	521	018
	ENG	READ	CQT-V	CQT-I	CQT-N	SEX	TRANS	CRED	GPA	TIME

MATH Subjects, N = 127

READ	486									
CQT-V	498	593								
CQT-I	518	585	530							
CQT-N	455	378	327	496						
SEX	041	-101	-102	-131	-053					
TRANS	103	-086	017	-064	-111	024				
CRED	-032	119	170	177	116	-137	-023			
GPA	340	465	318	470	331	-015	-237	209		
TIME	112	089	-011	172	026	-049	-002	-015	075	
SCORE	295	305	198	295	439	028	-070	-084	459	303
	ENG	READ	CQT-V	CQT-I	CQT-N	SEX	TRANS	CRED	GPA	TIME