DOCUMENT RESUME

ED 360 329 TM 020 126

AUTHOR Andrada, Gilbert N.; Linden, Kathryn W.
TITLE Effects of Two Testing Conditions on Classroom Achievement: Traditional In-Class versus Experimental Take-Home Conditions.
PUB DATE Apr 93
NOTE 15p.; Paper presented at the Annual Meeting of the American Educational Research Association (Atlanta, GA, April 12-16, 1993).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Class Activities; Classroom Techniques; Cognitive Processes; *College Students; Comparative Testing; Higher Education; *Objective Tests; Psychometrics; Test Construction; *Testing Problems; Test Items; Test Reliability; *Test Validity; *Thinking Skills
IDENTIFIERS Blooms Taxonomy; *Take Home Tests

ABSTRACT

The psychometric properties of objective tests administered in two testing conditions were compared, using an experimental take-home testing condition and a traditional in-class testing condition. Subjects were 290 college students in a basic educational psychology course who took a test developed and tested the previous semester. Two equivalent 30-item tests (Form A and Form B) were randomly distributed. Students in Group 1 (n=141) took Form A in the in-class testing condition and Form B in the take-home testing condition. Students in Group 2 (n=149) took Form B in class and Form A at home. Results indicate that carefully constructed objective tests designed to measure higher-order thinking can function effectively under take-home conditions. Test items in the three taxonomic classifications (based on the Taxonomy of Cognitive Objectives of B. Bloom, 1956) did not have differential impacts on the test in the two conditions. Most student-reported variables had no influence on test performance. The psychometric properties of a test can remain intact even when students have ample time and course materials available provided that higher-order thinking skills are being measured. Seven tables present study findings. (SLD)

Effects of Two Testing Conditions on Classroom Achievement: Traditional In-Class versus Experimental Take-Home Conditions*

Gilbert N. Andrada and Kathryn W. Linden

Purdue University

Presented at the annual meeting of the American Educatinal Research Association, April 15, 1993, Atlanta, Georgia.

Effects of Two Testing Conditions on Classroom Achievement: Traditional In-Class versus Experimental Take-Home Conditions*

Gilbert N. Andrada and Kathryn W. Linden
Purdue University

Nearly 20 years ago, Robert Ebel suggested a dual use of the same classroom achievement test (Ebel, 1972). This plan included administration of a test consecutively to students in an in-class condition and to the same students in a take-home condition, the purpose being to promote "learning during testing" as well as to provide reliable and valid measurements of achievement. Given a dual use for a test, Ebel claimed that a composite of the two test scores provides a more valid measure of achievement than would the test administered only in one testing condition, thus capturing the "quickness of intellect" in the in-class condition and "persistence of effort" in the take-home condition. Another investigator (Foley, 1981) reported using "repetitive testing" (in-class followed by take-home testing with the same instrument) in psychology courses for many years in order to foster "the educational process beyond that of memorization tapped through recognition and recall to an understanding of the meaning of item content" (p.244). This procedure extends the instructional process for students to achieve course objectives. However, neither Ebel nor Foley offered empirical data to support this position.

The take-home testing issue has been investigated in a number of studies, but the results are inconclusive. The controls necessary to yield conclusive results about the psychometric properties of these tests were not maintained uniformly. For example, Weber and his colleagues (1983a, 1983b) used tests with low internal reliability, and Marsh (1984) did not control for the possible contamination of the in-class testing group by the take-home testing group. Results of the study conducted by Linden and Mazzuca (1977) supported the learning value of take-home testing, but the psychometric properties of the objective classroom achievement test employed in their study were not replicated fully in a subsequent administration of the same instrument. A large majority of each student group in on the Linden and Mazzuca study reported preference for the take-home testing condition because of its value as a good learning experience. The results of these studies indicate that a controlled investigation of the value of take-home testing is warranted, especially when the potential benefits can be as great as Ebel (1972) predicted and Foley (1981) supported.

The primary objective of the present investigation was to compare the psychometric properties of objective tests administered in two testing conditions: an experimental take-home testing condition and a traditional in-class testing condition. Of particular concern was the ability of the tests in the two testing conditions to detect significant differences in educational achievement levels among students. A second objective involved examining the effect of testing condition on scores from subsets of items classified according to knowledge, comprehension, and application levels of the Taxonomy of Cognitive Objectives (Bloom et al., 1956). In addition, factors that might contribute to changes in scores observed between the two testing conditions were investigated. These factors included the effect of books, helpfulness of books, time spent studying and taking tests under the two testing conditions, and preference for take-home tests versus in-class tests. Based on these objectives, three research questions were proposed for the present study: (1) Can take-home objective classroom tests function as effectively and consistently in delineating differential achievement levels as does the same test given in an in-class condition? (2) Do test items reflecting various taxonomic levels impact the results differentially for a test administered in the two testing conditions? and (3) What factors influence the performances of students under the two testing conditions?

Methods

All students in eight divisions of a basic educational psychology course at a large midwestern university during the fall semester of 1991 served as subjects (N = 290) for this study. The course was organized in two 2-hour blocks, with four divisions meeting on Monday-Wednesday and four on Tuesday-Thursday sequences. Course examinations during the semester were scheduled at night in order to control for time of day and day of week effects.

Instruments

In a preliminary phase of this study, a 60-item classroom test on topics in educational psychology was administered during the previous semester (spring, 1991) to 323 undergraduate students in order to provide the data for developing equivalent tests to be used for this study. Test items reflecting the objectives of instruction for the first third of the semester were based on five content categories and the first three taxonomic levels proposed by Bloom and his colleagues (1956). Two 30-item subtests were constructed, each including approximately half the items in each cell of the original Table of Specifications. Test form equivalency was determined initially from these spring 1991 data: 1) Means of score distributions for the two subtests did not differ significantly ($p \leq .01$); 2) Cronbach alpha reliability estimates for the two subtests were significant and similarly high in magnitude; 3) The Pearson product-moment correlation between subtests was significant and high in magnitude; and 4) Items in the two subtests functioned similarly with regard to difficulty level and discrimination power (Andrada, 1991). Therefore, these two 30-item tests demonstrated equivalent psychometric properties and were viewed to be relevant for this investigation.

Procedures

The two equivalent 30-item tests, color coded and labeled Form A and Form B, were distributed randomly during a night examination session to students in the basic educational psychology course in the fall of 1991 (N = 290). Group 1 ($n = 141$) took Form A in the in-class testing condition and Form B in the take-home testing condition. Group 2 ($n = 149$) took Form B in class and Form A at home. The difference in sample sizes resulted from confusion in handing out the two test forms at the beginning of the in-class testing session. More Form

B than Form A tests were inadvertently given to the students who came late to the testing session.

On completion of the in-class testing condition, students were given the appropriate alternate take-home test, together with a questionnaire that elicited responses for 11 factors thought to be related to test-taking performances (Zoller & Ben-Chaim, 1988). Included in the questionnaire were factors of time spent studying for class and for the test, predicted performances on the two forms of the test administered in the two testing conditions, time spent on taking each test, attitudes toward the take-home testing experience, and preference for testing condition. Tests, answer sheets, and questionnaires were to be returned to a designated classroom by 5 p.m. on the following day. All students met this deadline.

Research Design

Three independent variables were involved in this investigation: (1) testing condition (in-class and take-home); (2) test form (Form A and Form B); and (3) taxonomic classification of items (knowledge, comprehension, and application). Scores on Form A and Form B comprised the dependent variable. Descriptive statistics and correlations between testing conditions for each test form were derived in order to answer the first research question. The second research question was investigated by means of analysis of variance designs. A two-factor ANOVA design, with two levels of testing condition by three levels of taxonomic classification of items, was employed to compare the relative performances of subjects. In order to use all students as one sample, linear T-scores for each test form under each testing condition were derived, and improvement was expressed as the difference between in-class and take-home T-scores. Changes in raw score performances between testing conditions were also examined. Percentage scores based on subscores for each taxonomic level were used for these analyses because of the somewhat unequal numbers of items in these categories on Form A and Form B. A one-way ANOVA design was used to determine the possible impact of each of the selected factors thought to influence performance. The .05 level of significance was held critical for all tests on the data.

Results

The design of this experiment was constrained by the need to utilize equivalent forms of the test in order to deter cheating during the in-class administration of the test. Consequently, analysis of the data using the entire class as one sample was contingent on the finding that Form A and Form B were equivalent for both groups of students in both testing conditions, even though equivalence of forms had been determined by the data derived from a similar group of students the previous semester. Descriptive statistics for both test forms administered under the two testing conditions for the present study group are presented in Table 1.

No significant difference between the two forms administered in each testing condition was found with regard to mean scores and internal reliability estimates. Mean scores for both Form A and Form B administered in class were significantly lower than corresponding mean scores derived from the take-home testing condition, as would be expected. The score distributions for each test under both testing conditions were highly reliable and not significantly different from the mathematical normal distribution. Moreover, scores derived from both in-class and take-home testing conditions (Form A in-class by Form B take-home and the reverse) were highly correlated ($r_{AB} = .65$ and $r_{BA} = .71$ for in-class and take-home conditions, respectively).

Item analysis results for Form A and Form B taken under both testing conditions are presented in Table 2 and Table 3. Form B items were somewhat more effective than Form A items in discriminating differences in achievement under the in-class testing conditions: 28 discriminating items for Form B; 22 discriminating items for Form A. However, Form A discriminated better than Form B under the take-home testing condition: 23 discriminating items for Form A; 15 discriminating items for Form B. Some differences in item difficulty level were also observed between test forms under the two testing conditions. Nevertheless, item difficulty and discrimination values were similarly effective for both test forms administered in both testing conditions.

Significant main effects for both testing condition and thinking process level of items were identified by the two-factor ANOVA model (see Table 4). Results of the Scheffé post hoc analyses of means indicated that the means of subtest scores were significantly higher in the take-home condition than in the in-class condition. As expected, the mean for application items was higher than the mean for comprehension items, which in turn was higher than the mean for knowledge items. Moreover, the magnitude of subtest scores increased uniformly from knowledge, to comprehension, to application items between the in-class and take-home conditions. Also examined was the effect of testing condition (in-class or take-home) by test form. This two-factor ANOVA model yielded a significant main effect for testing condition but not for test forms (see Table 5). Students' scores on both test forms taken in the traditional in-class testing condition were significantly higher than their corresponding scores for the take-home testing condition.

Several factors were investigated that might influence student performances differently under the two testing conditions. No significant difference in student performances was found for (1) number of hours that students studied per week for this examination, (2) student attitudes toward take-home testing as a good learning experience, and (3) preference of students for take-home tests over in-class tests. Only two factors were found to be significant: (1) time spent responding to the take-home test compared to the in-class test response time; and (2) the helpfulness of books.

A one-way ANOVA with three levels of time difference (spent less time, spent about the same amount of time and spent more time on the take-home test as compared to the in-class test) was used in order to determine the degree to which students' reports of time spent during testing were related to improvement in test performance (reported as linear T-scores). The model yielded a significant main effect for test-taking time ($F(2,286) = 4.35$, $p < .05$; see Table 6). Results of the Scheffé test indicated that the source of the significant effect was between the SPENT LESS TIME and SPENT MORE TIME categories. Students who spent less time taking their take-home test compared to taking their test in-class test had lower gain scores than those of students who reported spending more time on their take-home test than their in-class test. Students who reported spending the same amount of time in both testing conditions did not differ significantly from the other two groups.

Students' ratings of the helpfulness of books were obtained by using a four-point scale: 1 = Very Helpful; 2 = Helpful; 3 = Made Little Difference; and 4 = Counter-Productive. A one-way ANOVA with three levels of helpfulness ratings (helpful, not helpful, and counter-productive) was used to determine the degree to which ratings of this factor were related to improvement in students' performances from the in-class test to the take-home test (reported as linear T-scores). The model yielded a significant main effect ($F(2,271) = 7.07$, $p < .05$; see Table 7). The Scheffè test indicated a significant difference was found between the HELPFUL category and both NOT HELPFUL and COUNTER-PRODUCTIVE categories. Students who indicated that using books was helpful increased their scores on the take-home test significantly more than did students who reported that books were either not helpful or were counter-productive. The difference between scores of students who rated the use of books as being not helpful or counter-productive was not significant.

Discussion

Can objective classroom tests function effectively and consistently to delineate achievement levels of students when administered in both in-class and take-home testing conditions?

The results of this study clearly indicated that carefully constructed objective tests designed to measure higher-order thinking can function effectively under a take-home testing condition. Test analysis results for both forms of the test in both conditions met the criteria for the use of norm-referenced evaluation (APA/AERA/NCME, 1985). The psychometric properties of Form A and Form B were equally strong in both testing conditions. Each test in each testing condition had moderately high internal reliability and strong differentiation among levels of student achievement. The correlation between in-class and take-home test scores indicated that, although students obtained higher scores in the take-home condition as compared to their in-class test scores, the ordinal ranking of students remained quite intact.

Equivalence of the two forms of the test was further evidenced by the non-significant main effect for test form. Moreover, no interaction between test form and testing condition was found. The two main effects found in these analyses were for testing condition and process level of items. The increase in test performance between in-class and take-home testing conditions was expected, if for no other reason than the fact that answers to knowledge-level items could easily be found in the textbooks and class notes. Students had been encouraged to utilize the take-home testing condition as a re-learning experience. Therefore, they were expected to use textbooks and course notes while responding to their respective test form during the take-home testing administration. The important issue then became the question of whether or nót this improvement in test scores would reduce the ability of the test to detect differences between students. Some gains in the psychometric properties of both test forms were observed for the take-home condition, most notably in the increased reliability of the tests compared to the same test forms given in the traditional in-class, closed-book testing condition. However, as discussed above, the psychometric properties of the take-home tests were not affected adversely.

Uniform gains in each process level subtest scores across test forms from in-class to take-home testing conditions were observed, indicating that gains in students' scores because of being able to look up answers to knowledge items were similarly found for responses to higher-order items. These results suggest that new learning requiring higher-order thinking skills had occurred during the take-home testing experience. Consequently, instruments designed to measure these types of thinking skills can be employed in a take-home testing condition to promote meaningful learning.

Do test items reflecting various taxonomic levels impact the results differentially for a test administered in the two testing conditions?

The answer to this question was negative. Items in all three taxonomic classifications were found to be significantly easier for both forms in the take-home testing condition than for the in-class condition. Answers to factual knowledge-level items could easily be found in course materials. However, selecting the correct response to the comprehension-level items require that students understand the relationship between facts and ideas, and application items require that students use the facts and relationships between facts to respond correctly to a new situation. The increased success on the comprehension and application items in the two test forms suggests that higher-order learning can be promoted by the take-home testing condition. Application items in both testing conditions were consistently the most difficult items, followed by comprehension items, and then knowledge items. These findings are similar to those reported by Kalish (1958) for the open-book testing condition. As expected, knowledge-level items did not contribute to differentiation among student achievement levels in the take-home testing condition, but comprehension and application items on both Form A and Form B were sufficient to produce score distributions that permitted accurate differentiation of student achievement levels.

What factors influence the performances of students under the two testing conditions?

Most of the student-reported variables investigated in this study had no significant impact on test scores, including preference for testing condition and attitude toward take-home tests as a good learning experience. A majority of students reported that they learned a great deal more during the take-home testing condition about the content being measured than in their preparation for a traditional in-class testing experience. It seems that the use of a take-home test provided another type of learning experience for many of the students. There is, however a cost for this learning experience in the currency of a teacher's time and measurement resources.

Driven by the findings reported by Kalish (1958) and by Weber and colleagues (1983b), the helpfulness of book materials on open-book test scores was investigated. Kalish (1958) reported that student ratings of the degree of help gained from book materials were not accompanied by similar gains from their in-class scores to their take-home test scores. In, other words, students who reported that they gained "no help" from book materials improved as much as students who reported that the books provided "much help." Contrary to the Kalish findings, the group of students in our study who reported that their books were helpful had a higher mean gain score than students reporting that books were either not helpful or counter-productive. These contradictory results might be attributed to differences in the types of thinking skills and/or test content reflected in Kalish's test compared to those employed in this study. Because this analysis was conducted using gain scores only, the ability levels of the students could not be investigated. It is possible that students who reported the helpfulness of books may have been students who scored poorly on the in-class test, thus making more meaningful gains on their take-home exam than other students. In any case, the ability to find answers contained in text material is a skill that every student should learn, and apparently take-home testing, at least in part, should include this skill in measuring student achievement.

Another objective for this investigation was to determine whether or not different amounts of preparation in terms of hours of study per week and number of hours spent studying for a specific exam had an effect on in-class and take-home test scores. Because students were not forewarned about the take-home examination, it seemed reasonable to expect that their preparation for this exam was not different than their typical preparation for an exam of this nature. Differences in the amount of time spent in preparation for this exam, as reported by students, did not significantly affect the amount of gain that students made between their in-class and take-home test scores. These results were congruent with those of Feldhusen (1961) who found that the type of examination employed had little effect on the amount of time spent reviewing for the exam.

Weber and colleagues (1983b) and Kalish (1958) reported that students spent a significantly greater amount of time taking a take-home test versus taking an in-class test. However, the greater amount of time spent taking the take-home test was not accompanied by systematically higher scores on the take-home test. In other words, students who spent the greatest amount of time on their take-home tests did not always have the highest take-home test scores. It was expected that the same type of result would be found in our investigation. At issue was the relationship between the amount of time used to complete the take-home test, as reported by students, compared to the amount of time spent taking the in-class test. This comparison was utilized in order to control for the different rates at which students take exams. Students who reported that they spent more time taking the take-home test than they did taking the in-class test earned higher scores than students who reported that they spend as much time or less time taking the take-home test than they spent taking the in-class test. These results are similar to those reported by Weber and Kalish. Moreover, the "persistence of effort" proposed by Ebel (1972) appears to be verified by these results.

Preference for take-home tests compared to in-class tests was not found to have a significant relationship to the gains in test scores from the in-class to the take-home testing conditions, although both Feldhusen (1961) and Zoller (1988) had reported that students preferred the take-home format of testing over both in-class and in-class open-book formats. However, differences in test scores for the two testing conditions were not investigated by these researchers. Test scores for students in our study were apparently not influenced by their preferences for testing condition.

Several researchers investigating take-home tests have pointed out the benefit to students of "learning while testing" (Feldhusen, 1961; Ebel, 1972; Linden and Mazzuca, 1977; and Marsh 1983b). However, the increases in scores between in-class and take-home tests were found to have no relationship to the examinees' attitudes toward take-home tests as good learning experiences. These insignificant results could have been influenced by the fact that students were not informed in advance about the nature of these testing conditions. Many students complained later about not wanting to be "surprised" with new testing procedures, even though it was explained that informing them in advance might influence them to change their test preparation habits. In discussions of these results with the students in each of the eight divisions of the course, it was clear that students like the idea of having objective take-home tests but want to know **in advance** the nature of the testing condition to be employed.

Conclusions

Take-home objective tests can be at least as effective in delineating differences in student achievement as in-class objective tests if a majority of the items measure higher-order thinking skills than what is required for factual knowledge-level items. In other words, the psychometric properties of an objective test can remain intact even when students have ample time and course materials available to them, provided that higher-order thinking skills are being measured. Including predominately knowledge-level items in an objective test to be administered in a take-home testing condition would conflict with the basic purpose of using tests to promote additional learning. Consequently, alternative testing and evaluation methods utilizing a take-home component can be explored with some confidence in the ability of a test designed to measure higher-order thinking skills to produce useful information about student achievement. The potential educational gains suggested by Ebel (1972) were evidenced in the present study, but additional investigations need to be undertaken in order to explore this question further.

Because a large majority of students reported that they learned a great deal more about the content being measured during the take-home testing condition than in their preparation for a traditional in-class testing experience, it was concluded that the use of a take-home test can provide a different type of learning experience for students. There is, however, a cost for this learning experience in the currency of a teacher's time and measurement resources. The most useful types of items for a take-home test are items that test higher-order thinking skills (e.g., comprehension, application, analysis, etc.). Items that test these skills are much more difficult to write than items measuring knowledge-level material. Furthermore, in the interest of test-security, items once used as take-home items are not as secure as items used on tests that are collected and stored immediately after testing has been concluded. Security issues require that the frequent take-home test consumer must have either a large item pool from which tests can be constructed or the time and skill needed to generate new items easily. However, our use of objective take-home tests strongly indicates that this security threat is not as damaging as one might expect.

Course instructors reported that, in reviewing these test results with students, many students expressed considerable ownership of their response choices and explanations of the answers to test items sparked healthy and productive discussions on the topics being assessed. Anyone who is considering the use of a take-home objective test is cautioned to prepare in advance not only a rationale of why an answer to each item is the best answer but also a rationale for why each of the distractors is less appropriate than the keyed answer. Student ownership of their response choices makes the task of successfully explaining why the incorrect alternatives are indeed not the best answer quite difficult at times. Instructors must be well prepared to defend their take-home tests.

An observation made by Ebel (1971), Foley (1981), and Linden and Mazzuca (1977) is supported by the results of this study. The time spent in testing need not be time spent away from learning. As these investigators have suggested, alternative testing conditions can be utilized for the dual purposes of combining a good testing situation with a good learning situation. The effects of objective tests, designed to assess higher-order thinking skills and administered as take-home tests, should be investigated further with different age groups and different courses.

References

AERA/APA/NCME (1985). *Standards for educational and psychological testing.* Washington, DC: American Psychological Association.

Andrada, G. N. (1991). *An experimental evaluation of the take-home examination.* Unpublished Master's Thesis, Purdue University.

Bloom, B., Englehart, M., Furst, E., Hill, W., & Krathwohl, O. (1956). *Taxonomy of educational objectives: The classification of educational goals: Handbook 1. The Cognitive domain.* New York: David McKay.

Ebel, R. L. (1972). *Essentials of educational measurement.* Englewood Cliffs, NJ: Prentice-Hall Inc.

Foley, D. P. (1981). Instructional potential of teacher-made tests. *Teaching of Psychology, 8,* 243-244.

Kalish, R. A. (1958). An experimental evaluation of the open-book exam. *Journal of Educational Psychology, 49,* 335-337.

Linden, K. W., & Mazzuca, S. A. (1977). Using tests to promote learning: Comparisons of two testing conditions. Presented at the annual meeting of the American Psychological Association, San Francisco.

Marsh, R. (1984). A comparison of take-home versus in-class exams. *Journal of Educational Research, 78,* 111-113.

Weber, L. J., McBee, J. K., & Krebs, J. E. (1983a). Cheating: A problem with take-home exams? Paper presented at the Annual Meeting of the National Council on Measurement in Education, Montreal, Canada. (ERIC Document Reproduction Service No. ED 242 730).

Weber, L. J., McBee, J. K., & Krebs, J. E. (1983b). Take home tests: An experimental study. *Research in Higher Education, 18* (2), 473-483.

Zoller, U. & Ben-Chaim, D. (1988). Interaction between examination type, anxiety state, and academic achievement in college science; an action-oriented research. *Journal of Research in Science Teaching, 26,* 67-77.

Table 1

Test Statistics for Form A and Form B: In-Class and Take-Home Testing Conditions

	Form A In-Class	Form B Take-Home	Form B In-Class	Form A Take-Home
No. Examinees:	141	140	149	149
No. Items:	30	30	30	30
Range:	21	15	21	20
Mean Score:	19.13	23.75	19.99	23.23
Standard Deviation:	4.28	3.05	4.36	3.72
Cronbach alpha:	0.70*	0.61*	0.72*	0.70*
Error of Measurement:	2.53	1.92	2.30	2.03
Average Item Difficulty:	64%	77%	67%	79%
Max Corr. Between forms		0.65		0.71
Obs. Correlation Form A/Form B		0.54*		0.56*
Normality Test:	0.93*	1.00*	0.93*	1.02*

* Significant at or beyond the α=.05 level

Table 2

Form A In-Class and Take-Home Item Analysis Summaries*

	Discrimination Index (H - L %)				
Difficulty Index (%)	.79 -.60	.59 -.40	.39 -.20	.19 to -.19	Total
Easy 1.0 to .75	0 (0)	1 (0)	5 (12)	4 (6)	10 (18)
Average .74 to .26	1 (3)	12 (4)	2 (4)	4 (1)	19 (12)
Hard .25 to .00	0 (0)	0 (0)	1 (0)	0 (0)	1 (0)
Totals	1 (3)	13 (4)	8 (16)	8 (7)	30 (30)

* Numbers of items for Take-Home Condition are in parentheses.

Table 3

Form B In-Class and Take-Home Item Analysis Summaries*

	Discrimination Index (H - L %)				
Difficulty Index (%)	.79 -.60	.59 -.40	.39 -.20	.19 to -.19	Total
Easy 1.0 to .75	0 (0)	3 (2)	8 (8)	1 (12)	12 (22)
Average .74 to .26	1 (1)	9 (3)	5 (1)	1 (3)	16 (8)
Hard .25 to .00	0 (0)	0 (0)	2 (0)	0 (0)	2 (0)
Totals	1 (1)	12 (5)	15 (9)	2 (15)	30 (30)

* Numbers of items for Take-Home Condition are in parentheses.

Table 4

Testing Situation by Item Taxonomic Level Classification ANOVA Table

Source	*df*	Sum of Squares	Mean Square	F-Ratio	Prob>F
A	1	72548.52	72548.52	252.44	0.0000
B	2	35373.82	17686.91	61.54	0.0000
AB	2	573.0518	286.5259	1.00	0.3690
S(AB)	1731	497466.8	287.387		
TOTAL(Adj)	1736	605977.6			

Table 5

Testing Situation by Test Form

Source	*df*	Sum of Squares	Mean Square	F-Ratio	Prob>F
A	1	72712.49	72712.49	236.37	0.0000
B	1	87.62051	87.62051	0.28	0.5936
AB	1	221.9357	221.9357	0.72	0.3975
S(AB)	1733	533119.1	307.6278		
TOTAL(Adj)	1736	605977.6			

Table 6

Test Scores by Time Spent Taking the Tests in Take-Home and In-Class Testing Conditions

Source	*df*	Sums of Squares	Mean Square	F-Ratio	Prob>F
A	2	765.9533	382.9767	4.350	.0137
S(A)	286	25160.27	87.97297		
TOTAL(Adj)	288	25926.22			

Table 7

Test Scores by Levels of Helpfulness of Books

Source	*df*	Sums of Squares	Mean-Square	F-Ratio	Prob>
A	2	1198.971	599.485	57.070	.001
S(A)	271	22971.85	84.76698		
TOTAL(Adj)	273	24170.82			