



GRADUATE RECORD EXAMINATIONS

A SUMMARY OF THE RESEARCH LEADING TO THE REVISION OF THE FORMAT OF THE GRADUATE RECORD EXAMINATIONS APTITUDE TEST IN OCTOBER 1981

Cheryl L. Wild Spencer S. Swinton Madeline M. Wallmark

GRE Board Research Report GREB No. 80-1aR ETS Research Report 82-54

November 1982

This report presents the findings of a research project funded by and carried out under the auspices of the Graduate Record Examinations Board.



EDUCATIONAL TESTING SERVICE, PRINCETON, NJ

- Altman, R. A. and Wallmark, M. M. A Summary of Data from the Graduate Programs and Admissions Manual. 74-1R, 1975.
- Baird, L. L. An Inventory of Documented Accomplishments. 77-3R, 1979.
- Baird, L. L. Cooperative Student Survey (The Graduates [\$2.50 each], and Careers and Curricula). 70-4R, 1973.
- Baird, L. L. The Relationship Between Ratings of Graduate Departments and Faculty Publication Rates. 77-2aR, 1980.
- Baird, L. L. and Knapp, J. E. The Inventory of Documented Accomplishments for Graduate Admissions: Results of a Field Trial Study of Its Reliability, Short-Term Correlates, and Evaluation. 78-3R, 1981.
- Burns, R. L. Graduate Admissions and Fellowship Selection Policies and Procedures (Part I and II). 69-5R, 1970.
- Centra, J. A. How Universities Evaluate Faculty Performance: A Survey of Department Heads. 75-5bR, 1977. (\$1.50 each)
- Centra, J. A. Women, Men and the Doctorate. 71-10R, 1974. (\$3.50 each)
- Clark, M. J. The Assessment of Quality in Ph.D. Programs: A Preliminary Report on Judgments by Graduate Deans. 72-7aR, 1974.
- Clark, M. J. Program Review Practices of University Departments. 75-5aR, 1977. (\$1.00 each)
- Clark, M. J and Centra, J. A. Conditions Influencing the Career Accomplishments of Ph.Ds. 76-2R, 1982.
- DeVore, R. and McPeek, M. A Study of the Content of Three GRE Advanced Tests. 78-4R, 1982.
- Donlon, T. F. Annotated Bibliography of Test Speededness. 76-9R, 1979.
- Flaugher, R. L. The New Definitions of Test Fairness In Selection: Developments and Implications. 72-4R, 1974.
- Fortna, R. O. Annotated Bibliography of the Graduate Record Examinations. 1979.
- Frederiksen, N. and Ward, W. C. Measures for the Study of Creativity in Scientific Problem-Solving. 1978.
- Hartnett, R. T. Sex Differences in the Environments of Graduate Students and Faculty. 77-2bR, 1981.
- Hartnett, R. T. The Information Needs of Prospective Graduate Students. 77-8R, 1979.

- Hartnett, R. T. and Willingham, W. W. The Criterion Problem: What Measure of Success in Graduate Education? 77-4R, 1979.
- Knapp, J. and Hamilton, I. B. The Effect of Nonstandard Undergraduate Assessment and Reporting Practices on the Graduate School Admissions Process. 76-14R, 1978.
- Lannholm, G. V. and Parry, M. E. Programs for Disadvantaged Students in Graduate Schools. 69-1R, 1970.
- Miller, R. and Wild, C. L. Restructuring the Graduate Record Examinations Aptitude Test. GRE Board Technical Report, 1979.
- Powers, D. E. and Swinton, S. S. Effects of Self-Study of Test Familiarization Materials for the Analytical Section of the GRE Aptitude Test. 79-9R, 1982.
- Reilly, R. R. Critical Incidents of Graduate Student Performance. 70-5R, 1974.
- Rock, D. and Werts, C. An Analysis of Time Related Score Increments and/or Decrements for GRE Repeaters across Ability and Sex Groups. 77-9R, 1979.
- Rock, D. A. The Prediction of Doctorate Attainment in Psychology, Mathematics and Chemistry. 69-6aR, 1974
- Schrader, W. B. Admissions Test Scores as Predictors of Career Achievement in Psychology. 76-1aR, 1978.
- Schrader, W. B. GRE Scores as Predictors of Career Achievement in History. 76-1bR, 1980.
- Swinton, S. S. and Powers, D. E. A Study of the Effects of Special Preparation on GRE Analytical Scores and Item Types. 78-2R, 1982.
- Wild, C. L. Summary of Research on Restructuring the Graduate Record Examinations Aptitude Test. 1979.
- Wild, C. L. and Durso, R. Effect of Increased Test-Taking Time on Test Scores by Ethnic Group, Age, and Sex. 76-6R, 1979.
- Wilson, K. M. A Study of the Validity of the Restructured GRE Aptitude Test for Predicting First-Year Performance in Graduate Study. 78-6R, 1982.
- Wilson, K. M. The GRE Cooperative Validity Studies Project. 75-8R, 1979.
- Wiltsey, R. G. Doctoral Use of Foreign Languages: A Survey. 70-14R, 1972. (Highlights \$1.00, Part I \$2.00, Part II \$1.50).
- Witkin, H. A.; Moore, C. A.; Oltman, P. K.; Goodenough, D. R.; Friedman, F.; and Owen, D. R. A Longitudinal Study of the Role of Cognitive Styles in Academic Evolution During the College Years. 76-10R, 1977 (\$5.00 each).

A Summary of the Research Leading to the Revision of the Format of the Graduate Record Examinations Aptitude Test in October 1981

Cheryl L. Wild Spencer S. Swinton Madeline M. Wallmark

GRE Board Research Report GREB No. 80-1aR

November 1982

A Summary of the Research Leading to the Revision of the Format of the Graduate Record Examinations Aptitude Test in October 1981

Background

Consideration of a new format of the Graduate Record Examinations (GRE) Aptitude Test began in early 1974. The goal was to broaden the Aptitude Test and thus allow examinees to demonstrate a wider array of academic talents. Research was conducted on the feasibility of shortening the verbal and quantitative sections from 75 to 50 minutes and on seven proposed item types for a new reasoning module. In October 1977, the restructured GRE Aptitude Test was introduced. This test consisted of a 50-minute verbal section, a 50-minute quantitative section, two 25-minute analytical sections and a 25-minute variable section. Users were cautioned about the experimental nature of the analytical measure and were asked not to place undue weight on it until further research was conducted.

Meanwhile, the GRE Board Research Committee has funded a number of studies of the analytical measure and the restructured test as a whole. Although validity data for the analytical measure were quite promising, a problem in predicting the difficulty of the analytical questions when they were presented in the operational sections (III and IV) from the questions' difficulty in the trial section (V) was first noticed after introduction of a new test form in October 1978. Further studies were conducted to identify the reason for this lack of predictability, and the preliminary finding of a "practice effect" was reported to the GRE Board in September 1979. Since the GRE Aptitude Test has generally been designed to measure abilities learned over a long period of time, the apparent existence of a short-term practice effect raised an important policy question for the Board.

At the same time, another issue that could affect the format of the Aptitude Test was being discussed. On July 13, 1979, the New York State test disclosure law was signed. This law required that any GRE Aptitude Test administered in the state of New York after January 1, 1980, be disclosed within 30 days of reporting the test scores (amendments have since been passed that exclude low-volume administrations from the If similar test disclosure legislaimmediate disclosure requirement). tion were to be passed by other states or by the federal government, or if the GRE Board were to decide as a long term policy to disclose three or more Aptitude Test editions a year, the current equating method (i.e., the statistical method of assuring that scores earned on the multiple editions are comparable by spiraling, or administering a new test form and one or two old test forms at the same administration) would likely need to be supplemented or replaced by other methods. By changing the format of the GRE Aptitude Test, the Board could allow for flexibility in the equating procedure even if no immediate change in equating occurred.

Between September 1979 and September 1980, the GRE Board and its committees discussed possible options for score equating, scoring, test format, test content, and test disclosure. Research studies were funded to help inform the decisions about the test. The purpose of this paper is to summarize the research leading to, and decisions about, the reformatting of the GRE Aptitude Test in October 1981. Discussions and research results on the following four issues related to the GRE Aptitute Test are presented:

- 1. What is the appropriate format for each section of the GRE Aptitude Test?
- 2. What test disclosure plans are anticipated for 1981-82?
- 3. Should the GRE Aptitude Test continue to be formula scored or should rights-only scoring be introduced?
- 4. Should the analytical score continue to be reported, and if so, what item types should be included?

Test disclosure plans, especially those on the number of forms to be disclosed per year, have been included because they directly affect the development of questions. Recommendations concerning the format and content of the test have direct implications on which equating methodologies are feasible. For this reason, the impact of each recommendation on the different equating methodologies is mentioned throughout the discussion and a general summary of the equating options is presented in the full report.

Format

What is the appropriate format for each section of the GRE Aptitude Test?

The GRE Aptitude Test administered between October 1977 and September 1981 was composed of one 50-minute verbal section, one 50-minute quantitative section, two 25-minute analytical sections, and one 25-minute section that was used for item trials. After review of several disclosed tests, the identification of the trial section becomes fairly obvious to examinees. Although this didn't present a problem for equating by spiraling (it can present administrative problems for test center supervisors), it would present a problem for other methods of equating that would be required at disclosure administrations and may be needed to supplement spiraling at nondisclosure administrations. The variable section could be made less transparent by constructing a test with equally timed sections and changing the location of the variable section. One variable section should be sufficient for pretesting and equating if only two or three test forms are disclosed a year.

Seven 25-minute sections would represent a test of the same length as the original test. However, review of data from 25-minute verbal sections

suggested that the test was somewhat more speeded than desirable, especially if the option of equating using item response theory were to be kept open. Thus 30-minute sections were indicated for the verbal measure.

Review of 25-minute quantitative sections suggested that the 25-minute sections were not quite as highly correlated with the operational section as would be desirable for anchor equating. Since the correlations were about as high as possible given the reliability of the sections, the best way to raise the correlations would be to add questions to the operational 55-item section. Since the analytical reasoning questions are somewhat more correlated to quantitative comparison questions and since data interpretation questions are more time consuming, it was suggested that five regular mathematics questions be added. This would also require lengthening the time for the quantitative measure to allow for two 30-minute sections. The proposed analytical measure is composed of item types that require a great deal of time per question and two 30-minute sections would be indicated.

The data therefore suggested that the GRE Aptitude Test should be composed of seven 30-minute sections - two verbal, two quantitative, two analytical, and one variable section for pretesting and possibly equating. This has the effect of lengthening the test by 35 minutes.

Test Disclosure Plans 1981-82

What test disclosure plans are anticipated for 1981-82?

Discussions about test disclosure are related to future plans for the Aptitude Test because the number of forms disclosed directly affects the number of questions to be developed. Three alternative plans for test disclosure were presented, ranging from the least to the most disclosure. The ultimate goal of the first plan was to reach a steady state (1983-84) in which two Aptitude Test editions are developed each year, two test editions are disclosed each year, and four editions are secure at the end of the year. The goal of the second plan was to ultimately disclose three tests a year, develop three tests a year, and maintain four secure editions at the end of the year. The third option, maximum disclosure, had as its goal disclosing five test editions a year, one after each national administration. Each plan assumed that during 1981-82 only two or three test editions would be disclosed (plus a sample test) while four or five new editions were introduced. In this way some secure tests would be available for handicapped, Monday, and special administrations.

Depending on the number of examinees per year, it might be necessary to have two variable sections in the Aptitude Test for pretesting and equating if full disclosure were desired. Given the already long testing period, the small percent of examinees requesting copies of their actual test, and the earlier expressed reluctance of the Board to move to full disclosure without further information on its impact, the full disclosure option was not considered feasible at this time.

Disclosure of two Aptitude Tests a year would mean offering the test in New York at only two of the five national administrations and would represent a decrease from the current level. Although this may be desirable during 1981-82 in order to build a pool of secure test forms, the long term consequences of such a cutback in services may not be desirable. For these reasons, the second option, with a goal of ultimately disclosing three test forms a year, was chosen. This represents the same long-term publication rate as planned for 1980-81.

Scoring

Should the GRE Aptitude Test continue to be formula scored or should rights-only scoring be introduced?

In April 1980 the Minority Graduate Education Committee, the Research Committee, and the Board discussed the issues related to formula versus rights-only scoring. Formula scoring (coupled with formula directions) has the advantage of slightly reduced error variance in the reported scores while rights-only scoring (similarly coupled with rights directions) has the advantage that examinees know just what the best test-taking strategies are. The committees and the Board expressed a general preference for rights-only scoring, if the technical issue of impact on reported scores could be addressed.

Although changing the scoring procedure may result in some discontinuity in the score scale, preliminary results of a study of the SAT (Angoff and Schrader, 1981) indicated that such a discontinuity would be small for the Aptitude Test. The new format test could be linked to the earlier test by administering 30-minute equating sections with directions for rights scoring in April 1981 and again in October 1981.

One issue that was discussed before making a decision was whether it was appropriate to offer the GRE Aptitude Test and GRE Advanced Tests with different directions and scoring. An argument could be made that the Aptitude Test and Advanced Tests differ in their direct relation to undergraduate curricula. The Aptitude Test is not related to specific course work; examinees might be expected to have some knowledge about all the questions. Since the Advanced Tests cover a wide range of possible undergraduate preparation, no single individual would be expected to have studied all the areas covered in an Advanced Test. Thus it might be argued that formula-scoring directions would appropriately discourage random guessing on Advanced Tests.

Although preliminary analysis of data in the Angoff and Schrader study suggested that changing from formula to rights scoring for the Advanced Tests might be feasible without seriously affecting the score scale, a number of technical questions need to be answered before a final recommendation is made for those tests. Since the future of some Advanced Tests is currently in question, any recommendation to change the Advanced Test scoring procedures is premature.

The Board decided to introduce rights-only scoring (and directions) for the GRE Aptitude Test in October 1981 and continue formula scoring for the Advanced Tests. Staff were asked to continue to investigate the desirability of changing the scoring directions for the GRE Advanced Tests in future years.

Analytical Score

Should the analytical score continue to be reported and, if so, what item types should be included?

Three alternative answers to this question were reviewed:

- Continue to offer the test as then structured but provide practice materials and directions for special review to all examinees.
- 2. Modify the test to reduce the practice effect.
- 3. Drop the analytical measure as soon as possible.

Research on practice effect for verbal, quantitative, and analytical item types suggested that the "practice effect" for two of the analytical item types is relatively large. Powers and Swinton (1982) conducted a study that provided practice materials to a subset of examinees. Examinees were told about the practice effect and encouraged to study the materials provided. Results of the study show that the average analytical scores increased as the amount of encouragement increased. Scaled score differences between average analytical scores for control and for experimental groups were as high as 53 points. Average differences between examinees receiving the full set of practice materials and those receiving only the <u>Information Bulletin</u> were 33 points, not taking into account the effect of encouragement. The effect of encouragement irrespective of the amount of practice materials received was about 20 points. Another small study of the effect of a test preparation course yielded an average score increase of over 60 scaled score points (Swinton & Powers, 1982).

The rationale for the first option (to provide practice materials widely) was that the practice materials should equalize the effects of practice. Results of the Powers and Swinton study suggest that simply providing practice materials will not equalize the practice effect. Operationally this option also presents problems in relation to potential equating options. The current method of equating (spiraling) is not affected by the practice effect problem, as long as the analytical sections precede the trial questions in the variable section. However, disclosure may require use of alternate equating methods that would be sensitive to the practice effect.

The third option also had some disadvantages. The available validity evidence on the analytical measure suggested that the test was providing

valid and useful results. The analytical measure has provided good public relations in terms of "broadening skills" measured by the test. Change that narrows the focus of the test may not be well received, especially in light of the good validity evidence available. An advantage of dropping the analytical measure was that it would cut the cost of producing new tests substantially.

The second option, modifying the analytical test to reduce the practice effect, thus appeared to be the most feasible alternative. The rationale for selecting this option was that the test was producing valid and useful results but that the existence of a significant practice effect in the absence of a method of insuring that all examinees have the benefit of that practice could affect the fairness of the instrument. In considering this option, inclusion of each of the item types then in the analytical section plus one new item type (supporting conclusions) was evaluated. An estimate of the validity of each item type was obtained by looking at the relationship of the item type to self-reported undergraduate grade-point averages. Although this gives an estimate of the concurrent validity of the individual item types, an estimate of the predictive validity of a modified test is not available. A description of each item type and summary of the information about the item are presented below:

1. Analysis of Explanations

These items are based on brief narratives establishing a situation in which an action is taken in order to have a specific effect. A later result, which may or may not be related to the action, is described in a brief statement. Each item is a statement that must be evaluated in terms of the situation and the result. The examinee must determine whether the statement is inconsistent with the situation or result, a possible adequate explanation of the result, deducible from the situation or result, or relevant to or irrelevant to an explanation of the result.

Factor analytic studies of the GRE Aptitude Test have shown that this item type is factorially complex, with options A and E highly related to reading comprehension. However, it does contribute to an analytical factor and correlates relatively highly with self-reported undergraduate grade-point averages. Directions are quite complex, which may partially explain the relatively large practice effect found for this item type. In addition, this item type has been found to respond to coaching (Swinton & Powers, 1982).

2. Logical Diagrams

This item type presents several circle diagrams of possible relationships. Each item consists of three nouns and the student is asked to select the circle diagram that best characterizes the relationship of the three. Factor analytic studies suggest that this item type is contributing to the analytical factor (Powers, Swinton, Thayer, & Yates, 1978; Swinton & Powers, 1980; Rock, Werts & Grandy, 1982). It does not correlate highly with verbal or quantitative scores. Correlations with self-reported undergraduate grade-point averages are slightly lower than for the analysis of explanations item type. This item type also has been found to have a relatively large practice effect (about two-thirds of that found in analysis of explanations) and to respond to coaching (Swinton & Powers, 1982).

3. Analytical Reasoning - Type 1

These items are based on a brief set of statements expressing relationships among abstract symbols (letters) or among sets of rules governing processes or procedures having few concrete referents. The examinee is asked to draw inferences from and sometimes critically assess those statements.

The correlation of these items with verbal and quantitative scores is relatively low. With undergraduate grade-point averages the correlation is about the same as the logical diagrams and supporting conclusions item types. Analytical reasoning item types exhibit little practice effect, even though they appear at the end of a somewhat speeded section. The analytical reasoning item type is one of the two most highly rated by test development staff in terms of logical defensibility of the questions. One drawback of this item type is that it is time consuming for examinees.

4. Analytical Reasoning - Type 2

These items are based on brief arguments or statements presenting evidence or opinions. The questions require that the student recognize unstated presuppositions, logical flaws, methods of persuasion, conclusions logically following from given arguments, and the like.

Research on this item type for the Law School Admission Test program showned it to have a high criterion validity. This item type does exhibit a high correlation with self-reported undergraduate grade-point average, but it is about as highly correlated with the verbal as with the analytical score. It has a relatively low correlation with the quantitative score. No practice effect has been found for this item type and it is one of the two most highly rated analytical item types by the test development specialists asked to rank the question types. However, this item type is difficult and time consuming to write.

5. Supporting Conclusions

A passage is presented with several numbered sentences. Following the passage are several statements. For each statement the examinee must decide which of the numbered sentences are sufficient to supply all the information in each statement. This item type correlates moderately with the verbal score and less with the quantitative score. It has a fairly high correlation with the analytical measure. A small practice effect was found for this item type. Test development staff have expressed concerns about the logical defensibility of this item type. Although it may have promise in the future, more study is needed before it is included as an operational item type.

Summary. The research on these five item types suggests that the two analytical reasoning item types are the most appropriate for a revised analytical section. The rankings for each item type based on selected criteria are presented below, where (+) indicates a favorable rating, (0) a neutral rating, and (-) an unfavorable rating. These ratings are relative only to the item types studied here. Note also that four of the five item types had already been screened (Miller & Wild, 1979) on some of the criteria listed below as well as some not included in this study (e.g., item difficulty, efficiency, and face validity). For this reason, very few negative ratings have been given.

Test Development

Item Type	Practice Effect	Indepe V	ndence <u>Q</u>	Corre	elation UGPA	Rating For Logical Defensibility
Analysis of Explanations	_	0	+	+	+	0
Logical Diagrams	-	0	+	0	0	0
*Analytical Reasoning ~ Type l	+	+	+	0	0	+
*Analytical Reasoning - Type 2	+	-	+	+	+	+
Supporting Conclusions	0	0	+	+	0	0

^{*}Item types proposed for the revised analytical measure.

The two analytical reasoning item types thus appear to be the most promising for a new analytical module. The two also compensate for each other — analytical reasoning type 1 is more independent of the verbal measure while type 2 correlates more highly with self-reported undergraduate grade-point averages. In order not to inflate the correlation of the

analytical and verbal scores, staff proposed that the new measure include about 25 percent analytical reasoning type 2 and 75 percent analytical reasoning type 1 questions. These would be divided into two parallel separately timed sections (30 minutes each). Such a module could be expected to have a reliability of about .88 and intercorrelations with the verbal and quantitative measures in the range of .65 to .75.

Although such a section would be feasible and useful, it will represent a major change from the original analytical measure. In fact the item types in the revised measure constituted only 21 percent of the original measure. As a result, the predictive validity of the new analytical measure will have to be reassessed. Also due to the major change in the content, scores earned on the new test will not be comparable to those earned prior to the revision. This should not present a major problem to users, however, since the analytical score has been considered experimental up to this time. It was recommended that scores for the revised analytical measure be rescaled in relation to verbal and quantitative scores, rather than by attempting to link the new analytical measure to the old form.

Summary of the Revisions

The changes in the GRE Aptitude Test introduced in October 1981 have to do with the content of the analytical measure, the format and timing of the sections, the number of test forms to be disclosed per year, and the scoring method. The changes are as follows:

- 1. The analytical measure contains analytical reasoning items, about three-fourths type 1 and one-fourth type 2.
- 2. The goal of the test disclosure plan will be to publish three test editions a year. For 1981-82, five tests were introduced, but only two published.
- 3. The GRE Aptitude Test consists of seven 30-minute sections -- two verbal, two quantitative, two analytical, and one variable.
- 4. Scoring is based only on the number of correct responses (i.e., rights-only scoring).

References

- Angoff, W. H., & Schrader, W. B. A study of alternative methods for equating rights scores to formula scores. ETS Research Report 81-8. Princeton, NJ: Educational Testing Service, 1981.
- Miller, R,. & Wild, C. L. (Eds.) Restructuring the Graduate Record

 Examinations Aptitude Test. GRE Board Technical Report. Princeton,

 NJ: Educational Testing Service, 1979.
- Powers, D. E., & Swinton, S. S. Effects of self-study of test familiarization materials for the analtytical sections of the GRE Aptitude Test. GRE Board Research Report No. 79-9R.

 Princeton, NJ: Educational Testing Service, 1982.
- Powers, D. E., Swinton, S. S., Thayer, D., & Yates, A. A factor analytic investigation of seven experimental analytical item types. GRE Board Professional Report No. 77-1P. Princeton, N.J.: Educational Testing Service, 1978.
- Rock, D. A., Werts, C., & Grandy, J. Construct validity of the Graduate
 Record Examinations Aptitude Test across populations—an empirical
 confirmatory study. GRE Board Professional Report No. 78-1P.
 Princeton, NJ: Educational Testing Service, 1982.
- Swinton, S. S., & Powers, D. E. A factor analytic study of the restructured GRE Aptitude Test. GRE Board Professional Report No. 77-6P. Princeton, NJ: Educational Testing Service, 1980.
- Swinton, S. S., and Powers, D. E. A study of the effects of special preparation on GRE analytical scores and item types. GRE Board Research Report No. 78-2R. Princeton, NJ: Educational Testing Service, 1982.

- Boldt, R. R. Comparison of a Bayesian and a Least Squares Method of Educational Prediction. 70-3P, 1975.
- Campbell, J. T. and Belcher, L. H. Word Associations of Students at Predominantly White and Predominantly Black Colleges. 71-6P, 1975.
- Campbell, J. T. and Donlon, T. F. Relationship of the Figure Location Test to Choice of Graduate Major. 75-7P, 1980.
- Carlson, A. B.; Reilly, R. R.; Mahoney, M. H.; and Casserly, P. L. The Development and Pilot Testing of Criterion Rating Scales. 73-1P, 1976.
- Carlson, A. B.; Evans, F.R.; and Kuykendall, N. M. The Feasibility of Common Criterion Validity Studies of the GRE. 71-1P, 1974.
- Centra, J. A. Graduate Degree Aspirations of Ethnic Student Groups Among GRE Test-Takers. 77-7P, 1980.
- Donlon, T. F. An Exploratory Study of the Implications of Test Speededness. 76-9P, 1980.
- Donlon, T. F.; Reilly, R. R.; and McKee, J. D. Development of a Test of Global vs. Articulated Thinking: The Figure Location Test. 74-9P, 1978.
- Echternacht, G. Alternate Methods of Equating GRE Advanced Tests. 69-2P, 1974.
- Echternacht, G. A Comparison of Various Item Option Weighting Schemes/A Note on the Variances of Empirically Derived Option Scoring Weights. 71-17P, 1975.
- Echternacht, G. A Quick Method for Determining Test Bias. 70-8P, 1974.
- Evans, F. R. The GRE-Q Coaching/Instruction Study. 71-5aP, 1977.
- Fredericksen, N. and Ward, W. C. Development of Measures for the Study of Creativity. 72-2P, 1975.
- Kingston, N. and Dorans, N. Effect of the Position of an Item Within a Test on Item Responding Behavior: An Analysis Based on Item Response Theory. 79-12bP, 1982.
- Kingston, N. M. and Dorans, N. J. The Feasibility of Using Item Response Theory as a Psychometric Model for the GRE Aptitude Test. 79-12P, 1982.
- Levine, M. V. and Drasgow, F. Appropriateness Measurement with Aptitude Test Data and Esimated Parameters. 75-3P, 1980.

- McPeek, M.; Altman, R. A.; Wallmark, M.; and Wingersky, B. C. An Investigation of the Feasibility of Obtaining Additional Subscores on the GRE Advanced Psychology Test. 74-4P, 1976.
- Pike, L. Implicit Guessing Strategies of GRE Aptitude Examinees Classified by Ethnic Group and Sex. 75-10P, 1980.
- Powers, D. E.; Swinton, S.; Thayer, D.; and Yates, A. A Factor Analytic Investigation of Seven Experimental Analytical Item Types. 77-1P, 1978.
- Powers, D. E.; Swinton, S. S.; and Carlson, A. B. A Factor Analytic Study of the GRE Aptitude Test. 75-11P, 1977.
- Reilly, R. R. and Jackson, R. Effects of Empirical Option Weighting on Reliability and Validity of the GRE. 71-9P, 1974.
- Reilly, R. R. Factors in Graduate Student Performance. 71-2P, 1974.
- Rock, D. A. The Identification of Population Moderators and Their Effect on the Prediction of Doctorate Attainment. 69-6bP, 1975.
- Rock, D. A. The "Test Chooser": A Different Approach to a Prediction Weighting Scheme. 70-2P, 1974.
- Rock, D., Werts, C., and Grandy, J. Construct Validity of the GRE Aptitude Test Across Populations—An Empirical Confirmatory Study. 78-1P, 1982.
- Sharon, A. T. Test of English as a Foreign Language as a Moderator of Graduate Record Examinations Scores in the Prediction of Foreign Students' Grades in Graduate School. 70-1P, 1974.
- Stricker, L. J. A New Index of Differential Subgroup Performance: Application to the GRE Aptitude Test. 78-7P, 1981.
- Swinton, S. S. and Powers, D. E. A Factor Analytic Study of the Restructured GRE Aptitude Test. 77-6P, 1980.
- Ward, W. C. A Comparison of Free-Response and Multiple-Choice Forms of Verbal Aptitude Tests. 79-8P, 1982.
- Ward, W. C.; Frederiksen, N.; and Carlson, S. B. Construct Validity of Free-Response and Machine-Scorable Versions of a Test of Scientific Thinking. 74-8P, 1978.
- Ward, W. C. and Frederiksen, N. A Study of the Predictive Validity of the Tests of Scientific Thinking. 74-6P, 1977.