An Examination of the Speededness of the General Aptitude Test Battery Power Tests

Report

By

Christopher E. Sager, Norman G. Peterson,
and Scott H. Oppler

Addendum

Please note that the General Aptitude Test Battery (Forms E& F) referred to within this report has been renamed the Ability Profiler (Forms 1& 2). The name of the assessment was changed to reflect: 1) the focus on reporting a profile of score results from the instrument for career exploration purposes; 2) the technical improvements made to the assessment compared to previous forms of the instrument; and 3) the capacity to use the Ability Profiler in conjunction with other instruments to promote whole person assessment for career exploration.

Abstract

Concern has been expressed regarding the General Aptitude Test Battery's (GATB) use of relatively stringent time limits on tests (i.e., speeded tests) measuring constructs that are conventionally measured by tests with generous time limits (i.e., nonspeeded tests). This criticism was directed at the Computation, Three-Dimensional Space, Vocabulary, and Arithmetic Reasoning GATB power tests. This investigation was designed to address three major issues: (1) the extent to which nonspeeded versions of the tests referred to above measure the same constructs as the current speeded versions of these tests; (2) compared to the current speeded versions, the extent to which nonspeeded versions have the capacity to reduce mean subgroup differences; and (3) the effect of potential changes in instructions and item formats on speededness and mean subgroup differences. The investigation consisted of a field experiment with two levels of a within-groups independent variable (i.e., test speededness) and two levels of a between-groups independent variable (i.e., old instructions/format versus new instructions/format). Data were collected from 1,742 subjects at 6 state employment agencies. The data include a sufficient representation of relevant subgroups to compare the mean subgroup differences for speeded and nonspeeded versions of the GATB power tests. The results indicate that (a) with the exception of a speeded component, the speeded and nonspeeded versions of the GATB power tests measure the same constructs, (b) compared to the speeded versions some of the nonspeeded versions show a small reduction in some mean subgroup differences, and (c) the studied changes in instructions and format do not substantially affect the speededness of GATB power tests or the mean subgroup differences on these tests. Recommendations regarding nonspeeded operational versions of these tests are made.

Table of Contents

Table of Contents (continued)

Tables

Table of Contents (continued)

Figures

An Examination of the Speededness of the
General Aptitude Test Battery Power Tests

Introduction

The General Aptitude Test Battery (GATB) power tests (i.e., Computation, Three-Dimensional Space, Vocabulary, and Arithmetic Reasoning) were designed to measure psychological constructs generally measured with power tests; however, these tests appear to be more speeded than tests that measure similar constructs in other modern multi-aptitude test batteries (Peterson, 1993). Part of the definition of a power test is that it allows all examinees sufficient time to consider and attempt all items. Scores on power tests are often viewed as assessments of aptitude that are uninfluenced by rate of work.

The National Research Council's (NRC) Committee on the GATB (Hartigan & Wigdor, 1989) expressed three major concerns about the relative speededness of the GATB power tests. Peterson (1993, p. 1) outlines these:

- the meaning of constructs measured by speeded power tests may be different from the meaning conventionally attached to those constructs;

- the speed component of the tests may cause the tests to be differentially valid for different racial or ethnic groups;

- "the severe time limits of the GATB subtests might produce an adverse psychological reaction in examinees as they progress through the examination and might thereby reduce the construct validity of the subtests" (p. 106).

For a complete review of the literature relevant to the committee's concerns and other issues associated with the speededness of the GATB power tests see this report's companion document titled Review of Issues Associated with Speededness of GATB Tests, (Peterson, 1993).

Objectives

In an effort to address the concerns outlined above, the investigation described in this report had six primary objectives:

1. Evaluate the feasibility of constructing nonspeeded versions of the GATB power tests that can be administered within the time limits of the current operational forms.

2. Assess the extent to which speededness influences the magnitude of subgroup difference in scores on the GATB power tests. The subgroups of interest in this investigation are Whites, Blacks, Hispanics, Males, Females, Under 40 Years of Age, and 40 Years of Age and Over.

3. Assess the extent to which nonspeeded versions of the GATB power tests measure the same constructs as operational (i.e., speeded) versions of these tests.

4. Assess the extent to which the relationships among the speeded and nonspeeded tests are the same across the relevant subgroups.

5. Assess the effects on test scores of changes to instructions, item formats, and answer sheet formats anticipated for future operational forms of the GATB power tests. This objective is included because of concerns that these changes might affect the speededness of these tests.

6. Based on the evaluations and assessments outlined in the first five objectives, make recommendations about the number of items and time limits for future operational forms of the GATB power tests.

Method

<u>Research Design</u>

This investigation included a within-groups independent variable and a between-groups independent variable. Table 1 illustrates the research design. The within-groups variable was speededness. This independent variable allows for inferences about the effects of speededness on performance on the GATB power tests (i.e., Computation, Three-Dimensional Space, Vocabulary, and Arithmetic Reasoning). The speed test Name Comparison (a test designed to measure a speed construct) was included to allow for inferences about how the speededness of power tests affects their relationships with a test measuring a speed construct. The between-groups variable was old instructions/format versus new instructions/format. This independent variable allows for inferences about the effects of potential changes in instructions and format anticipated for future operational GATB Forms E and F.

<u>Sample</u>

Participants were recruited and data were collected at six state employment service offices:

- Baltimore, Maryland (Downtown Office)
- Baltimore, Maryland (East Point Office)
- Brownsville, Texas
- Chattanooga, Tennessee
- El Paso, Texas (Airport Office)
- Frederick, Maryland

Participation was restricted to individuals who (a) were 16 years of age or older, (b) had not taken the GATB within the last 12 months, and (c) were fluent in English.

The data collection sites listed above were selected by the U.S. Department of Labor to ensure that there would be a sufficient number of examinees from each of a number of subgroups to support the investigation of subgroup differences. The goal was to include 300 Whites, 300 Blacks, and 300 Hispanics in each of the two between-groups conditions, for a total of 1,800 subjects. It was assumed that this sample would include a sufficient number of Females to address Male/Female differences and a sufficient number of individuals 40 years of age and over to address Under 40 Years of Age-40 Years of Age and Over differences.

<u>Measures</u>

Tables 2 and 3 depict the measures that were administered to the examinees. The speeded old instructions/format tests were the actual operational Form A power tests (i.e., Computation, Three-Dimensional Space, Vocabulary, and Arithmetic Reasoning) and the speeded test Name Comparison. The speeded new instructions/format tests consisted of items sampled from these operational Form A tests. The nonspeeded old and new instructions/format tests ·consisted of items sampled from the operational Form D power tests.

3

Table 1.   Experimental Design[a].

Old Instructions/Format	New Instructions/Format

Speeded Tests
 	
■   Old Instructions/Format
 	
■   Old Number of Items in Each Test
 	
■   Old Time Limits for Each Test
 	
■   Tests
    -   Name Comparison
    -   Computation
    -   Three-Dimensional Space
    -   Vocabulary
    -   Arithmetic Reasoning
	Speeded Tests
	
	■ New Instructions/Format
	
	■ Number of Items suggested by the Forms E & F Study
	
	■ Old Time Limits for Each Test
	
	■ Tests
	    - Name Comparison
	    - Vocabulary
	    - Arithmetic Reasoning
	    - Three-Dimensional Space
	    - Computation

Nonspeeded Tests
 	
■   Old Instructions/Format
 	
■   Reduced Number of Items in Each Test (i.e., nonspeeded versions of the power tests)
 	
■   Old Time Limits for Each Test Except for a Extended Time Limit for Arithmetic Reasoning
 	
■   Tests
    -   Computation
    -   Three-Dimensional Space
    -   Vocabulary
    -   Arithmetic Reasoning
	Nonspeeded Tests
	
	■ New Instructions/Format
	
	■ Reduced Number of Items in Each Test (i.e., nonspeeded versions of the power tests)
	
	■ Old Time Limits for Each Test Except for a Extended Time Limit for Arithmetic Reasoning
	
	■ Tests
	    - Vocabulary
	    - Arithmetic Reasoning
	    - Three-Dimensional Space
	    - Computation

Note: [a] = Each double lined box represents the two levels of the between-groups independent variable. The top and bottom halves of each box represent the two levels of the within-groups independent variable. This table shows the speeded tests being administered to the examinees first, followed by the nonspeeded tests; in this investigation the order of presentation of speeded and nonspeeded tests was counter-balanced.

Table 2.   Data Collected From Examinees in the Old Instructions/Format Condition.

Activity		Number of Items	Time in Minutes: (Instr. + Admin.) = Subtotal
Start-up		–	10
Speeded Tests	Name Comparison	150	4 + 6 = 10
	Computation	50	4 + 6 = 10
	Three-Dimensional Space	40	4 + 6 = 10
	Vocabulary	60	4 + 6 = 10
	Arithmetic Reasoning	25	4 + 7 = 11
Break		–	10
Non-Speeded Tests	Computation	12	4 + 6 = 10
	Three-Dimensional Space	15	4 + 6 = 10
	Vocabulary	14	4 + 6 = 10
	Arithmetic Reasoning	10	4 + 11 = 15
Debriefing/Subject Payment		–	15
Total Time			2 hrs. 11 mins.

Table 3.   Data Collected From Examinees in the New Instructions/Format Condition.

Activity		Number of Items	Time in Minutes: (Instr. + Admin.) = Subtotal
Start-up		--	10
Speeded Tests	Name Comparison	90	4 + 6 = 10
	Vocabulary	50	4 + 6 = 10
	Arithmetic Reasoning	24	4 + 7 = 11
	Three-Dimensional Space	35	4 + 6 = 10
	Computation	40	4 + 6 = 10
Break		--	10
Non-Speeded Tests	Vocabulary	14	4 + 6 = 10
	Arithmetic Reasoning	10	4 + 11 = 15
	Three-Dimensional Space	15	4 + 6 = 10
	Computation	12	4 + 6 = 10
Debriefing/Subject Payment		--	15
Total Time			2 hrs. 11 mins.

Old instructions/format versus new instructions/format. In the old instructions/format condition, the speeded and nonspeeded versions of the tests were administered with the present operational GATB instructions, the items were in their present (i.e., operational) format, and the answer sheets followed the current format.

In the new instructions/format condition, the speeded and nonspeeded versions of the tests were administered with instructions similar to those that were anticipated for future forms of the GATB, as of September, 1993, and that were used in the Forms E and F item pretest study. In addition the instructions for the power tests in the new instructions/format condition included the following statement:

> You should provide an answer to every question you read, even if you are not sure of the answer. If you don't know the answer to a question, **take your best guess** and go on to the next one. Don't answer questions that you don't have a chance to read and consider.

The speeded new instructions/format, nonspeeded new instructions/format, and the nonspeeded old instructions/format booklets are in an unattached appendix.

The formats of the items administered in the new instructions/format condition were edited to reflect anticipated item format changes (e.g., in the Vocabulary test the four words within each item will be listed vertically instead of horizontally), and the answer sheets also followed the anticipated format. Additionally, as indicated in Table 3, the power tests were administered in a different order to represent the change in test order anticipated for operational Forms E and F.

Another aspect of the new instructions/format condition was that the number of items in the speeded version of each test was smaller than the number of items in its corresponding speeded old instructions/format test. The actual differences can be observed by comparing the numbers of items depicted in the "Speeded Tests" portions of Tables 2 and 3. This smaller number of items was achieved for the speeded new instructions/format Computation, Three-Dimensional Space, and Vocabulary tests by eliminating every i^{th} operational Form A item. For example, the speeded new instructions/format Computation test included all but every 5th item of the speeded old instructions/format Computation test. The speeded new instructions/format Arithmetic Reasoning test consisted of all but the last item of the speeded old instructions/format Arithmetic Reasoning test. Finally, the new instructions/format Name Comparison test consisted of the first 90 items of the old instructions/format Name Comparison test.

Speededness. As indicated in Tables 1, 2, and 3, speeded and nonspeeded versions of the GATB power tests were administered to all examinees. The speeded tests included one speed GATB test (i.e., Name Comparison) and the four GATB power tests (i.e., Computation, Three-Dimensional Space, Vocabulary, and Arithmetic Reasoning), all administered under

7

speeded conditions (i.e., the existing operational time limits and approximate numbers of items). As indicated above, the speeded tests consist of items from operational Form A.

The nonspeeded tests were edited versions of the operational Form D power tests. For the purposes of this investigation, the nonspeeded version of a power test would optimally contain the number of items that 90% of the examinees would be able to complete within the test's time limit. In Hartigan and Wigdor (1989), "A power test is defined operationally as one where 90 percent of examinees have sufficient time to complete all of the test items" (p. 103); however, there are a number of definitions of a power test, see Peterson (1993). To create nonspeeded versions of the GATB power tests, it was decided that instead of increasing each test's time limit, it would be preferable to reduce the number of items on each test. This is because (a) a considerable increase in the GATB administration time is currently not considered operationally feasible and (b) based on the results from administering such nonspeeded versions, the Spearman-Brown Prophecy formula could be used to estimate the reliability of nonspeeded versions of any length. However, another issue is the desire to avoid a nonspeeded version of a power test that has so few items that its capacity to represent its intended content domain is severely diminished. Although the goal of constructing nonspeeded versions of the GATB power tests was to achieve "power" status by some formal definition, the most important outcome was the creation of versions of the tests that were much less speeded than the operational versions, but that were in all other respects as similar to the operational versions as possible.

Information about the number of items completed by 90% of the examinees, under current time limits and numbers of items, was provided in a memorandum by Steve Mellon (June 22, 1993). The problem is that these data were based on speeded administrations of the tests; therefore, the data probably represent overestimates of the number of items that 90% of the examinees could have completed if the items were administered under nonspeeded conditions.

The Forms E & F study data provided another source of information about the number of items that could have been completed by 90% of the examinees. Part of the Forms E & F study consisted of administering long (i.e., more than the operational number of items) versions of the power tests under nonspeeded conditions. These data provided an estimate of the average amount of time required to complete each item under nonspeeded conditions and the capacity to estimate the internal-consistency reliability of nonspeeded tests of different lengths. The Spearman-Brown Prophecy formula was used to make these estimates. These internal-consistency reliability estimates provided additional information on which to base the decisions regarding the number of items to include in the nonspeeded tests.

The decisions regarding the number of items to include in the nonspeeded tests and their associated time limits were based on the issues and data discussed in the previous three paragraphs and a September 23, 1993 meeting with Department of Labor officials. The information used to make these decisions is presented in Appendix A. Tables 2 and 3 indicate the number of items and time limit for each nonspeeded test.

8

The process of selecting items from the operational Form D tests consisted of six steps.

1. The number of items in each operational Form A power test was divided by the number of items to be included in the nonspeeded version of the test. This value will be referred to as i.

2. Every i^{th} item in the Form A test was identified.

3. Every i^{th} item in the Form D test was identified.

4. Comparisons were made between each identified Form D item (Step 3) and its corresponding Form A item (Step 2) to ensure that they were isomorphic in content.

5. Substitutions were made in each subset of Form D items to exclude items with seriously outdated content (e.g., items referring to stereotypical sex roles).

6. Each subset of Form D items was then examined to ensure that they complied with the distribution requirements recommended in a draft version of a test development guide (U.S. Department of Labor, in prep.).

This process resulted in a nonspeeded version of each of the power tests, using items culled from operational Form D power tests, that were as parallel as possible to its corresponding speeded operational Form A power test. Note that the items in the operational forms of the GATB power tests are ordered by increasing difficulty. This "i^{th} item" process was used to ensure content and difficulty comparability; a desirable side effect is that it also operated to reduce the probability of ceiling effects in the nonspeeded versions.

As indicated above, speededness was a within-groups independent variable in this investigation. Tables 1, 2, and 3 show the speeded tests being administered to the examinees first, followed by the nonspeeded tests; in this study the order of presentation of the speeded and nonspeeded tests was counter-balanced across sessions so that carryover effects would be equalized.

Procedures

American Institutes for Research (AIR) personnel collected data at the six locations listed above. AIR assigned one data collector to each location. This individual's primary responsibilities were to schedule participants and administer the speeded and nonspeeded forms of the GATB tests to the participants according to the schedules shown in Tables 2 and 3. Two of the less experienced data collectors were accompanied by more senior personnel for the first two days of their data collection.

9

The schedule consisted of two test administrations each day; the first was from approximately 9:00 a.m. to 11:30 a.m., and the second was from approximately 1:00 p.m. to 3:30 p.m.. Each day the data collector consulted a schedule that randomly determined the administration of the new instructions/format condition during the morning and the old instructions/format condition in the afternoon or vice versa. Again, the order of presentation of the speeded and nonspeeded tests was counterbalanced. Each subject was reimbursed $15.00 for his/her participation in compensation for transportation expenses.

Results

This section includes the results of analyses directed toward (a) removing "problem" examinees from the data base to be analyzed, when appropriate, (b) examining the effects of speededness and instructions/format on performance on the GATB power tests, (c) examining and comparing differences in performance on the speeded and nonspeeded versions of the power tests across relevant subgroups (e.g., Whites and Blacks) and instructions/format, and (d) assessing the extent to which the speeded and nonspeeded versions of these tests measure the same psychological constructs.

Examinees Removed for Purposes of Further Analyses

The data and the data collection logs were examined to identify examinees whose data should not be included in further analyses. The database started with complete data for 1,742 examinees. 61 examinees were removed for the purpose of the analyses.

- Six examinees were removed because of comments in the data collection logs (e.g., a few examinees continued responding to test items after they were told to stop).

- 45 examinees were removed because of excessive construction noise in the adjacent room at the Baltimore, Maryland (Downtown Office) during two sessions.

- 10 examinees were removed because they filled in two alternatives per item on several items. Most of these examinees apparently misunderstood the Vocabulary test instructions.

Analyses were performed on 1,681 examinees (867 examinees in the old instructions/format condition and 814 examinees in the new instructions/format condition). The number of examinees in the old instructions/format condition is greater than the number of examinees in the new instructions condition because all of the 45 examinees removed due to construction noise were in the new instructions/format condition. The data were examined for out-of-range values; none were discovered.

Speededness

The Educational Testing Service (ETS) uses a number of rules-of-thumb to assess the extent to which a "nonspeeded" test is actually nonspeeded (Donlon, 1973). One of these rules-of-thumb states that a test is speeded if fewer than 100% of the examinees reach 75% of the items and if fewer than 80% of the examinees reach all of the items. A common definition of the number of items reached by an examinee is the last item responded to by the examinee (i.e., the highest numbered item to which the examinee responded). This report uses a more conservative index; it is the number of items attempted by each examinee, (i.e., the number of items that the examinee did not leave blank).

Table 4 shows the percentage of examinees who attempted 75% of the items and the percentage of examinees who attempted all of the items on each nonspeeded power test in both conditions. For example, in the old instructions/format condition, 98.2% of the examinees attempted 75% of the Three-Dimensional Space items, and 88.1% of the examinees attempted all of the Three-Dimensional Space items. Table 4 indicates that for the total groups in each condition the nonspeeded versions of the Three-Dimensional Space, Vocabulary, and Arithmetic Reasoning tests reasonably satisfy the ETS rule-of thumb for nonspeeded tests. However, Computation seems to fall short of satisfying this rule. While a fairly high percentage of each total group of examinees attempted 75% of the Computation items in each condition (91.0% and 88.2%), only about half of the examinees attempted all of the Computation items in each condition (53.5% and 43.1%).

Figures 1 through 4 represent another way to examine the speededness of this investigation's "nonspeeded" tests. Figure 1 plots the percent of examinees who attempted each possible number of items on the nonspeeded Computation tests in each condition. In this figure the number of items attempted is on the horizontal axis and percentage of examinees is on the vertical axis. The diagonal solid dark line in the upper-right corner represents the ETS rule-of-thumb. In Figure 1 the line starts at 75% of the items (i.e., 10) and 100% of the examinees and ends at all of the items (i.e., 12) and 80% of the examinees. The extent to which a plotted line is below the ETS "line" is one method of assessing the extent to which the test is speeded. Figure 1 indicates that the old instructions/format and the new instructions/format nonspeeded Computation tests do not differ substantially in speededness and that they both miss a strict definition of a nonspeeded test. Figures 2 through 4 provide the same information for the nonspeeded versions of the Three-Dimensional, Vocabulary, and Arithmetic Reasoning tests in both conditions. These figures indicate that the instructions/format variable has virtually no effect on the speededness of these tests and that all three tests come very close to satisfying the ETS rule-of-thumb for nonspeeded tests. Figures showing these comparisons across subgroups are presented in Appendix B.

Tables 5 and 6 represent a third way to assess the speededness of the tests used in this investigation that allows a direct comparison of the relative speededness of the operational and nonspeeded versions. Table 5 shows the means and standards deviations of five alternative scores for all the tests in the old instructions/format conditions. The Correct score is simply the examinee's raw test score. The Wrong score is the number of items that the examinee responded to incorrectly. The Answered score is the number of items the examinee attempted (i.e., responded to). The Proportion Attempted is the number of items answered divided by the number of items possible. The Accuracy score is the number of items correct (i.e., raw score) divided by the number of items answered. The columns showing mean Proportion Attempted and Accuracy are in bold. These two columns of Table 5 indicate that the nonspeeded Three-Dimensional Space, Vocabulary, and Arithmetic Reasoning old instructions/format tests are consistent with the traditional definition of a power test (Peterson, 1993). That is, examinees were able to attempt a very high percentage of the items and the items are relatively difficult. The nonspeeded Computation old instructions/format test misses this definition because the mean proportion attempted is only 0.92 (somewhat smaller than the other nonspeeded tests) and the items are on average not very difficult (mean Accuracy = 0.75). However, the nonspeeded Computation old instructions/format test is much less

12

Table 4. Percentage Attempting 75% of the Items and All of the Items in the Nonspeeded Versions of the Power tests, by Condition and Subgroup.

Condition	Computation (12 Items)		3-D Space (15 Items)		Vocabulary (14 Items)		Arithmetic Reasoning (10 Items)	
	75% of Items	All Items	75% of Items	All Items	75% of Items	All Items	75% of Items	All Items
Total Group								
Old Instr./Format	91.0%	53.5%	98.2%	88.1%	97.2%	85.8%	98.0%	84.8%
New Instr./Format	88.2%	43.1%	97.2%	82.2%	97.4%	87.6%	97.8%	87.8%
Whites								
Old Instr./Format	94.5%	57.4%	97.9%	86.4%	97.4%	88.5%	99.6%	88.1%
New Instr./Format	93.1%	51.9%	96.1%	77.5%	99.1%	90.5%	98.7%	90.0%
Blacks								
Old Instr./Format	86.3%	43.8%	97.8%	85.6%	96.2%	77.6%	98.4%	81.2%
New Instr./Format	77.4%	31.8%	96.1%	79.5%	95.1%	79.9%	96.8%	85.9%
Hispanics								
Old Instr./Format	93.4%	60.8%	98.7%	91.7%	98.3%	92.7%	96.7%	85.7%
New Instr./Format	95.0%	47.7%	99.3%	89.6%	98.6%	93.2%	98.2%	88.5%
Males								
Old Instr./Format	89.5%	51.0%	98.3%	86.5%	96.8%	83.7%	98.1%	85.0%
New Instr./Format	87.2%	41.1%	97.2%	81.9%	97.0%	87.0%	98.3%	88.5%
Females								
Old Instr./Format	93.3%	57.4%	98.2%	90.9%	97.9%	89.1%	97.9%	84.2%
New Instr./Format	90.0%	47.6%	97.0%	83.0%	98.2%	88.6%	96.7%	86.3%
< 40								
Old Instr./Format	92.0%	56.9%	98.3%	88.7%	98.3%	86.3%	98.8%	86.9%
New Instr./Format	90.6%	46.6%	97.9%	84.6%	97.9%	90.1%	98.6%	90.3%
≥ 40								
Old Instr./Format	88.3%	43.5%	97.7%	86.0%	94.4%	85.0%	95.8%	79.0%
New Instr./Format	82.3%	34.6%	95.2%	76.6%	96.5%	81.4%	95.7%	81.8%
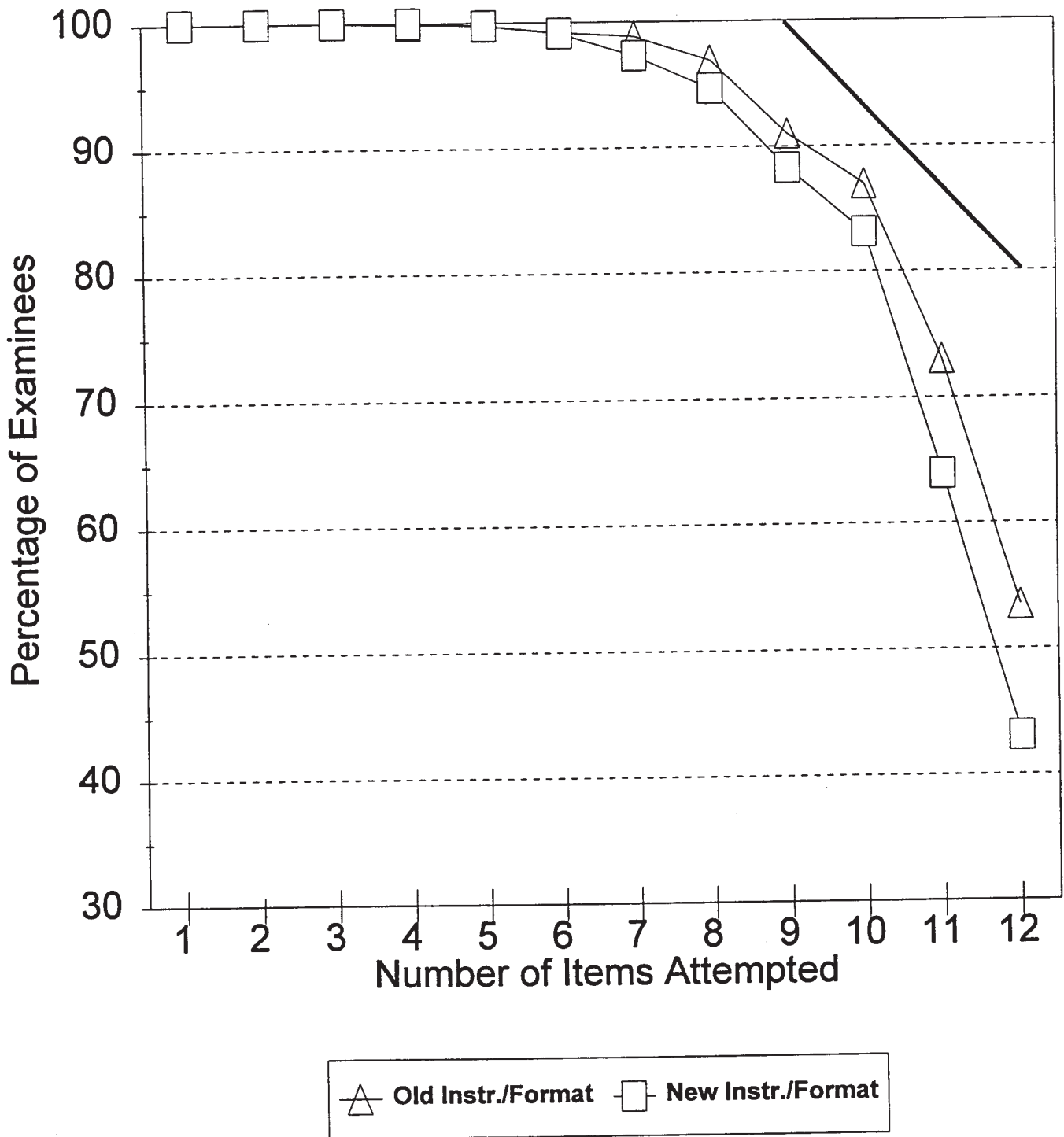
Note: The diagonal solid dark line in the upper-right corner starts at 75% of the items and
100% of the examinees and ends at all of the items and 80% of the examinees.

Figure 1. Percentage of Examinees Attempting Nonspeeded Computation Items in the Old
Instructions/Format and the New Instructions/Format Conditions.
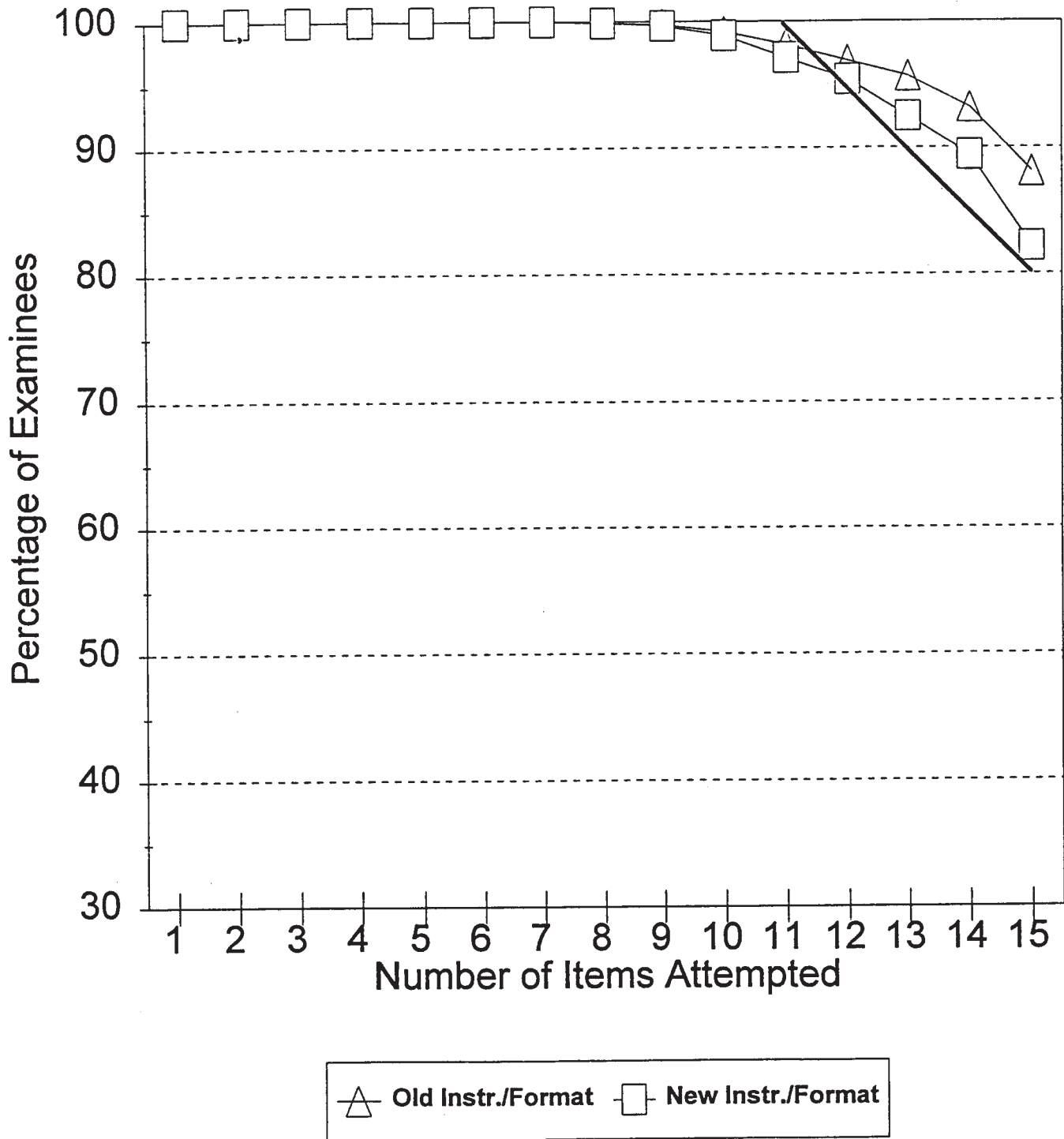
Note: The diagonal solid dark line in the upper-right corner starts at 75% of the items and 100% of the examinees and ends at all of the items and 80% of the examinees.

Figure 2. Percentage of Examinees Attempting Nonspeeded Three-Dimensional Space Items in the Old Instructions/Format and the New Instructions/Format Conditions.
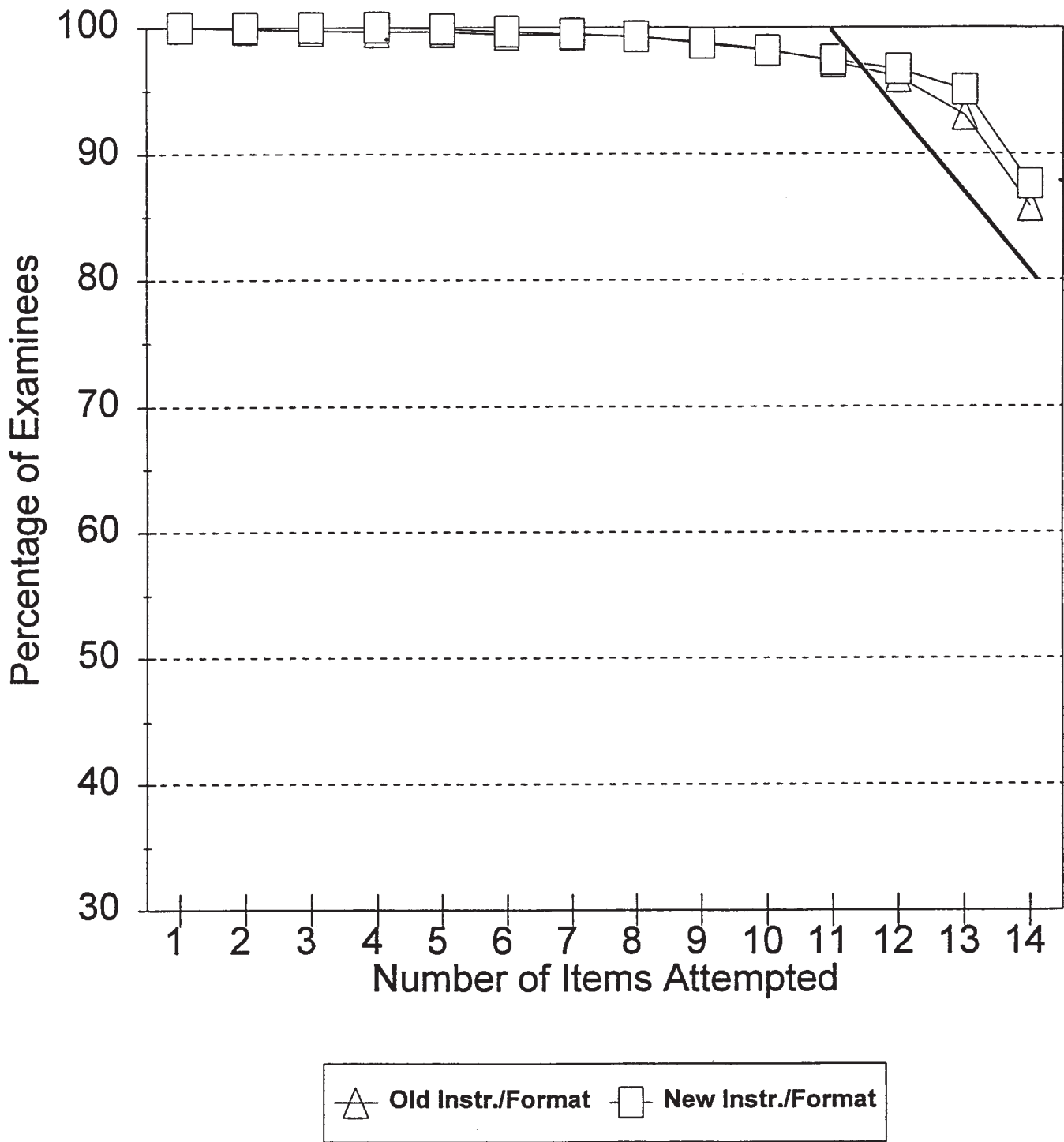
Note: The diagonal solid dark line in the upper-right corner starts at 75% of the items and 100% of the examinees and ends at all of the items and 80% of the examinees.

Figure 3. Percentage of Examinees Attempting Nonspeeded Vocabulary Items in the Old Instructions/Format and the New Instructions/Format Conditions.
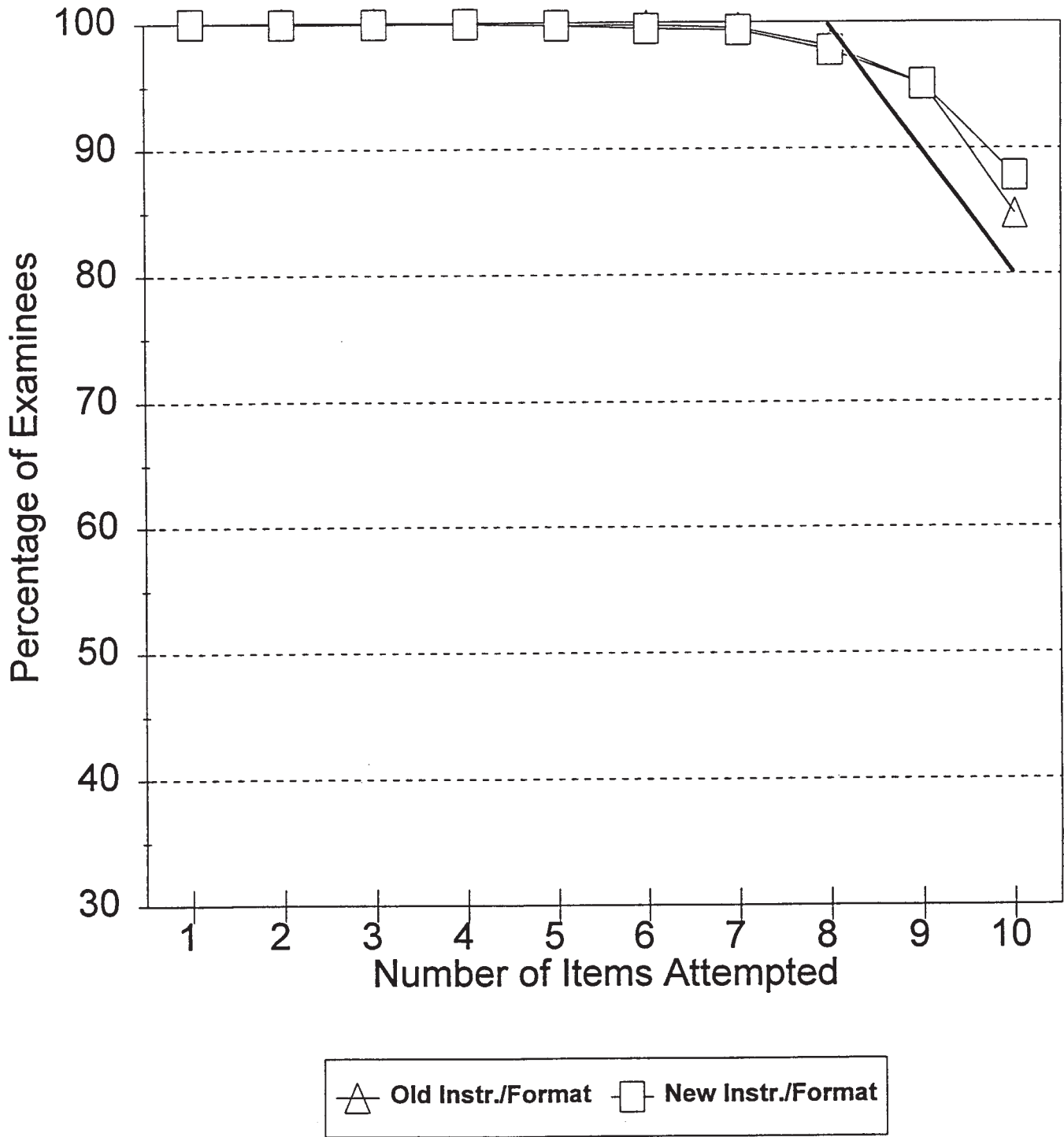
Note: The diagonal solid dark line in the upper-right corner starts at 75% of the items and 100% of the examinees and ends at all of the items and 80% of the examinees.

Figure 4. Percentage of Examinees Attempting Nonspeeded Arithmetic Reasoning Items in the Old Instructions/Format and the New Instructions/Format Conditions.

Table 5.   Five Alternative Scores on Old Instr./Format: Total Sample (N=865).

Test	Correct		Wrong		Answered		Proportion Attempted		Accuracy	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Speeded Name Comparison	40.28	12.67	5.31	6.83	45.59	13.99	0.30	0.09	0.88	0.12
Speeded Computation	19.52	6.10	3.63	4.16	23.15	5.58	0.46	0.11	0.84	0.75
Nonspeeded Computation	8.21	2.47	2.78	2.22	10.99	1.45	0.92	0.12	0.75	0.20
Speeded Three Dimensional	14.83	6.15	12.09	8.58	26.93	7.31	0.67	0.18	0.58	0.24
Nonspeeded Three Dimensional	7.04	2.94	7.67	3.09	14.71	0.99	0.98	0.07	0.48	0.20
Speeded Vocabulary	15.27	7.23	10.39	7.83	25.66	7.83	0.43	0.13	0.60	0.23
Nonspeeded Vocabulary	7.18	3.29	6.49	3.22	13.66	1.19	0.98	0.08	0.52	0.23
Speeded Arithmetic Reasoning	8.48	3.44	5.03	4.37	13.51	4.15	0.54	0.17	0.65	0.24
Nonspeeded Arithmetic Reasoning	5.12	2.06	4.65	2.12	9.77	0.66	0.98	0.07	0.53	0.21

Note: Proportion Attempted = Number Answered / Number of Items.
Accuracy = Number Correct / Number Answered.
The sample size in this table is two smaller than the sample size indicated in the Sample section (N = 867) because two examinees who did not answer any items on a test were eliminated for the purposes of this table. This was done so that calculation of the Accuracy score would not result in a value of 0 in the denominator.

18

Table 6.   Five Alternative Scores on New Instr./Format: Total Sample   (N=812).

Test	Correct		Wrong		Answered		Proportion Attempted		Accuracy	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Speeded Name Comparison	44.96	13.49	5.87	6.15	50.83	13.67	0.56	0.15	0.88	0.12
Speeded Computation	17.71	4.89	3.56	3.85	21.27	4.72	0.55	0.12	0.83	0.15
Nonspeeded Computation	8.31	2.38	2.39	1.99	10.69	1.59	0.89	0.13	0.77	0.19
Speeded Three Dimensional	14.42	5.76	13.60	7.74	28.02	6.32	0.80	0.18	0.53	0.22
Nonspeeded Three Dimensional	7.33	2.99	7.22	3.18	14.55	1.20	0.97	0.08	0.51	0.21
Speeded Vocabulary	15.40	6.69	12.90	8.29	28.30	8.03	0.57	0.16	0.55	0.21
Nonspeeded Vocabulary	7.08	3.18	6.64	3.08	13.72	1.05	0.98	0.07	0.51	0.23
Speeded Arithmetic Reasoning	8.83	3.52	5.50	4.43	14.33	4.31	0.60	0.18	0.64	0.23
Nonspeeded Arithmetic Reasoning	5.30	2.05	4.49	2.06	9.79	0.71	0.98	0.07	0.54	0.21

Note: Proportion Attempted = Number Answered / Number of Items.
Accuracy = Number Correct / Number Answered.
The sample size in this table is two smaller than the sample size indicated in the Sample section (N = 814) because two examinees who did not answer any items on a test were eliminated for the purposes of this table.   This was done so that calculation of the Accuracy score would not result in a value of 0 in the denominator.

19

speeded than the speeded version of this test (mean proportion attempted 0.92 versus 0.46). These columns also indicate that speeded Name Comparison old instructions/format test is consistent with the traditional definition of a speeded test (Peterson, 1993). That is, examinees were not able to complete all of the items and the items are relatively easy. Finally, these columns suggest that the speeded old instructions/format tests do not satisfy the definition of a power test or a speeded test. That is, a large proportion of the examinees were not able to complete the test and the items are relatively difficult. The exception is the speeded Computation old instructions/format test that approximates the properties of a speeded test. Table 6 shows the same results for the new instructions/format condition. The values in Table 6 are substantially the same as the values shown in Table 5. When comparing the values in Tables 5 and 6 recall that the speeded new instructions/format tests contain slightly fewer items than the speeded new instructions/format tests. This difference across conditions could explain the generally higher mean proportion of items attempted values in the new instructions/format condition (Table 6). Appendix C contains these tables for both conditions within each subgroup.

Summary. The results presented above consistently indicate that the nonspeeded versions of the Three-Dimensional Space, Vocabulary, and Arithmetic Reasoning GATB power tests in the old instructions/format and new instructions/format conditions are actually nonspeeded. They reasonably satisfy the ETS rule-of-thumb discussed above and they come fairly close to satisfying Hartigan and Wigdor's (1989) operational definition of a power test that indicates that 90% of examinees must have sufficient time to complete all of the test items (see Table 4). In any case, they are considerably less speeded than the operational versions of the power tests. However, the nonspeeded versions of Computation in the old instructions/format and new instructions/format conditions are not judged to be nonspeeded in the absolute sense (i.e., they do not satisfy the ETS rule-of-thumb or the Hartigan and Wigdor operational definition of a nonspeeded test). More importantly, for this investigation, they are considerably less speeded than the operational versions.

Subgroup Differences

Means and standard deviations. Tables 7 through 10 show the means and standard deviations for each version of each power test (i.e., speeded old instructions/format, nonspeeded old instructions/format, speeded new instructions/format, and nonspeeded new instructions/format). The tables include these values for the total group and by subgroup for each condition. The tables also indicate the number of items in each version of each power test. For example, Table 7 shows that there are scores for 235 Whites on the 50 item speeded old instructions/format Computation test and that this subgroup has a mean score of 21.61 and a standard deviation of 6.18 on this test. (Note: The means and standard deviations for the speeded new instructions/format Computation test in Table 7 are based on 39 items, although Table 3 indicates that this test contains 40 items. Item number 12 of this test was not scored for this investigation's analyses because, due to a test creation error, this item is a duplicate of item 10.)

20

Table 7. Means and Standard Deviations for Each Version of the Computation Test, by Subgroup.

Sub-group	n	Speeded Old Instr./Format (50 Items)		Nonspeeded Old Instr./Format (12 Items)		n	Speeded New Instr./Format (39 Items)		Nonspeeded New Instr./Format (12 Items)	
		Mean	SD	Mean	SD		Mean	SD	Mean	SD
All	867	19.53	6.10	8.21	2.47	814	17.68	4.93	8.30	2.38
White	235	21.61	6.18	8.77	2.23	231	19.48	4.66	8.94	2.18
Black	313	17.75	6.15	7.66	2.59	283	16.07	5.03	7.54	2.54
Hispanic	301	19.79	5.47	8.34	2.40	279	17.81	4.49	8.52	2.21
Male	533	19.20	6.14	8.06	2.51	540	17.53	5.06	8.15	2.43
Female	329	20.10	5.99	8.48	2.38	271	18.00	4.65	8.59	2.28
< 40	648	19.59	5.99	8.28	2.45	577	17.94	4.66	8.41	2.30
≥ 40	214	19.39	6.42	8.00	2.52	231	17.11	5.49	8.04	2.58

Table 8. Means and Standard Deviations for Each Version of the Three-Dimensional Space Test, by Subgroup.

Sub-group	n	Speeded Old Instr./Format (40 Items)		Nonspeeded Old Instr./Format (15 Items)		n	Speeded New Instr./Format (35 Items)		Nonspeeded New Instr./Format (15 Items)	
		Mean	SD	Mean	SD		Mean	SD	Mean	SD
All	867	14.82	6.16	7.03	2.94	814	14.41	5.76	7.33	2.99
White	235	16.72	6.51	7.94	3.10	231	16.08	6.05	8.33	3.06
Black	313	12.29	5.31	6.11	2.65	283	11.99	5.13	6.40	2.76
Hispanic	301	15.97	5.75	7.23	2.82	279	15.48	5.30	7.44	2.87
Male	533	15.14	6.41	7.16	3.06	540	14.68	5.77	7.52	3.05
Female	329	14.26	5.69	6.83	2.73	271	13.91	5.71	6.98	2.81
< 40	648	15.06	6.31	7.13	2.96	577	15.01	5.66	7.61	2.97
≥ 40	214	14.14	5.62	6.75	2.87	231	12.99	5.77	6.68	2.94

Table 9. Means and Standard Deviations for Each Version of the Vocabulary Test, by Subgroup.

Sub-group	n	Speeded Old Instr./Format (60 Items)		Nonspeeded Old Instr./Format (14 Items)		n	Speeded New Instr./Format (50 Items)		Nonspeeded New Instr./Format (14 Items)	
		Mean	SD	Mean	SD		Mean	SD	Mean	SD
All	867	15.28	7.23	7.18	3.29	814	15.36	6.73	7.08	3.19
White	235	18.89	7.66	8.46	3.36	231	18.75	6.85	8.45	3.26
Black	313	13.06	6.62	6.71	3.26	283	13.00	5.98	6.20	2.92
Hispanic	301	14.76	6.33	6.65	2.97	279	14.79	6.08	6.71	2.94
Male	533	15.25	7.46	7.32	3.42	540	15.33	6.79	7.05	3.26
Female	329	15.33	6.89	6.93	3.06	271	15.49	6.62	7.16	3.07
< 40	648	14.82	6.96	6.88	3.11	577	14.99	6.38	6.89	2.96
≥ 40	214	16.76	7.85	8.13	3.63	231	16.43	7.42	7.63	3.65

Table 10. Means and Standard Deviations for Each Version of the Arithmetic Reasoning Test, by Subgroup.

Sub-group	n	Speeded Old Instr./Format (25 Items)		Nonspeeded Old Instr./Format (10 Items)		n	Speeded New Instr./Format (24 Items)		Nonspeeded New Instr./Format (10 Items)	
		Mean	SD	Mean	SD		Mean	SD	Mean	SD
All	867	8.46	3.46	5.11	2.06	814	8.81	3.53	5.30	2.06
White	235	10.26	3.70	6.03	2.22	231	10.66	3.37	6.10	2.25
Black	313	7.21	3.17	4.60	1.87	283	7.42	3.22	4.82	1.86
Hispanic	301	8.39	2.93	4.96	1.89	279	8.62	3.21	5.05	1.85
Male	533	8.63	3.65	5.33	2.11	540	8.85	3.67	5.45	2.09
Female	329	8.19	3.11	4.76	1.94	271	8.71	3.25	4.97	1.95
< 40	648	8.40	3.41	4.99	1.97	577	8.89	3.42	5.24	2.04
≥ 40	214	8.68	3.59	5.47	2.28	231	8.68	3.79	5.44	2.11

Mean subgroup differences. In this investigation mean subgroup difference were operationally defined as the mean of the scores for examinees in the referent subgroup (e.g., Whites) minus the mean of the scores for examinees in the nonreferent subgroup (e.g., Blacks) divided by the standard deviation of the referent subgroup's scores. In this investigation White-Black, White-Hispanic, Male-Female, and Under 40 Years of Age-40 Years of Age and Over mean subgroup differences were examined. The mean subgroup differences for each version of each power test are presented in Tables 11 through 14. Each table contains these values for a particular power test. The first five rows of each table refer to the old instructions/format versions of the test and the second five rows refer to the new instructions/format versions. The first of these five rows contains the four mean subgroup differences for the speeded version of the test. For example, the first row of Table 11 shows that the White-Black, White-Hispanic, Male-Female, and Under 40 Years of Age-40 Years of Age and Over mean subgroup differences for the speeded old instructions/format Computation test are 0.62, 0.29, -0.15, and 0.03, respectively. The second row contains these values for the nonspeeded version of the test. For example, the second row of Table 11 shows that the White-Black, White-Hispanic, Male-Female, and Under 40 Years of Age-40 Years of Age and Over mean subgroup differences for the nonspeeded old instructions/format Computation test are 0.50, 0.19, -0.17, and 0.11, respectively.

The third of these five rows contains "corrected" estimates of what the mean subgroup differences for the nonspeeded version of the test would be if the nonspeeded version was as reliable as the speeded version. The "corrected" mean subgroup differences were calculated for the nonspeeded tests because they consisted of many fewer items than the speeded version of these tests. Tests with fewer items are less reliable and can therefore exhibit comparatively smaller mean subgroup differences merely as an artifact of this relatively lower unreliability. This corrected-for-unreliability mean subgroup difference was calculated in the following manner:

1. The parallel-forms reliabilities of the speeded versions of each power test were retrieved from U.S. Department of Labor (1985).

2. The internal consistency reliabilities of the nonspeeded versions of each power test were calculated. These values are presented in Appendix D.

3. The Spearman-Brown Prophecy formula was used to determine how much longer the nonspeeded version of each power test would need to be to obtain an internal consistency reliability equal to the parallel-forms reliability of its speeded version. This value will be referred to as K.

4. Referring to Gulliksen (1950), the subgroup means of each K length test were calculated. $Mean_K = Mean (K)$, where $Mean_K$ is the estimated subgroup mean on the nonspeeded test that is K times longer than the original, and Mean is the subgroup mean on the original nonspeeded test.

Table 11. Estimates of Mean Subgroup Differences for the Speeded and Nonspeeded Versions of Computation in the Old Instructions/Format and New Instructions/Format Conditions.

Computation Score		White - Black	White - Hispanic	Male - Female	< 40 - ≥ 40
Old Instr./ Format	Speeded	0.62	0.29	-0.15	0.03
	Nonspeeded	0.50	0.19	-0.17	0.11
	Corrected Nonspeeded	0.54	0.21	-0.18	0.12
	Difference	-0.12	-0.10	-0.02	0.08
	Corrected Difference	-0.08	-0.08	-0.03	0.09
New Instr./ Format	Speeded	0.73	0.36	-0.09	0.18
	Nonspeeded	0.64	0.19	-0.18	0.16
	Corrected Nonspeeded	0.70	0.21	-0.20	0.17
	Difference	-0.09	-0.17	-0.09	-0.02
	Corrected Difference	-0.03	-0.15	-0.11	-0.01

Note: The first five rows refer to the old instructions/format Computation tests; the second five rows refer to the new instructions/format Computation tests. The first of these five rows contains the mean subgroup differences for the speeded version of the test; the second row contains the mean subgroup differences for the nonspeeded version of the test. The third row contains "corrected" estimates of what the mean subgroup differences for the nonspeeded version of the test would be if the nonspeeded version was as reliable as the speeded version. The fourth row contains the difference between the uncorrected mean subgroup differences for the nonspeeded version of the test and the mean subgroup differences for the speeded version. In this row a negative number indicates that the relative performance of members of the nonreferent subgroup (in comparison to the performance of members of the referent subgroup) is more favorable on the nonspeeded version of the test than on the speeded version. Similarly, the fifth row compares the corrected mean subgroup differences for the nonspeeded version of the test to the mean subgroup differences for the speeded version of the test.

Table 12. Estimates of Mean Subgroup Differences for the Speeded and Nonspeeded Versions of Three-Dimensional Space in the Old Instructions/Format and New Instructions/Format Conditions.

Three-Dimensional Space Score		White - Black	White - Hispanic	Male - Female	< 40 - ≥ 40
Old Instr./ Format	Speeded	0.68	0.12	0.14	0.15
	Nonspeeded	0.59	0.23	0.11	0.13
	Corrected Nonspeeded	0.64	0.25	0.12	0.14
	Difference	-0.09	0.11	-0.03	-0.02
	Corrected Difference	-0.04	0.13	-0.02	-0.01
New Instr./ Format	Speeded	0.68	0.10	0.13	0.36
	Nonspeeded	0.63	0.29	0.18	0.31
	Corrected Nonspeeded	0.68	0.31	0.19	0.34
	Difference	-0.05	0.19	0.05	-0.05
	Corrected Difference	0.00	0.09	0.06	-0.02

Note: The first five rows refer to the old instructions/format Three-Dimensional Space tests; the second five rows refer to the new instructions/format Three-Dimensional Space tests. The first of these five rows contains the mean subgroup differences for the speeded version of the test; the second row contains the mean subgroup differences for the nonspeeded version of the test. The third row contains "corrected" estimates of what the mean subgroup differences for the nonspeeded version of the test would be if the nonspeeded version was as reliable as the speeded version. The fourth row contains the difference between the uncorrected mean subgroup differences for the nonspeeded version of the test and the mean subgroup differences for the speeded version. In this row a negative number indicates that the relative performance of members of the nonreferent subgroup (in comparison to the performance of members of the referent subgroup) is more favorable on the nonspeeded version of the test than on the speeded version. Similarly, the fifth row compares the corrected mean subgroup differences for the nonspeeded version of the test to the mean subgroup differences for the speeded version of the test.

Table 13. Estimates of Mean Subgroup Differences for the Speeded and Nonspeeded Versions of Vocabulary in the Old Instructions/Format and New Instructions/Format Conditions.

Vocabulary Score		White - Black	White - Hispanic	Male - Female	< 40 - ≥ 40
Old Instr./ Format	Speeded	0.76	0.54	-0.01	-0.28
	Nonspeeded	0.52	0.54	0.11	-0.40
	Corrected Nonspeeded	0.55	0.57	0.12	-0.43
	Difference	-0.24	0.00	0.12	-0.12
	Corrected Difference	-0.21	0.03	0.13	-0.15
New Instr./ Format	Speeded	0.84	0.58	-0.02	-0.23
	Nonspeeded	0.69	0.53	-0.03	-0.25
	Corrected Nonspeeded	0.74	0.57	-0.04	-0.27
	Difference	-0.15	-0.05	-0.01	-0.02
	Corrected Difference	-0.10	-0.01	-0.02	-0.04

Note: The first five rows refer to the old instructions/format Vocabulary tests; the second five rows refer to the new instructions/format Vocabulary tests. The first of these five rows contains the mean subgroup differences for the speeded version of the test; the second row contains the mean subgroup differences for the nonspeeded version of the test. The third row contains "corrected" estimates of what the mean subgroup differences for the nonspeeded version of the test would be if the nonspeeded version was as reliable as the speeded version. The fourth row contains the difference between the uncorrected mean subgroup differences for the nonspeeded version of the test and the mean subgroup differences for the speeded version. In this row a negative number indicates that the relative performance of members of the nonreferent subgroup (in comparison to the performance of members of the referent subgroup) is more favorable on the nonspeeded version of the test than on the speeded version. Similarly, the fifth row compares the corrected mean subgroup differences for the nonspeeded version of the test to the mean subgroup differences for the speeded version of the test.

26

Table 14. Estimates of Mean Subgroup Differences for the Speeded and Nonspeeded Versions of Arithmetic Reasoning in the Old Instructions/Format and New Instructions/Format Conditions.

Arithmetic Reasoning Score		White - Black	White - Hispanic	Male - Female	< 40 - ≥ 40
Old Instr./ Format	Speeded	0.82	0.51	0.12	-0.08
	Nonspeeded	0.64	0.48	0.27	-0.24
	Corrected Nonspeeded	0.75	0.56	0.32	-0.29
	Difference	-0.18	-0.03	0.15	-0.16
	Corrected Difference	-0.07	0.05	0.20	-0.21
New Instr./ Format	Speeded	0.96	0.61	0.04	0.06
	Nonspeeded	0.57	0.47	0.23	-0.10
	Corrected Nonspeeded	0.66	0.54	0.27	-0.11
	Difference	-0.39	-0.14	0.19	-0.16
	Corrected Difference	-0.30	-0.07	0.23	-0.17

Note: The first five rows refer to the old instructions/format Arithmetic Reasoning tests; the second five rows refer to the new instructions/format Arithmetic Reasoning tests. The first of these five rows contains the mean subgroup differences for the speeded version of the test; the second row contains the mean subgroup differences for the nonspeeded version of the test. The third row contains "corrected" estimates of what the mean subgroup differences for the nonspeeded version of the test would be if the nonspeeded version was as reliable as the speeded version. The fourth row contains the difference between the uncorrected mean subgroup differences for the nonspeeded version of the test and the mean subgroup differences for the speeded version. In this row a negative number indicates that the relative performance of members of the nonreferent subgroup (in comparison to the performance of members of the referent subgroup) is more favorable on the nonspeeded version of the test than on the speeded version. Similarly, the fifth row compares the corrected mean subgroup differences for the nonspeeded version of the test to the mean subgroup differences for the speeded version of the test.

5. Referring to Gulliksen (1950), the referent subgroup standard deviations of each K length test were calculated using the formula,

$$SD_K = SD (K + K (K - 1) r_{xx})^{1/2},$$

where SD_K is the estimated referent subgroup standard deviation on the nonspeeded test that is K times longer than the original, SD is the referent subgroup standard deviation on the original nonspeeded test, and r_{xx} is the internal consistency reliability of the original nonspeeded test.

6. The corrected nonspeeded mean subgroup differences were calculated by subtracting the nonreferent subgroup's $Mean_K$ from the referent subgroup's $Mean_K$ and dividing it by the referent subgroup's SD_K.

For example, the third row of Table 11 shows that the White-Black, White-Hispanic, Male-Female, and Under 40 Years of Age-40 Years of Age and Over corrected mean subgroup differences for the nonspeeded old instructions/format Computation test are 0.54, 0.21, -0.18, and 0.12, respectively.

The fourth row of each set of five rows in Tables 11 through 14 contains the difference between the _uncorrected_ mean subgroup differences for the nonspeeded version of the test and the mean subgroup differences for the speeded version. In this row a negative number indicates that the relative performance of members of the nonreferent subgroup (in comparison to the performance of members of the referent subgroup) is more favorable on the nonspeeded version of the test than on the speeded version. For example, the value -0.12 in the fourth row of Table 11 shows that the White-Black mean subgroup difference, that favors Whites on the speeded and nonspeeded versions of old instructions/format Computation, is 0.12 standard deviations less for the nonspeeded test as compared to the speeded test. Likewise the value 0.08 in the fourth row of Table 11 shows that the Under 40 Years of Age-40 Years of Age and Over mean subgroup difference, that favors examinees under 40 years of age on the speeded and nonspeeded versions of old instructions/format Computation, is 0.08 standard deviations greater for the nonspeeded test as compared to the speeded test. Similar to the fourth row, each fifth row contains the difference between the _corrected_ mean subgroup differences for the nonspeeded version of the test and the mean subgroup differences for the speeded version.

Tables 11 through 14 indicate that Blacks score lower than Whites on the speeded and nonspeeded old instructions/format and new instructions/format versions of the power tests; but that the White-Black mean subgroup difference is generally smaller for the nonspeeded versions, even after correcting for the relative unreliability of the nonspeeded tests. There are no substantial differences across instructions/format.

These tables indicate that Hispanics also score lower than Whites on the speeded and nonspeeded old instructions/format and new instructions/format versions of the power tests; however, the differences are smaller than the White-Black differences. The uncorrected and corrected mean subgroup differences for the nonspeeded versions of Computation are smaller than the mean subgroup differences for the speeded versions; however, this effect does not

exist for Vocabulary and Arithmetic Reasoning and reverses itself for Three-Dimensional Space. There are no substantial differences across instructions/format.

Tables 11 through 14 indicate that the sex and age differences are generally small. The Computation tests favor Females. The Three-Dimensional Space tests favor Males. The old instructions/format Vocabulary tests tend to favor Males, while there are no sex differences on the new instructions/format Vocabulary tests. The Arithmetic Reasoning tests favor Males and the effect is more pronounced for the nonspeeded versions. Generally, there are no substantial sex differences across instructions/format. The Computation tests favor examinees under 40 years of age. The Three-Dimensional Space tests favor examinees under 40 years of age; this effect was greater in the new instructions/format condition than in the old instructions/format condition. The Vocabulary tests favor examinees 40 years of age and over; this effect is more substantial for the nonspeeded version of the old instructions/format version than it is for the other versions. The same is true for the Arithmetic Reasoning tests; they generally favor examinees 40 years of age and over, and the effect is more substantial for the nonspeeded version of the old instructions/format test.

Additional descriptive comparisons. Additional information about subgroup differences for these tests can be derived from examining the figures in Appendix B and the tables in Appendix C. Appendix B contains figures that plot the percentage of examinees who attempted each possible number of items on each nonspeeded test within condition and across the subgroups. The structure of these figures is described in detail above, in the Speededness section of this report. In each figure, the extent to which the plotted line for subgroup A is below the plotted line for subgroup B is the extent to which the test is more speeded for subgroup A than for subgroup B. Figures B-1 through B-8 indicate that the nonspeeded Computation tests are slightly more speeded for Blacks than they are for Whites and Hispanics; otherwise the White, Black, and Hispanic plots are very similar. Figures B-9 through B-16 compare Males and Females. These figures indicate that the nonspeeded Computation tests are slightly more speeded for Males than they are for Females; otherwise the Male and Female plots are very similar for the remaining nonspeeded tests. Figures B-17 through B-24 compare examinees under 40 years of age and examinees 40 years of age and over. These figures indicate that all of the nonspeeded tests are slightly more speeded for the 40 Years of Age and Over subgroup than they are for the Under 40 Years of Age subgroup. It is important to note that all of the differences discussed in this paragraph are fairly small and are not always in the same direction as the test score differences discussed above.

Appendix C contains the means and standards deviations of five alternative scores for all the tests in the old instructions/format and new instructions/format conditions, by subgroup. The structure of these tables is described above, in the Speededness section of this report. The results in these tables are generally consistent with the results shown in Appendix B. The comparison of the Proportion Attempted and Accuracy scores, however, do provide additional information. Tables C-1, C-2, C-8, and C-9 indicate that on the speeded versions of the power tests, Blacks respond to fewer items than Whites and with less accuracy. On the nonspeeded versions, Blacks respond to a similar number of items as Whites, but still with less accuracy. This suggests that the nonspeeded versions have the potential to reduce White-Black mean subgroup differences that are due to differences in rate-of-responding, but not to reduce White-Black mean subgroup differences due to differences in accuracy. Tables C-1,

C-3, C-8, and C-10 indicate that Hispanics and Whites respond to a similar number of items on both speeded and nonspeeded versions of the power tests, but that Hispanics are generally less accurate. Tables C-4, C-5, C-11, and C-12 indicate that generally Males and Females respond to a similar number of items on both the speeded and nonspeeded versions of the power tests. However, Males are a bit more accurate. Tables C-6, C-7, C-13, and C-14 indicate that examinees under 40 years of age and examinees 40 years of age and over respond to a similar number of items on both the speeded and nonspeeded versions of the power tests. However, the 40 Years of Age and Over subgroup is more accurate on the Vocabulary and the Arithmetic Reasoning tests.

Analyses of variance: Nonspeeded tests. The effects of instructions/format and subgroup membership on nonspeeded test performance were tested for significance. The results of these tests are summarized in Table 15. Each analysis is a univariate 2 x 2 between groups Analysis of Variance. The levels of the first independent variable in all analyses are Old Instructions/Format and New Instructions/Format; the levels of the second independent variable in each analysis are the two subgroups referred to in the second column of the table (e.g., Whites and Blacks). Columns three through five of this table summarize results of analyses for which raw test score is the dependent variable; columns six through eight summarize results of analyses for which the number of items attempted is the dependent variable.

For clarification, one of the analyses summarized in Table 15 is described here. Columns three, four, and five of the first row summarize the results of one of the univariate 2 x 2 between groups Analysis of Variance. In this analysis the first independent variable is instructions/format; one level is old instructions/format and the other level is new instructions/format. The second independent variable is subgroup membership; one level is Whites and the other level is Blacks. The dependent variable is examinee raw score on the nonspeeded Computation test. The result in the third column shows that the main effect for instructions/format is nonsignificant. This result indicates that the mean raw score on the nonspeeded version of Computation for examinees in the old instructions/format condition is not significantly different from the mean raw score on the nonspeeded version of Computation for examinees in the new instructions/format condition. In contrast, the result in the fourth column shows that the main effect for subgroup is significant, in favor of Whites. This result indicates that the mean raw score on the nonspeeded versions of Computation for White examinees is significantly greater than the mean raw score on the nonspeeded versions of Computation for Black examinees. Finally, the result in the fifth column shows that the interaction between the two independent variables is nonsignificant. This indicates that the magnitude of the difference between the White and Black raw score means is not dependent on whether the examinees were in the old instructions/format condition or the new instructions/format condition.

30

The results of these analyses are consistent with the data presented thus far. That is, the main effects for instructions/format and subgroup are consistent with the results presented in earlier tables. The two significant interaction effects, however, need to be interpreted with caution. First, Table 15 represents 32 separate analyses; it is likely that some interaction effects will be significant by chance. Table 15 shows only two significant interactions, about the number that would be expected by chance. This result suggests that the mean subgroup differences for the nonspeeded tests are not affected by instructions/format. Second, each analysis is based on a relatively large sample size (e.g., the White-Black analyses are based on a total sample size of 1062); therefore, relatively small interaction effects are likely to be statistically significant. Finally, the number of items attempted variable is extremely negatively skewed. (Univariate normality is an assumption underlying this type of analysis, and the consequences of violating it are not clear.) Appendix E contains the source tables and sample sizes for each Analysis of Variance summarized in Table 15.

A speeded test: Name Comparison. Tables 16 and 17 show the means, standard deviations, and mean subgroup differences for the old instructions/format and the new instructions/format versions of the speeded Name Comparison test. Table 17 indicates that Blacks score lower than Whites on the old instructions/format and new instructions/format versions of this test, and that the effect is somewhat more substantial for the new instructions/format version. Similarly, Hispanics score lower than Whites in both conditions and the effect is more substantial in the new instructions/format condition; however, the White-Hispanic mean subgroup difference is substantially smaller than the White-Black mean subgroup difference. Females score a little higher than Males on both versions of Name Comparison. There are virtually no age differences for the old instructions/format version; however, examinees under 40 years of age score higher on the new instructions/format version of Name Comparison than examinees 40 years of age and over.

Analyses of variance: Speeded tests. The effects of speededness and instructions/format on speeded test performance were tested for significance. The results of these tests are summarized in Table 18. Table 18 is organized similar to Table 15, except that columns three through five refer to the independent variable accuracy instead of raw score. This is primarily because the speeded old instructions/format and new instructions/format tests do not contain the same number of items. The results of these analyses are consistent with the data presented thus far; however, the significance of an effect under these circumstances needs to be interpreted with caution. Table 18 represents 32 separate analyses; it is likely that some effects will be significant by chance. For example, it is relevant to note that Table 18 shows only one significant interaction, fewer than would by expected by chance. This result suggests that the mean subgroup differences for the speeded tests are not affected by instructions/format. Additionally, each analysis is based on a relatively large sample size (e.g., the White-Black analyses are based on a total sample size of 1,062); therefore, relatively small effects are likely to be statistically significant. Appendix E contains the source tables for each Analysis of Variance discussed above.

31

Table 15.   Summary of Significant Effects (i.e., p ≤ .05) on Nonspeeded Test Performance in Instructions/Format by Subgroup Analyses of Variance.

Comparison		Raw Score			Number of Items Attempted		
		Instr./ Format Effect	Sub-group Effect	Instr./ Format by Group Interact.	Instr./ Format Effect	Sub-group Effect	Instr./ Format by Group Interact.
Comp.	White-Black	ns	W > B	ns	O > N	W > B	SIG
	White-Hispanic	ns	W > H	ns	O > N	ns	ns
	Male-Female	ns	F > M	ns	O > N	F > M	ns
	< 40-≥ 40	ns	U > O	ns	O > N	U > O	ns
3-Dim. Space	White-Black	ns	W > B	ns	O > N	ns	ns
	White-Hispanic	ns	W > H	ns	O > N	H > W	SIG
	Male-Female	ns	M > F	ns	O > N	ns	ns
	< 40-≥ 40	ns	U > O	ns	O > N	U > O	ns
Vocab.	White-Black	ns	W > B	ns	ns	W > B	ns
	White-Hispanic	ns	W > H	ns	ns	ns	ns
	Male-Female	ns	ns	ns	ns	ns	ns
	< 40-≥ 40	ns	O > U	ns	ns	U > O	ns
Arith. Reas.	White-Black	ns	W > B	ns	ns	W > B	ns
	White-Hispanic	ns	W > H	ns	ns	ns	ns
	Male-Female	ns	M > F	ns	ns	ns	ns
	< 40-≥ 40	ns	O > U	ns	ns	U > O	ns

Note: W = Whites; B = Blacks; H = Hispanics; M = Males; F = Females; O = 40 Years of Age and Over; U = Under 40 Years of Age; ns indicates that the effect is nonsignificant; > indicates that the first mean is significantly greater than the second mean; SIG indicates that the interaction is significant. Each analysis is a univariate 2 x 2 between Analysis of Variance. The levels of the first independent variable in all analyses are Old Instructions/Format and New Instructions/Format; the levels of the second independent variable in each analysis are the two subgroups referred to in the second column of the table (e.g., Whites and Blacks). These analyses are based on the sample sizes shown in Tables 7 through 10.

Table 16. Means and Standard Deviations of Raw Scores for each Version of the Name Comparison test, by Subgroup.

Sub-group	n	Old Instr./Format (50 Items)		n	New Instr./Format (39 Items)	
		Mean	SD		Mean	SD
All	867	40.28	12.67	814	44.96	13.49
White	235	44.14	12.59	231	48.53	13.24
Black	313	36.86	11.66	283	40.85	13.19
Hispanic	301	40.97	12.79	279	46.03	12.87
Male	533	38.55	12.27	540	43.33	12.65
Female	329	43.13	12.80	271	48.31	14.42
< 40	648	40.97	12.34	577	46.59	13.25
≥ 40	214	38.32	13.41	231	41.04	13.15

Table 17. Estimates of Mean Subgroup Differences for the Name Comparison test in the Old Instructions/Format and New Instructions/Format Conditions.

Name Comparison Score	White - Black	White - Hispanic	Male - Female	< 40 - ≥ 40
Old Instructions/Format	0.62	0.29	-0.15	0.03
New Instructions/Format	0.73	0.36	-0.09	0.18

Table 18. Summary of Significant Effects (i.e., p ≤ .05) on Speeded Test Performance in Instructions/Format by Group Analyses of Variance.

Comparison		Accuracy			Number of Items Attempted		
		Instr./ Format Effect	Sub-group Effect	Instr./ Format by Group Interact.	Instr./ Format Effect	Sub-group Effect	Instr./ Format by Group Interact.
Name Compar.	White-Black	ns	W > B	ns	N > O	W > B	ns
	White-Hispanic	ns	W > H	ns	N > O	ns	ns
	Male-Female	ns	F > M	ns	N > O	F > M	ns
	< 40-≥ 40	ns	ns	ns	N > O	U > O	SIG
Comp.	White-Black	ns	W > B	ns	O > N	W > B	ns
	White-Hispanic	ns	W > H	ns	O > N	W > H	ns
	Male-Female	ns	ns	ns	O > N	F > M	ns
	< 40-≥ 40	ns	ns	ns	O > N	U > O	ns
3-Dim. Space	White-Black	O > N	W > B	ns	ns	W > B	ns
	White-Hispanic	O > N	W > H	ns	N > O	H > W	ns
	Male-Female	O > N	M > F	ns	N > O	F > M	ns
	< 40-≥ 40	O > N	ns	ns	N > O	U > O	ns
Vocab.	White-Black	O > N	W > B	ns	N > O	W > B	ns
	White-Hispanic	O > N	W > H	ns	N > O	ns	ns
	Male-Female	O > N	ns	ns	N > O	F > M	ns
	< 40-≥ 40	O > N	O > U	ns	N > O	U > O	ns
Arith. Reas.	White-Black	ns	W > B	ns	N > O	W > B	ns
	White-Hispanic	ns	W > H	ns	N > O	W > H	ns
	Male-Female	ns	M > F	ns	N > O	ns	ns
	< 40-≥ 40	ns	O > U	ns	N > O	U > O	ns

Table 18. Continued.

Note: W = Whites; B = Blacks; H = Hispanics; M = Males; F = Females; O = 40 Years of Age and Over; U = Under 40 Years of Age; ns indicates that the effect is nonsignificant; > indicates that the first mean is significantly greater than the second mean; SIG indicates that the interaction is significant. Each analysis is a univariate 2 x 2 between Analysis of Variance. The levels of the first independent variable in all analyses are Old Instructions/Format and New Instructions/Format; the levels of the second independent variable in each analysis are the two subgroups referred to in the second column of the table (e.g., Whites & Blacks). Accuracy is a computed score (Number of Items Correct / Number of Items Attempted). These analyses are based on the sample sizes shown Appendix C.

Summary. The results presented above consistently indicate that the White-Black mean subgroup difference is substantial for all of the power tests. However, the nonspeeded versions of the Computation, Vocabulary, and Arithmetic Reasoning GATB power tests do show smaller White-Black mean subgroup differences than the speeded versions of these tests. This effect is somewhat diminished, but not eliminated, by correcting the nonspeeded mean subgroup differences for their relative unreliability. The results also consistently indicate a moderate White-Hispanic mean subgroup difference for all of the power tests. The nonspeeded power tests do not show a general reduction in the White-Hispanic mean subgroup difference compared to their speeded counterparts. There are no substantial sex effects. However, the nonspeeded versions of Arithmetic Reasoning do favor Males. There are small age differences favoring the Under 40 Years of Age subgroup on Computation and Three-Dimensional Space. However, somewhat larger differences favoring the 40 Years of Age and Over subgroup were observed for Vocabulary and Arithmetic Reasoning. (These differences are somewhat more substantial for the nonspeeded versions.)

Construct Equivalence.

Do GATB power tests given under current (i.e., speeded) and power (i.e., nonspeeded) conditions measure the same constructs? One way to address this question is to examine the correlations between the speeded and the nonspeeded versions of the tests within each instructions/format condition. Table 19 contains these correlations. The third column of values in this table shows the correlations between the speeded version of each power test (based on items from GATB operational Form A) and the nonspeeded version of each power test (based on items from GATB operational Form D). These correlations can be compared to the correlations in the first column of Table 19; they are correlations between the operational Form A and D power tests administered one week apart (U.S. Department of Labor, 1985). If the correlations in the third column were equal to the correlations in the first column, the inference would be strong that the nonspeeded versions of the power tests measure the same constructs as the speeded (operational) versions to the same extent that the parallel versions of the speeded tests measure the same constructs. However, the correlations in third column are somewhat smaller than the correlations in the first column (between 0.00 to 0.20 smaller, 0.13 smaller on average). These results do not allow a strong inference that the nonspeeded versions of the power tests measure the same constructs as the speeded versions to the same extent that the parallel versions of the speeded tests measure the same construct.

Taus and true score correlations. One potential reason for the speeded and nonspeeded versions of the power tests not correlating with each other to the same extent as the parallel versions of the speeded tests is that the nonspeeded test scores are based on fewer items and are therefore less reliable. If the correlations between the speeded and nonspeeded versions of the power tests were corrected for differences in reliability, would they be 1.00? Cronbach and Warrington (1951) proposed *tau* to address this issue. "Essentially, *tau* allows determination of the proportion of the reliable time limit score variance reflecting common factors in the no-time-limit parallel test" (Rindler, 1979, p. 266). As presented by Cronbach

36

Table 19. *Tau* and True Score Correlations Between Speeded and Nonspeeded Versions of the GATB Power Tests in the Old Instructions/Format and New Instructions/Format Conditions.

	Test	Speeded Reliability	Non-speeded Reliability	r Speeded/Non-Speeded	*Tau* 1	*Tau* 2	True Score r1	True Score r2
Old Instr./ Format	Computation	0.84	0.71	0.72	0.87	0.73	0.93	0.86
	Three-Dimensional Space	0.81	0.69	0.68	0.83	0.70	0.91	0.84
	Vocabulary	0.88	0.78	0.74	0.80	0.71	0.89	0.84
	Arithmetic Reasoning	0.77	0.56	0.63	0.92	0.67	0.96	0.82
New Instr./ Format	Computation	0.84	0.71	0.75	0.94	0.80	0.97	0.89
	Three-Dimensional Space	0.81	0.70	0.67	0.79	0.68	0.89	0.83
	Vocabulary	0.88	0.76	0.76	0.86	0.75	0.93	0.86
	Arithmetic Reasoning	0.77	0.57	0.57	0.74	0.55	0.86	0.74

Note: The speeded reliabilities are parallel forms reliabilities based on averages of two samples of over 500; Forms A and D administered one week apart (U.S. Department of Labor, 1985).
The nonspeeded reliabilities are internal consistency reliabilities were computed from present samples of 867 and 814.
Tau 1 equals the between forms correlation squared divided by the parallel forms reliability of the speeded versions of the tests times the internal consistency reliabilities of the nonspeeded versions of the tests.
Tau 2 equals the between forms correlation squared divided by the parallel forms reliability of the speeded versions of the tests squared.
True score correlations are the square roots of the respective *taus*.

and Warrington, the calculation of *tau* requires scores on two parallel forms of the speeded version of a test and two parallel forms of the nonspeeded version of the test on which to base estimates of parallel test reliabilities for the speeded and nonspeeded versions of the test. *Tau* equals

$$(r_{sn})^2 / ((r_{ss})(r_{nn})),$$

where r_{sn} is the correlation between the speeded and nonspeeded versions of a test, r_{ss} is the parallel forms reliability of the speeded version of the test, and r_{nn} is the parallel forms reliability of the nonspeeded version of the test. This investigation can only approximate *tau* because there is no parallel forms reliability estimate for the nonspeeded versions of the tests; an estimate of internal consistency reliability and the parallel forms reliabilities of the speeded versions of the tests are used as substitutes. The square root of this approximation of *tau* is an estimate of the true score correlation between the speeded and nonspeeded versions of a test.

Table 19 shows estimates of *tau* and the true score correlations for the power tests, by condition. The fourth column of values shows the estimates of *tau* (i.e., *tau* 1) that use the operational Forms A and D parallel forms reliabilities as estimates of the reliabilities of the speeded tests (the first column of values) and the internal consistency reliabilities as estimates of the reliabilities of the nonspeeded tests (the second column of values). The sixth column of values shows the estimated true score correlations (i.e., true score correlation 1) based on these *taus*. Because it is likely that the internal consistencies are underestimates of the parallel forms reliabilities of the nonspeeded tests, it is likely that *tau* 1 and true score correlation 1 slightly overestimate these parameters. The fifth row shows the estimates of *tau* (i.e., *tau* 2) that use the operational Forms A and D parallel forms reliabilities as estimates of the reliabilities of the speeded and the nonspeeded tests (the first column of values). The seventh column of values shows the estimated true score correlations (i.e., true score correlation 2) based on these *taus*. Because it is likely that the operational Forms A and D parallel forms reliabilities are substantial overestimates of the parallel forms reliabilities of the nonspeeded tests, it is likely that *tau* 2 and true score correlation 2 substantially underestimate these parameters.

The results in the last two columns of Table 19 show that the true score correlations are not 1.00. This means that even after correcting for unreliability in the speeded and nonspeeded tests, the inference cannot be made that the nonspeeded versions of the power tests measure the same constructs as the speeded (operational) versions of the power tests. However, the constructs being measured by the speeded and nonspeeded tests are similar; all but one of the estimated true score correlations are greater than 0.80.

Models. The constructs measured by the speeded and nonspeeded versions of the power tests can be compared in the context of maximum likelihood confirmatory factor analysis. In accordance with past empirical results and theory, a model was hypothesized; it is referred to here as Model 1. Model 1 attempts to account for the relationships among the speeded and nonspeeded tests in the old instruction/format condition by hypothesizing that,

with two exceptions, the speeded and the nonspeeded versions of each test are parallel (i.e., they measure the same constructs to the same extent). The first exception is that because the nonspeeded versions of the tests contain many fewer items than the speeded versions, their variances and covariances cannot be equal; therefore, the correlation matrix is analyzed instead of the covariance matrix. The second exception is that a Speed factor accounts for some of the common variance in the speeded tests. Model 1 makes the following hypotheses:

- All of the tests load on a General factor (i.e., a factor that represents g), and the speeded and nonspeeded versions of each test load equally on this factor. This equality constraint represents the hypothesis that the speeded and nonspeeded versions of each test measure g to the same extent.

- The speeded versions of each test load on a Speed factor.

- The speeded and nonspeeded versions of Computation and Arithmetic Reasoning load on a Math factor, and the speeded and nonspeeded versions of each of these tests load equally on this factor. This equality constraint represents the hypothesis that the speeded and nonspeeded versions of each of these tests measure Math to the same extent.

- The speeded and nonspeeded versions of Three-Dimensional Space load on a Three-Dimensional Space factor, and the speeded and nonspeeded versions of this test load equally on this factor. This equality constraint represents the hypothesis that the speeded and nonspeeded versions of this test measure Three-Dimensional Space to the same extent.

- The speeded and nonspeeded versions of Vocabulary load on a Vocabulary factor, and the speeded and nonspeeded versions of this test load equally on this factor. This equality constraint represents the hypothesis that the speeded and nonspeeded versions of this test measure Vocabulary to the same extent.

- The General and Speed factors are hypothesized to be orthogonal to each other and the other factors, and the specific factors (i.e., Math, Three-Dimensional Space, and Vocabulary) are hypothesized to be related to each other.

The parameter estimates for Model 1 are presented in Tables 20 through 22. Table 20 shows the estimated loadings of each test on each factor. Table 21 shows the estimated correlations among the factors. Table 22 shows the estimated uniquenesses.

Simplified versions of Model 1 were tested to assess the extent to which some of the parameters in Model 1 are necessary. Model 2 is the same as Model 1 except that it does not hypothesize a Speed factor. Model 3 is the same as Model 1 except that it does not hypothesize a Speed factor, and it hypothesizes that the specific factors are orthogonal. Model 4 is the same as Model 1 except that it does not hypothesize a General factor. Finally, Model 5 is the same as Model 1 except that it does not hypothesize a General or a Speed factor. Fit statistics for each of these Models are presented in Table 23. The fit statistics in

Table 20. Pattern Matrix Parameter Estimates for Model 1.

Test	Factor				
	General	Speed	Math	3 -D	Vocab.
Speeded Computation	0.79	0.34	0.28	0	0
Speeded 3-D Space	0.29	0.23	0	0.77	0
Speeded Vocab	0.42	0.27	0	0	0.75
Speeded Arith. Reas.	0.52	0.33	0.61	0	0
Nonspeeded Computation	0.79	0	0.28	0	0
Nonspeeded 3-D Space	0.29	0	0	0.77	0
Nonspeeded Vocab	0.42	0	0	0	0
Nonspeeded Arith. Reas.	0.52	0	0.61	0	0.75

Note: The speeded and nonspeeded versions of each test were constrained to load equally on the General factor and their particular specific factor (i.e., Math, Three-Dimensional Space, or Vocabulary.). 0 indicates that the loading was fixed to zero.

Table 21. Factor Correlation Matrix Parameter Estimates for Model 1.

Factors	Factors				
	General	Speed	Math	3-D	Vocab.
General	1.00				
Speed	0	1.00			
Math	0	0	1.00		
3-D	0	0	0.55	1.00	
Vocab.	0	0	0.73	0.51	1.00

Note: 0 indicates that the loading was fixed to zero; 1.00 indicates that the loading was fixed to one.

Table 22. Uniqueness Parameter Estimates for Model 1.

Test	Uniqueness
Speeded Computation	0.15
Speeded Three-Dimensional Space	0.29
Speeded Vocabulary	0.18
Speeded Arithmetic Reasoning	0.21
Nonspeeded Computation	0.32
Nonspeeded Three- Dimensional Space	0.29
Nonspeeded Vocabulary	0.26
Nonspeeded Arithmetic Reasoning	0.41

Table 23. Model Fits (Old Instructions/Format, Total Sample, N = 867).

Model	d.f.	χ^2	p	RMSEA	RMR	GFI
Model 1	13	25.20	0.02	0.03	0.03	0.99
Model 2	17	140.67	0.00	0.09	0.05	0.96
Model 3	20	143.41	0.00	0.08	0.05	0.96
Model 4	17	144.99	0.00	0.09	0.04	0.96
Model 5	21	224.87	0.00	0.11	0.06	0.94

Note: Model 1 is described in the text and by Tables 21 through 22;
Model 2 is the same as Model 1 minus the Speed factor;
Model 3 is the same as Model 1 minus the Speed factor, and the Specific factors are hypothesized to be orthogonal;
Model 4 is the same as Model 1 minus the General factor;
Model 5 is the same as Model 1 minus the General factor and the Speed factor;
RMSEA = Root Mean Square Error of Approximation;
RMR = Root Mean Square Residual;
GFI = Goodness of Fit Index.

Table 23 indicate that Model 1 fits reasonably well and substantially better than Models 2, 3, 4, or 5. Examination of Table 22 reveals a result that is consistent with the observation that the speeded tests are more reliable than their nonspeeded versions; that is, the uniquenesses of the nonspeeded tests are larger than the uniquenesses of the speeded tests. These results support the inference that with the exception of different variances and reliabilities and a Speed factor accounting for some of the variance in the speeded tests, the speeded and nonspeeded versions of the GATB power tests measure the same constructs to the same extent. Appendix F contains the correlation matrix on which these analyses are based.

An alternative model, using the Generalized Least Squares method, was tested by Lauress Wise (personal communication, December 13, 1993). This model attempts to account for the relationships among all of the speeded and nonspeeded tests in the old instruction/format condition, including Name Comparison. It hypothesizes five factors: (1) a Speed factor that Name Comparison and the speeded versions of Computation, Three-Dimensional Space, Vocabulary, and Arithmetic Reasoning load on; (2) a Computation factor that the speeded and nonspeeded versions of Computation load on; (3) a Three-Dimensional Space factor that the speeded and nonspeeded versions of Three-Dimensional Space load on; (4) a Vocabulary factor that the speeded and nonspeeded versions of Vocabulary load on; and (5) an Arithmetic Reasoning factor that the speeded and nonspeeded versions of Arithmetic Reasoning load on. All of the factors are hypothesized to be related and none of the test factor loadings are hypothesized to be equal. The model fit reasonably well (N = 867, d.f. = 13, χ^2 = 46.23, p = 0.00, Root Mean Square Residual (RMR) = 0.02, Goodness of Fit (GFI) = 0.99). The parameter estimates resulting from this model are consistent with the inference that the speeded and nonspeeded versions of the GATB power tests measure the same constructs, except that a Speed factor accounts for some of the variance in the speeded versions of these tests.

Do the inferences made based on the confirmatory factor analyses of the correlations among the speeded and nonspeeded versions of the old instructions/format tests extend to the new instructions/format tests? A model was tested that hypothesizes that the population correlations among the eight speeded and nonspeeded power tests in the new instructions/format conditions are identical to the population correlations among the eight speeded and nonspeeded power tests in the old instructions/format condition. The model fit very well (N = 867 in the old instructions/format condition, N = 814 in the new instructions/format condition, d.f. = 36, χ^2 = 33.13, p = 0.61, Root Mean Square Error of Approximation (RMSEA) = 0.00, Root Mean Square Residual (RMR) = 0.01, Goodness of Fit (GFI) = 0.99). This result supports the inference that the population correlations among the speeded and nonspeeded power tests in the old instructions/format condition are the same as the correlations among the speeded and nonspeeded tests in the new instructions/format condition. Appendix F contains the correlation matrices on which these analyses are based.

Do the inferences made based on the confirmatory factor analyses of the correlations among the speeded and nonspeeded versions of the tests for the total group extend to the relevant subgroups? That is, are the relationships among the tests the same across subgroups? To address this question models were tested within each instructions/format condition that hypothesized that: (a) the population correlations among the speeded and nonspeeded tests for Whites are identical to the population correlations among these tests for Blacks; (b) the population correlations among the speeded and nonspeeded tests for Whites are identical to the population correlations among these tests for Hispanics; (c) the population correlations among the speeded and nonspeeded tests for Males are identical to the population correlations among these tests for Females; and (d) the population correlations among the speeded and nonspeeded tests for examinees under 40 years of age are identical to the population correlations among these tests for examinees 40 years of age and over. The fit statistics for these models are presented in Table 24. Appendix F contains the correlations matrices on which these analyses are based.

Some of the chi-square values in Table 24 reject the models at the $p = 0.01$ level of significance. This result suggests that some the relationships among the speeded and nonspeeded versions of the GATB power tests are not the same across subgroups; however, it is important to note that a number of researchers have criticized the chi-square statistic for assessing the fit of a model (e.g., Cudeck & Henly, 1991). One of the criticisms has been that the chi-square statistic is extremely sensitive to sample size. That is, when the sample size is large almost no model will fit and when the sample size is small almost any model will fit. The sample sizes in this situation are not small. The RMSEA fit statistic addresses this problem explicitly (Steiger & Lind, 1980). This statistic is a measure of the misfit per degree of freedom of the model. One advantage that RMSEA has over the chi-square is that if the model fits perfectly in the population, RMSEA will equal zero regardless of the sample size. Joreskog and Sorbom (1993) suggest that a value of 0.05 or less indicates a close fit of the model. The RMSEA values in Table 24 are all 0.04 or smaller. Two other commonly used fit statistics (i.e., RMR and GFI) are also included in the table. The RMR fit statistic is a measure of the average of the fitted residuals (Joreskog & Sorbom, 1993). The fitted residuals are the differences between the elements of the sample correlation matrix and the elements of the matrix estimated using the parameters of the model. GFI is another statistic that measures the fit of the model. If the sample matrix and the estimated matrix that is based on the model are equal, then GFI is 1.00 (Joreskog & Sorbom, 1993).

The chi-square statistic, tested at the $p = 0.01$ level of significance, rejects some of the models referred to in Table 24. However, given sufficient sample size models that are imperfect in the smallest way will be rejected by this test. The RMSEA acknowledges this problem and allows us to view the fit of a model on a continuum that is less affected by sample size, and the RMSEA value of 0.05 has been chosen as a reasonable cutoff. Based on these considerations and the values of these statistics shown in Table 24, it is concluded that the results provide reasonable support for the inference that the relationships among the speeded and nonspeeded versions of the GATB power tests are the same across subgroups.

Table 24. Models Hypothesizing Identical Correlations Among the Eight Power Tests Across Subgroups in the Old Instructions/Format and the New Instructions/Format Conditions.

Model		N	d.f.	χ^2	P	RMSEA	RMR	GFI
Old Instr./ Format	White/Black	235/313	36	47.46	0.10	0.03	0.03	0.99
	White/Hispanic	235/301	36	70.35	0.00	0.04	0.05	0.98
	Male/Female	533/329	36	77.80	0.00	0.04	0.04	0.97
	Under/Over 40	648/214	36	45.23	0.14	0.02	0.06	0.96
New Instr./ Format	White/Black	231/283	36	43.46	0.18	0.02	0.03	0.98
	White/Hispanic	231/279	36	54.34	0.03	0.03	0.05	0.98
	Male/Female	540/271	36	59.28	0.01	0.03	0.05	0.97
	Under/Over 40	577/231	36	62.06	0.00	0.03	0.06	0.95

Note: Each model hypothesizes that each population correlation in the first subgroup is identical to its corresponding population correlation in the other group.
RMSEA = Root Mean Square Error of Approximation;
RMR = Root Mean Square Residual;
GFI = Goodness of Fit Index.

Relationships with a speed test. Model 1, discussed above, hypothesizes that a Speed factor accounts for some of the variance in the speeded power tests. Further evidence regarding the relative influence of a Speed factor on the power tests can be obtained by observing the correlations between Name Comparison (a deliberately speeded test) and the speeded and nonspeeded versions of these tests. These correlations are presented in Table 25 for the old instructions/format condition and in Table 26 for the new instructions/format condition. The second column in each table shows the uncorrected correlations, and the fourth column shows the correlations corrected for the unreliability of the power tests. Tables 25 and 26 indicate that even after correcting for unreliability, the correlations between the speeded versions of the power tests and Name Comparison are higher than the correlations between the nonspeeded versions and Name Comparison.

Summary. The goal of the preceding analyses was to assess the extent to which the nonspeeded versions of the GATB power tests measure the same constructs as the speeded (i.e., operational) versions. Comparing raw score correlations suggests that the speeded and nonspeeded versions of the tests do not measure the same constructs. Comparing the *taus*, the estimated true score correlations, and the reliabilities suggests that the speeded and nonspeeded versions measure nearly but not precisely the same constructs. Examination of these results and the results from the confirmatory factor analyses indicates that the nonspeeded versions of the power tests measure the same constructs as the speeded versions with one notable exception; a Speed factor accounts for some of the variance in the speeded tests. A comparison of the correlations between the speeded power tests and Name Comparison and the correlations between the nonspeeded power tests and Name Comparison supports the inference that the variance unique to the speeded power tests is indeed a Speed factor, and not a method factor. Finally, confirmatory analyses provide reasonable support for the inference that these results hold across the two instructions/format conditions and across the subgroups (i.e., Whites, Blacks, Hispanics, Males, Females, Under 40 Years of Age, and 40 Years of Age and Over).

Table 25. Correlations Between Name Comparison and the Speeded and Nonspeeded Power Tests in the Old Instructions/Format Condition: Uncorrected and Corrected for Unreliability in the Power Tests.

Power Test		Name Compar./ Power Test r	Power Test Reliability	Corrected Name Compar./ Power Test r
Speeded	Computation	0.64	0.84	0.70
	3-D Space	0.42	0.81	0.47
	Vocabulary	0.55	0.88	0.59
	Arith. Reas.	0.55	0.77	0.63
Non-Speeded	Computation	0.50	0.71	0.59
	3-D Space	0.34	0.69	0.40
	Vocabulary	0.40	0.78	0.46
	Arith. Reas.	0.35	0.56	0.47

Note: The speeded reliabilities are parallel forms reliabilities based on averages of two samples of over 500; Forms A and D administered one week apart.
The nonspeeded reliabilities are internal consistency reliabilities computed from the present sample of 867.

Table 26. Correlations Between Name Comparison and the Speeded and Nonspeeded Power Tests in the New Instructions/Format Condition: Uncorrected and Corrected for Unreliability in the Power Tests.

Power Test		Name Compar./ Power Test r	Power Test Reliability	Corrected Name Compar./ Power Test r
Speeded	Computation	0.64	0.84	0.70
	3-D Space	0.47	0.81	0.53
	Vocabulary	0.56	0.88	0.60
	Arith. Reas.	0.57	0.77	0.65
Non-Speeded	Computation	0.53	0.71	0.63
	3-D Space	0.31	0.70	0.37
	Vocabulary	0.44	0.76	0.50
	Arith. Reas.	0.36	0.57	0.48

Note: The speeded reliabilities are parallel forms reliabilities based on averages of two samples of over 500; Forms A and D administered one week apart.
The nonspeeded reliabilities are internal consistency reliabilities computed from the present sample of 814.

Discussion

Feasibility of Nonspeeded Tests

One objective of this investigation was to evaluate the feasibility of constructing nonspeeded versions of the GATB power tests that can be administered within approximately current operational time limits. The method section of this report describes an attempt to construct nonspeeded versions of the GATB power tests that were as parallel as possible to their speeded versions within the constraint of no substantial increases in test administration time. The results indicate that it is possible to create nonspeeded versions of the Three-Dimensional Space, Vocabulary, and Arithmetic Reasoning tests without making substantial changes to the operational time limits. The time limits for the nonspeeded versions of these tests, constructed for this investigation, were 6, 6, and 11 minutes, respectively. The nonspeeded Arithmetic Reasoning test was the only one whose time limit was extended beyond the operational time limit (i.e., from 7 to 11 minutes). These tests contained 15, 14, and 10 items, respectively. The results show that these tests do satisfy the criterion for a nonspeeded test that requires that examinees have sufficient time to attempt all of the items. The results are less supportive of a nonspeeded version of the Computation test. During the construction of the "nonspeeded" version of the Computation test the judgement was made that 10 items were required to adequately capture the content domain prescribed for that test. The problem is that the nonspeeded version of the Computation test (i.e., 10 items with a 6 minute time limit) did not satisfy the criterion that requires that examinees have sufficient time to attempt all of the items. Given the requirement that this test include at least 10 items, achieving a completely nonspeeded version of the Computation test would require increasing the time limit. However, this investigation's "nonspeeded" versions of the Computation test were considerably less speeded than the speeded versions (i.e., GATB Forms A, B, C, and D).

Speededness and Subgroup Differences

This investigation examined the extent to which speededness influences the magnitude of subgroup differences in scores on the GATB power tests. The results indicate that the White-Black mean subgroup differences are substantial for both the speeded and nonspeeded versions of the GATB power tests. The nonspeeded versions of the Computation, Vocabulary, and Arithmetic Reasoning tests do show somewhat smaller White-Black mean subgroup differences than the speeded versions of these tests. This effect is diminished, but not eliminated, by correcting the estimates of mean subgroup differences for the relative unreliability of the nonspeeded versions of the tests. The results indicate moderate White-Hispanic mean subgroup differences for the speeded and nonspeeded versions of the GATB power tests. The nonspeeded power tests do not show a general reduction in White-Hispanic mean subgroup differences compared to their speeded counterparts. The speeded and nonspeeded versions of the GATB power tests do not show substantial sex effects; however, the nonspeeded versions of Arithmetic Reasoning do favor Males. There are small age differences favoring the Under 40 Years of Age subgroup on Computation and Three-Dimensional Space; however, somewhat larger differences favoring the 40 Years of Age and Over subgroup were observed for Vocabulary and Arithmetic Reasoning. (This effect was somewhat larger for the nonspeeded versions of Vocabulary and Arithmetic Reasoning.) Taken as a whole the results of this investigation suggest that moving to nonspeeded versions

of the GATB power tests would not substantially increase mean subgroup differences for any of the subgroups studied in this investigation and might result in small reductions in the White-Black mean subgroup differences for three of the power tests.

Construct Equivalence

A major objective of this investigation was to assess the extent to which nonspeeded versions of the GATB power tests measure the same constructs as current (i.e., speeded) versions of these tests. The results of this investigation support the hypothesis that the speeded and nonspeeded versions of the GATB power tests measure the same constructs as the speeded versions with one conspicuous exception; a Speed factor accounts for some of the variance in the speeded tests. Peterson (1993) indicated that the NRC committee on the GATB (Hartigan & Wigdor, 1989) was concerned that, "the meaning of constructs measured by the speeded power tests may be different from the meaning conventionally attached to those constructs" (p. 1). These results indicate that scores on the current versions of the GATB power tests are influenced by a speeded component.

Relationships Among Tests Across Subgroups

An issue related to the constructs measured by the speeded and nonspeeded versions of the GATB power tests is the question of whether the relationships among these tests are the same across relevant subgroup comparisons. The results of this investigation provide reasonable support for the inference that the relationships among the speeded and nonspeeded versions of the GATB power tests are substantially the same across four important subgroup comparisons: (1) White-Black; (2) White-Hispanic; (3) Male-Female; and (4) Under 40 Years of Age-40 Years of Age and Over.

Peterson (1993) indicated that the NRC committee on the GATB (Hartigan & Wigdor, 1989) was concerned that the speeded component in the current versions of the GATB power tests might cause differential validity across subgroups. The results of this investigation cannot address issues related to differential criterion-related validity. However, in terms of construct validity, the results of this investigation do allow the inference that the relationships among the speeded and nonspeeded versions of the GATB power tests are substantially the same across the subgroups examined in this investigation.

Adverse Reactions Due to Speededness

The NRC committee on the GATB (Hartigan & Wigdor, 1989) was concerned that "the severe time limits of the GATB subtests might produce an adverse psychological reaction in examinees as they progress through the examination and might thereby reduce the construct validity of the subtests" (p. 106). They were specifically concerned that the people most likely to experience this effect would be individuals "least experienced with standardized tests, a group in which minority examinees will be overrepresented" (p.106). Peterson (1993) reviewed the relevant literature and concluded that there is very little existing research that addresses the question of whether test speededness results in adverse psychological reactions that reduce construct validity, or whether this effect varies across subgroups. The research that Peterson did discuss offers, at best, mixed support for the existence of such an effect.

The current research cannot address the issue of adverse psychological reactions to speeded tests directly; it can only speak to potential differential speeded test construct validity across subgroups. As mentioned above the results of this investigation suggest that the nonspeeded versions of three of the GATB power tests do show smaller White-Black mean subgroup differences than their speeded versions. This result suggests that for these tests, speededness may adversely affect the performance of Black examinees compared to White examinees. Such effects are small to nonexistent for the remaining subgroup comparisons. As mentioned above the results of this investigation also allow for the comparison of the relationships among the speeded and nonspeeded versions of the GATB power tests across subgroups. The results of those analyses generally support the hypothesis that the relationships among the speeded versions of the tests are the same across the subgroups examined in this investigation. In summary, with the exception of smaller White-Black mean group differences on three of the GATB power tests, the results suggest that speededness does not precipitate differential construct validity across subgroups.

Instructions/Format

One objective of this investigation was the assessment of the effects of changes to instructions, item formats, and answer sheet formats anticipated for future forms of the GATB power tests. The basis for this concern was the hypothesis that changes (i.e., improvements) in the instructions, item formats, and/or answer sheet formats might affect the speededness of the power tests. This concern lead to the inclusion, in this investigation, of an old (i.e., operational) instructions/format condition and a new (i.e., changes anticipated for new operational forms) instructions/format condition.

The results of this investigation show no substantial differences in test performance between the old instructions/format and new instructions/format conditions. While Analyses of Variance show some significant instructions/format effects in terms of number of items attempted, Figures 1 through 4 provide convincing evidence that differences in the speededness of the two nonspeeded versions of the GATB power tests used in this investigation are virtually nonexistent. Also, Tables 5 and 6 show that there are no substantial across condition differences in the number of items attempted on the speeded versions, thus showing that instructions/format has little effect on the speededness of the speeded tests. The results clearly indicate that instructions/format has no effect on nonspeeded GATB power test scores and little or no effect on speeded GATB power test scores. This investigation also provides evidence that mean subgroup differences for the speeded and nonspeeded GATB power tests are not affected by instructions/format. Finally, the results indicate that the relationships among the speeded and nonspeeded tests are the same in both instructions/format conditions.

Recommendations

One motivation for this investigation was that the NRC committee on the GATB criticized the GATB for the use of relatively stringent time limits for tests designed to measure constructs that are conventionally measured by tests with generous time limits. The following are recommendations regarding alterations to the GATB's power tests that attempt to address this criticism. These recommendations are based on (a) the Review of Issues

50

(Peterson, 1993), (b) the empirical research discussed in this report, and (c) comments on this research and recommendations made on the basis of this research provided by scientific/technical advisors during a conference held on December 13, 1993 to examine issues associated with the speededness of the GATB power tests.

Computation. It is recommended that this test remain speeded, that the plan to reduce the number of items in the test from 50 to 40 be implemented, and that its time limit remain 6 minutes. Based on information from past operational forms (U.S. Department of Labor, 1985), the parallel forms reliability of this test would be approximately 0.84. There are three primary justifications for this recommendation. The first is that computation can legitimately be viewed as a speed construct (Carroll, 1993). The second is that a nonspeeded version of this test would likely require a substantial increase in its current operational time limit. Recall that this investigation's "nonspeeded" version of the test has only 10 items (with a 6 minute time limit) and is still too speeded to be called a nonspeeded test. Finally, the items in current versions of the test require only basic numerical operations (i.e., addition, subtraction, multiplication, and division) and are therefore relatively easy. Because of these low item difficulties, it would be difficult to construct a nonspeeded version of this test with items sufficiently difficult to label it a "power" test.

Three-Dimensional Space. It is recommended that this test be changed to a version containing 20 items with an 8-minute time limit and that the items represent the range of difficulty in the current operational form of the test. A nonspeeded version of Three-Dimensional Space would require at least 20 items to achieve an acceptable level of reliability. Based on the internal consistency reliability of the 15-item nonspeeded new instructions/format Three-Dimensional Space test in this investigation, it is estimated that the internal consistency of this 20-item test would be approximately 0.76 (see Appendix D). There was agreement among the technical advisors that the construct measured by Three-Dimensional Space is one that should be uninfluenced by speed and that a purer measure of this construct (i.e., one not influenced by speed) would likely increase the GATB's potential classification efficiency.

Vocabulary. It is recommended that this test be changed to a version containing 14 items with a 6-minute time limit and that the items represent the range of difficulty in the current operational form of the test. There was agreement among the technical advisors that the construct measured by Vocabulary is one that should be uninfluenced by speed; furthermore, as mentioned above, a purer measure of this construct would likely increase potential classification efficiency. Based on the internal consistency reliability of the 14-item nonspeeded new instructions/format test in this investigation, it is estimated that the internal consistency of this test would be approximately 0.76 (see Appendix D).

Arithmetic Reasoning. It is recommended that this test be changed to a version containing 18 items with a 20-minute time limit and that the items represent the range of difficulty in the current operational form of the test. Based on the internal consistency reliability of the 10-item nonspeeded new instructions/format Arithmetic Reasoning test in this investigation, it is estimated that the internal consistency of this test would be approximately 0.70 (see Appendix D). As is the case for Three-Dimensional Space and Vocabulary there

was agreement that the measurement of Arithmetic Reasoning should be uninfluenced by speed and that such a measure would likely increase the GATB's potential classification efficiency. It is recognized that a 20-minute Arithmetic Reasoning test would be 13 minutes longer than the current operational 7 minute version. However, it was the judgement of the technical advisors that the construct measured by Arithmetic Reasoning is sufficiently central to the prediction of human behavior that it should be measured reliably and that 20 minutes would be the minimum time limit required to achieve adequate reliability using a nonspeeded measure of this construct..

Additional recommendations. The recommendations made above would add 15 minutes to the current battery (i.e., 2 minutes to Three-Dimensional Space and 13 minutes to Arithmetic Reasoning). It is recommended that to minimize the increase in administration time, consideration be given to dropping one of the speed tests, perhaps Form Matching. Form Matching is chosen because it measures nearly the same construct as Tool Matching, but is less reliable. This would save an estimated 11 minutes; thus, bringing the estimated net increase in administration time to only 4 minutes. Viewed in the context that the GATB administration time can vary as much as 30 minutes (personal communication, Mark Rothenberg, December 13, 1993), an estimated net increase of 4 minutes does not seem substantial. It is also recommended that power tests be administered before the paper-and-pencil speed tests. One potential order is (1) Arithmetic Reasoning, (2) Vocabulary, (3) Three-Dimensional Space, (4) Computation, (5) Name Comparison, and (6) Tool Matching. Table 27 contains estimates of what the correlations among Name Comparison and the four power tests would be if the recommendations outlined above were followed and anticipated changes to instructions and format were implemented. These estimates are based on using the Spearman-Brown Prophecy formula to estimate the correlations among the new instructions/format tests if Three-Dimensional Space, Vocabulary, and Arithmetic Reasoning test consisted of the recommended number of items.

Table 27. Estimates of the Correlations Among Tests if the Recommendations Were Followed and the Anticipated Instruction/Format Changes Were Implemented.

Test	Name Comp.	Computation	3-D Space	Vocabulary	Arith. Reas.
Name Comp.	1.00				
Computation	0.64	1.00			
3-D Space	0.32	0.38	1.00		
Vocabulary	0.44	0.53	0.40	1.00	
Arith. Reas.	0.40	0.60	0.46	0.61	1.00

References

Carroll, J. B. (1993). Human Cognitive Abilities: A Survey of Factor-Analytic Studies. New York: Cambridge University Press.

Cronbach, L. J., & Warrington, W. G. (1951). Time limit tests: Estimating their reliability and degree of speeding. Psychometrica, 14, 167-188.

Cudeck, R., & Henly, S. J. (1991). Model selection in covariance structures analysis and the "problem" of sample size: A clarification. Psychological Bulletin, 109, 512-519.

Donlon, T. T. (1973). Establishing time limits for tests. A paper presented at the Annual Meeting of the Northeast Educational Research Association, Ellenville, NY.

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

Hartigan, J. A. & Wigdor, A. K. (Eds.), (1989). Fairness in Employment Testing: Validity Generalization, Minority Issues, and the General Aptitude Test Battery. Washington, DC: National Academy Press.

Peterson, N. G. (1993). Review of issues associated with speededness of GATB tests. Washington, DC: American Institutes for Research.

Rindler, S. E. (1979). Pitfalls in assessing test speededness. Journal of Educational Measurement, 16, 261-270.

Steiger, J. H., & Lind, J. C. (1980). Statistically based tests for the number of common factors. Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA.

U.S. Department of Labor (1985). Manual for the USES General Aptitude Test Battery: Reliability and comparability forms C and D. Washington, DC: U.S. Department of Labor Employment and Training Administration, U.S. Employment Service.

U.S. Department of Labor (in prep.). Developing Items for the GATB. Washington, DC: U.S. Department of Labor Employment and Training Administration, U.S. Employment Service.

APPENDIX A

Information Used to Determine the Number of Items
to Include in this Investigation's Nonspeeded Versions of the
GATB Power Tests

Note: To create nonspeeded versions of the GATB power tests for use in this investigation, it was decided that instead of increasing each test's time limit, it would be preferable to reduce the number of items on each test. This is because (a) a considerable increase in the GATB administration time is currently not considered operationally feasible and (b) based on the results from administering such nonspeeded versions, the Spearman-Brown Prophecy formula could be used to estimate the reliability of nonspeeded versions of any length. However, another issue is the desire to avoid a nonspeeded version of a power test that has so few items that its capacity to represent its intended content domain is severely diminished.

Computation:

Calculations from an operational Form D study, based on a memorandum by Steve Mellon (June 22, 1993)

-   90% of the people completed 16 items, 22.5 sec. per item.

-   This would suggest 16 items.

Based on E & F Forms study data collected after July 16, 1993 (the data set included only scores on Forms administered after July 16, 1993)

-   50 Form A anchor items, $r_{xx} = .90$

-   90% of the examinees completed 84 items within the 67 min. time limit.

-   That is 47.86 sec. per item. (Because of pre-July 16 data, we know this is high. After July 16, the time limit for this test was extended from 52 min. to 67 min.; however, this change did not result in a substantial increase in the number of items completed by examinees.)

-   This would suggest 7.5 items (Far too few.).

-   Using Spearman-Brown on the above reliability estimate:

# of Items	r_{xx}
12	.68
13	.70
14	.71
15	.73
16	.74

The Decision

-   Based on the uncertainty about the amount of time it takes per item, the reliability estimates and the desire to represent the content, we decided to include 12 items within 6 min. (30 sec. per item). This allows for three addition items, three subtraction items, three multiplication items, and three division items.

Three-Dimensional Space:

Calculations from an operational Form D study, based on a memorandum by Steve Mellon (June 22, 1993)

- 90% of the people completed 16 items, 22.5 sec. per item.

- This would suggest 16 items.

Based on E & F Forms study data collected after July 16, 1993 (the data set included only scores on Forms administered after July 16, 1993)

- 50 Form A anchor items, r_{xx} = .88

- 94% of the examinees completed all 105 items within the 50 min. time limit.

- That is 28.57 sec. per item (This is an over estimate because so many people finished.).

- This would suggest 12.41 items.

- Using Spearman-Brown on the above reliability estimate:

# of Items	r_{xx}
12	.68
13	.70
14	.72
15	.73
16	.74
17	.75
18	.77

The Decision

- Based on the uncertainty about the amount of time it takes per item, the reliability estimates and the desire to represent the content, we decided to included 15 items within 6 min. (24 sec. per item). Based on content considerations, we believe 15 items is the number of items that it takes to capture the content domain.

Vocabulary:

Calculations from an operational Form D study, based on a memorandum by Steve Mellon (June 22, 1993)

- 90% of the people completed 15 items, 24 sec. per item.

- This would suggest 15 items.

Based on E & F Forms study data collected after July 16, 1993 (the Pacific data set included only scores on Forms administered after July 16, 1993)

- 50 Form A anchor items, r_{xx} = .91

- 94% of the examinees completed all 140 items within the 70 min. time limit.

- That is 30 sec. per item (This is an over estimate because so many people finished.).

- This would suggest 12 items.

- Using Spearman-Brown on the above reliability estimate:

# of Items	r_{xx}
12	.67
13	.69
14	.70
15	.72
16	.73
17	.74
18	.75

The Decision

- Based on the uncertainty about the amount of time it takes per item, the reliability estimates and the desire to represent the content, we decided to include 14 items within 6 min. (25.71 sec. per item). Based on content considerations, we believe 14 items is the number of items that it takes to capture the content domain.

Arithmetic Reasoning:

Calculations from an operational Form D study, based on a memorandum by Steve Mellon (June 22, 1993)

- 90% of the people completed 8 items, 52.5 sec. per item.

- This would suggest 8 items (within the 7 min. time limit).

Based on E & F Forms study data collected after July 16, 1993 (the data set included only scores on Forms administered after July 16, 1993)

- 50 Form A anchor items, $r_{xx} = .84$

- 90% of the examinees completed all 57 items within the 73 min. time limit.

- That is 76.84 sec. per item.

- This would suggest items 5.45 items (far too few).

- Using Spearman-Brown on the above reliability estimate:

# of Items	r_{xx}
8	.62
9	.65
10	.67
11	.69
12	.71

The Decision

- Based on the uncertainty about the amount of time it takes per item, the reliability estimates and the desire to represent the content, we decided to include 10 items within 11 min. (66 sec. per item). This represents adding 4 minutes on to the existing time limit. Based on content considerations, we believe 10 items is the number of items that it takes to capture the content domain.

APPENDIX B

**Plots of the Percentage of Examinees
Attempting each Possible Number of Items on each Nonspeeded Test
Within Condition and Across Subgroups**

Note: Figures B-1 through B-24 plot the percent of examinees who attempted each possible
number of items on the nonspeeded versions of each test within instructions/format
condition and across subgroups (i.e., Whites, Blacks, Hispanics, Males, Females,
Under 40 Years of Age, & 40 Years of Age and Over). In each of these figures the
number of items attempted is on the horizontal axis and percentage of examinees is on
the vertical axis. The diagonal solid dark line in the upper-right corner represents the
ETS rule-of-thumb; it starts at 75% of the items and 100% of the examinees and ends
at all of the items and 80% of the examinees. Figures B-1 through B-8 compare
Whites, Blacks, Hispanics, and the total group. Figures B-9 through B-16 compare
Males to Females. Finally, Figures B-17 through B-24 compare examinees under 40
years of age to examinees 40 years of age and over.
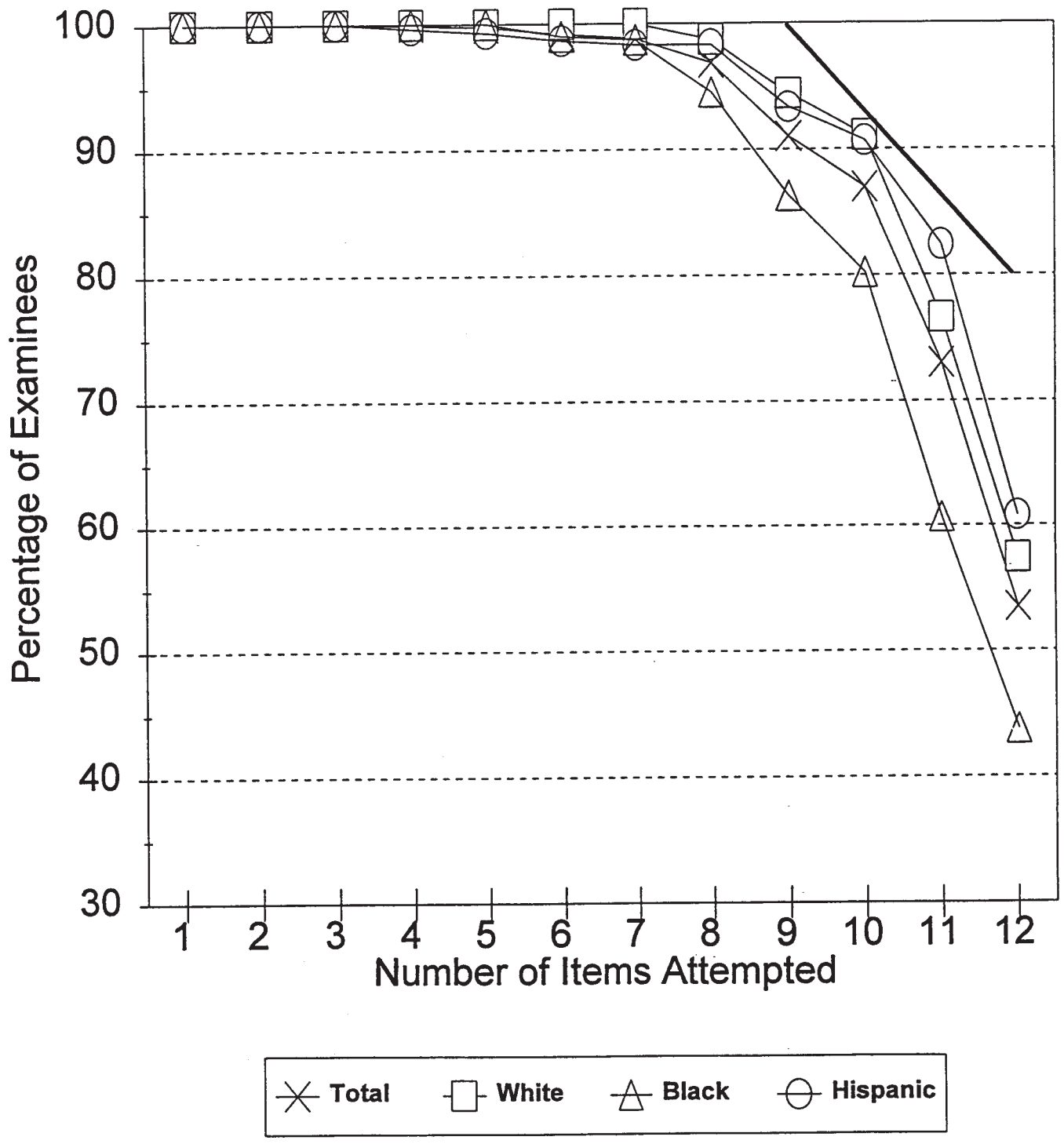
Figure B-1.   Percentage of Examinees in the White, Black, and Hispanic Subgroups, and in the Total Group Attempting Nonspeeded Computation Items in the Old Instructions/Format Condition.
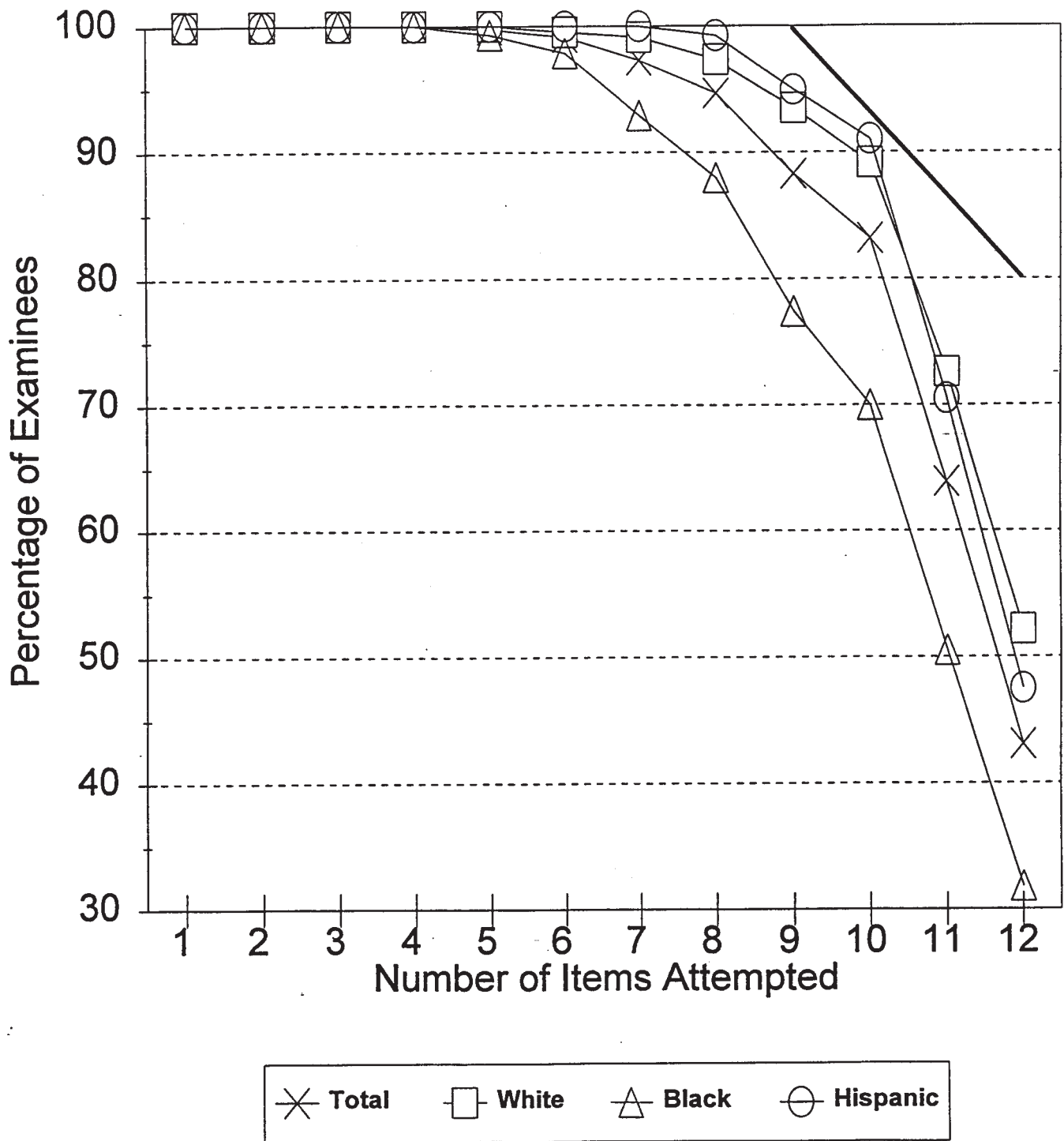
Figure B-2. Percentage of Examinees in the White, Black, and Hispanic Subgroups, and in the Total Group Attempting Nonspeeded Computation Items in the New Instructions/Format Condition.
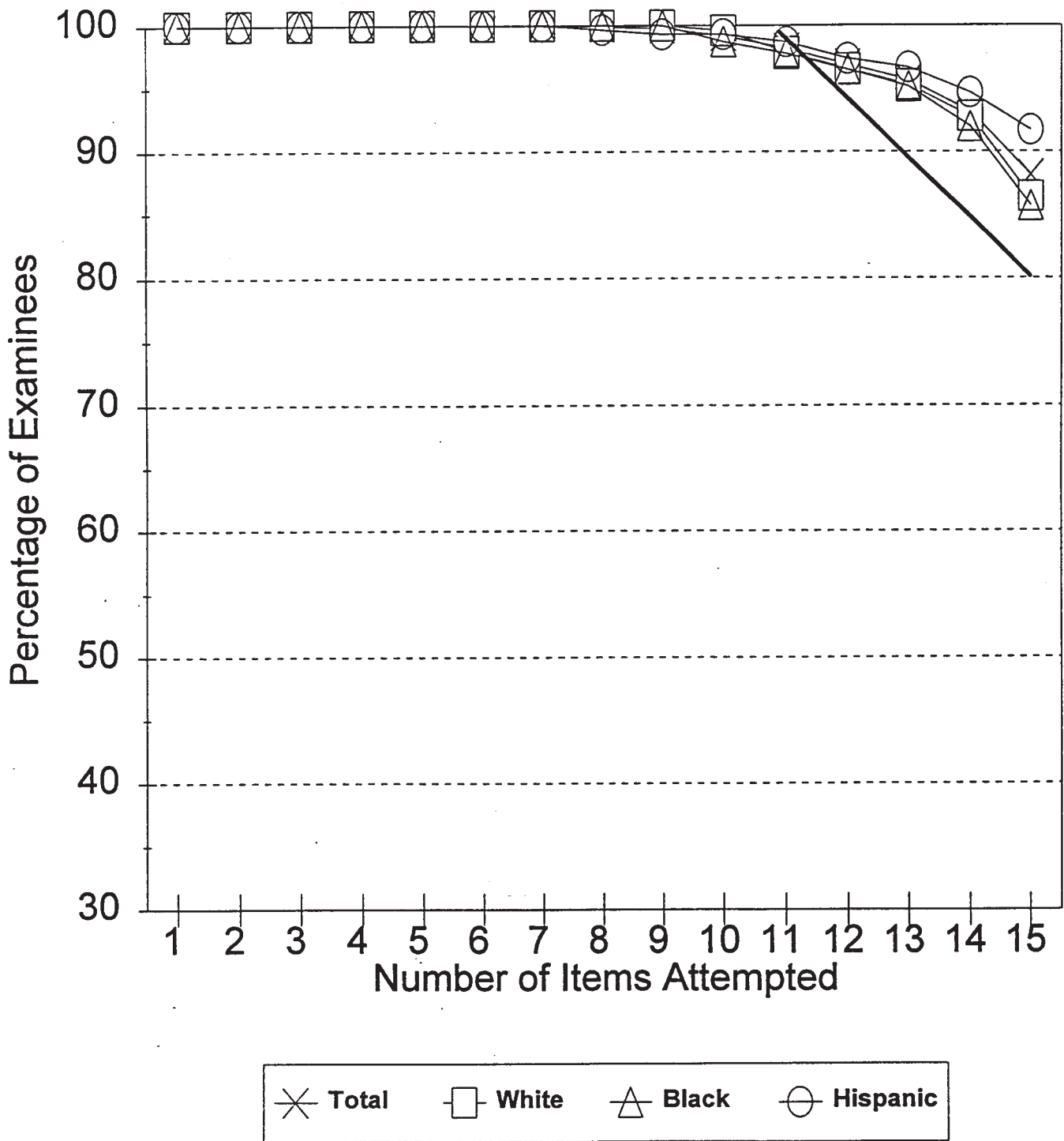
Figure B-3. Percentage of Examinees in the White, Black, and Hispanic Subgroups, and in the Total Group Attempting Nonspeeded Three-Dimensional Space Items in the Old Instructions/Format Condition.
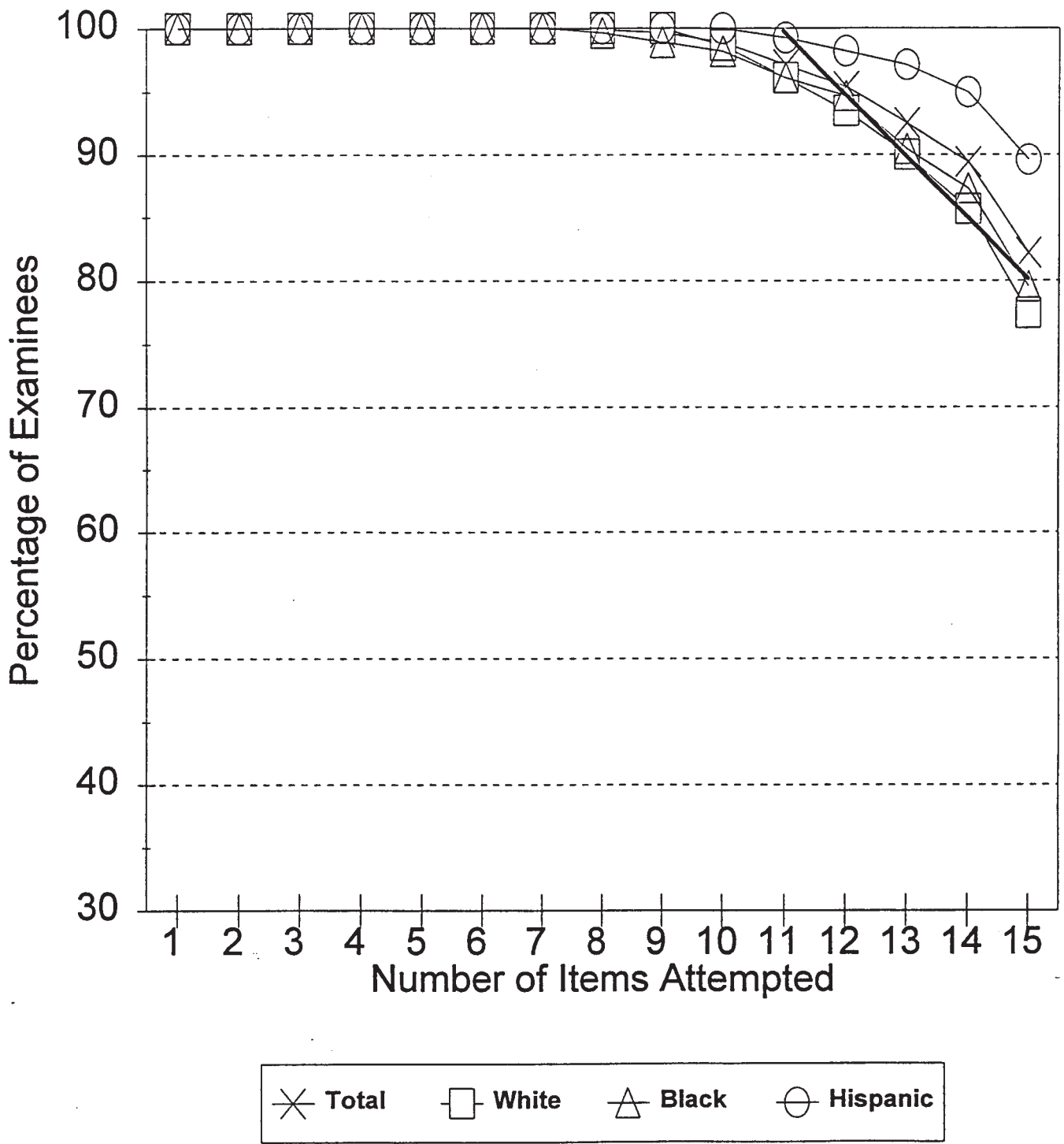
Figure B-4.   Percentage of Examinees in the White, Black, and Hispanic Subgroups, and in the Total Group Attempting Nonspeeded Three-Dimensional Space Items in the New Instructions/Format Condition.
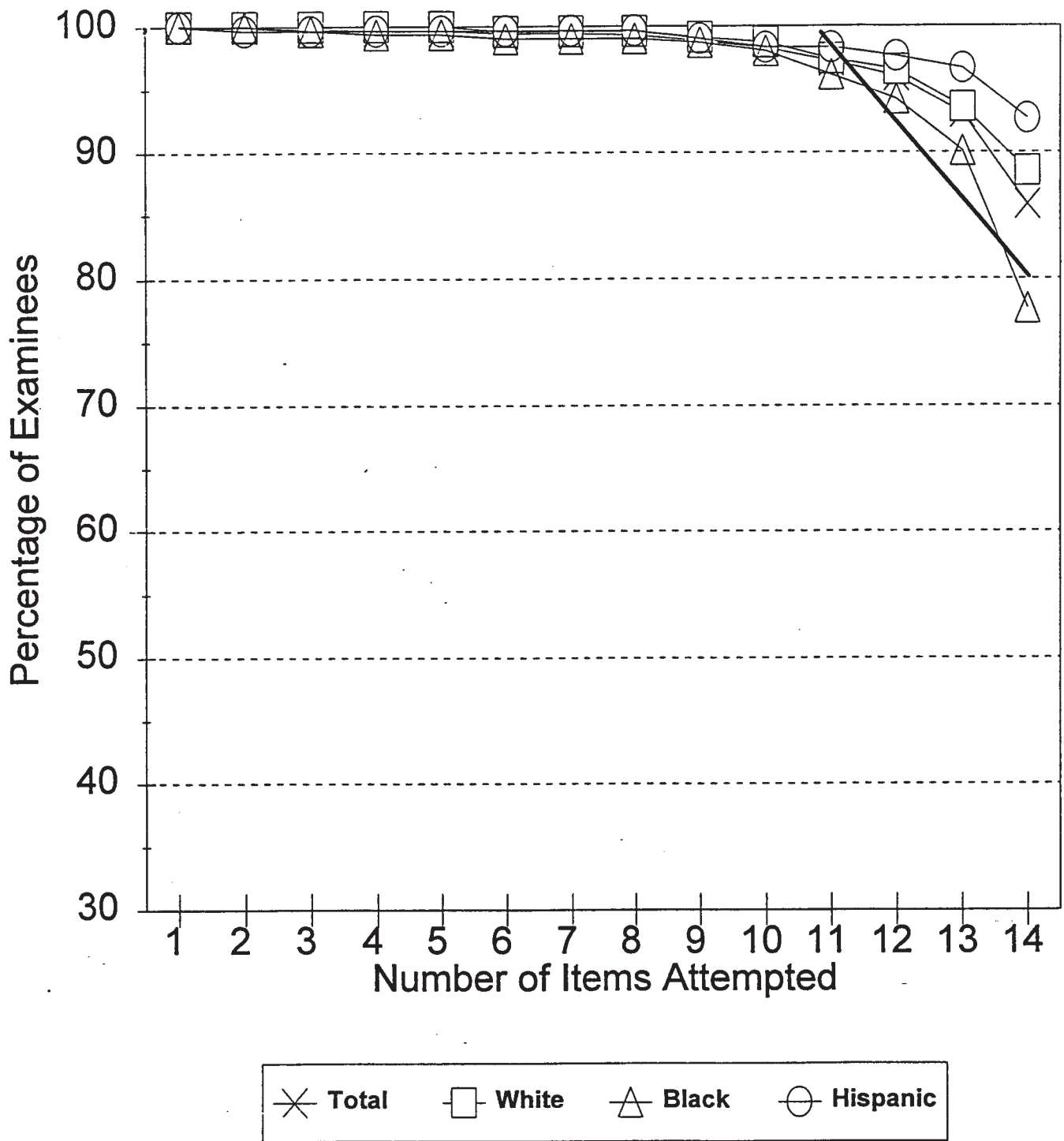
Figure B-5. Percentage of Examinees in the White, Black, and Hispanic Subgroups, and in the Total Group Attempting Nonspeeded Vocabulary Items in the Old Instructions/Format Condition.
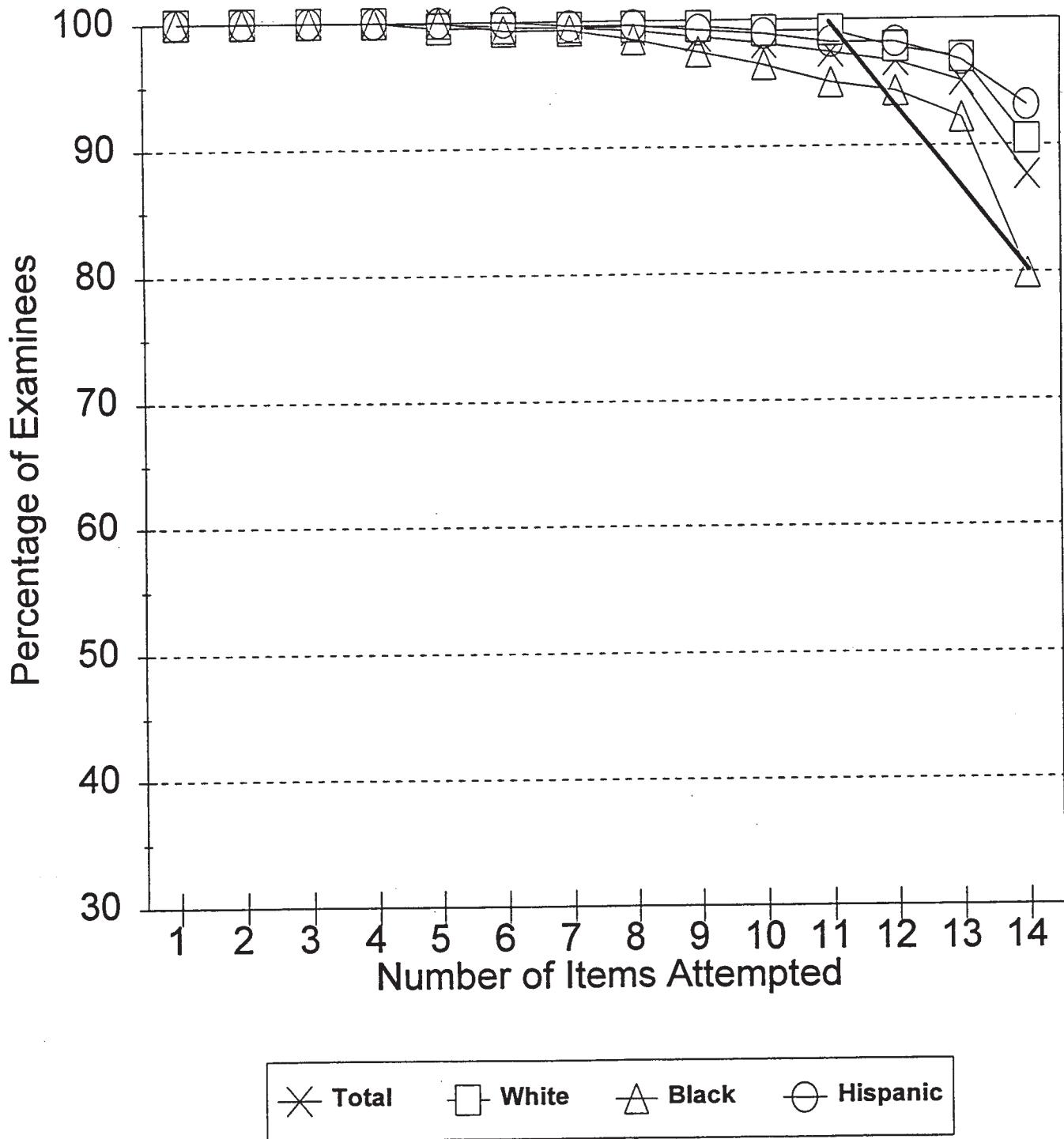
Figure B-6. Percentage of Examinees in the White, Black, and Hispanic Subgroups, and in the Total Group Attempting Nonspeeded Vocabulary Items in the New Instructions/Format Condition.
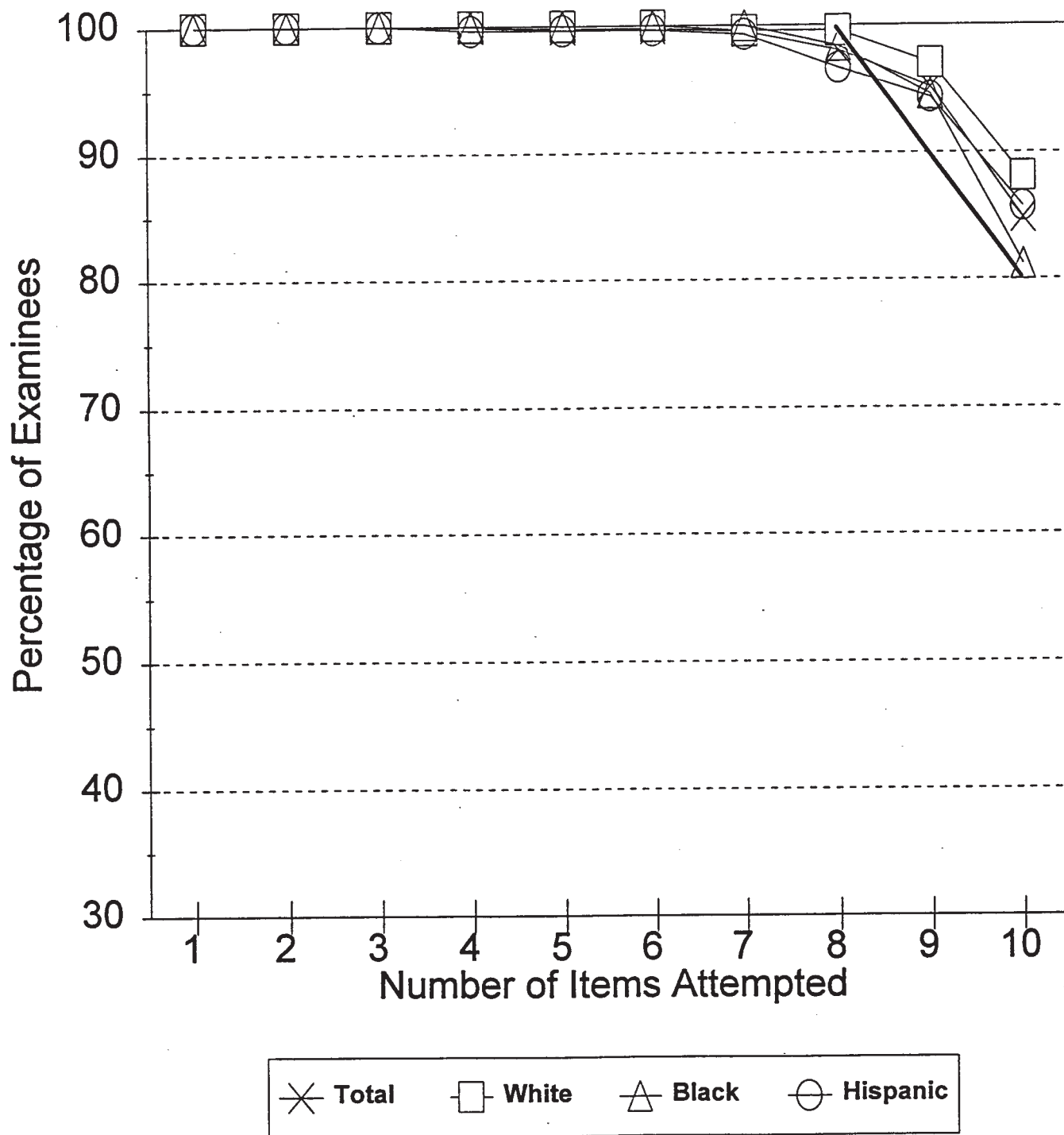
Figure B-7. Percentage of Examinees in the White, Black, and Hispanic Subgroups, and in the Total Group Attempting Nonspeeded Arithmetic Reasoning Items in the Old Instructions/Format Condition.
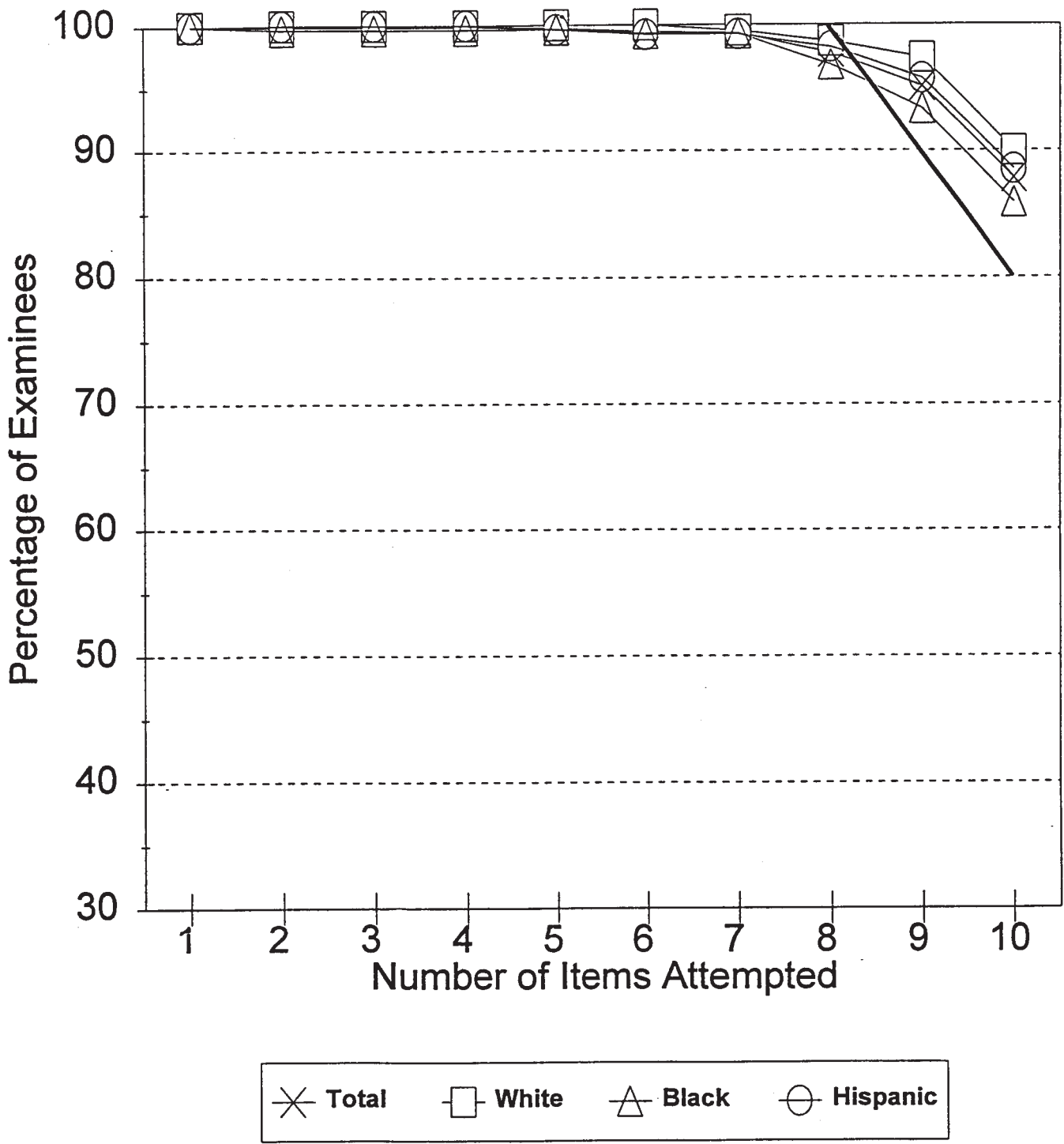
Figure B-8.   Percentage of Examinees in the White, Black, and Hispanic Subgroups, and in the Total Group Attempting Nonspeeded Arithmetic Reasoning Items in the New Instructions/Format Condition.
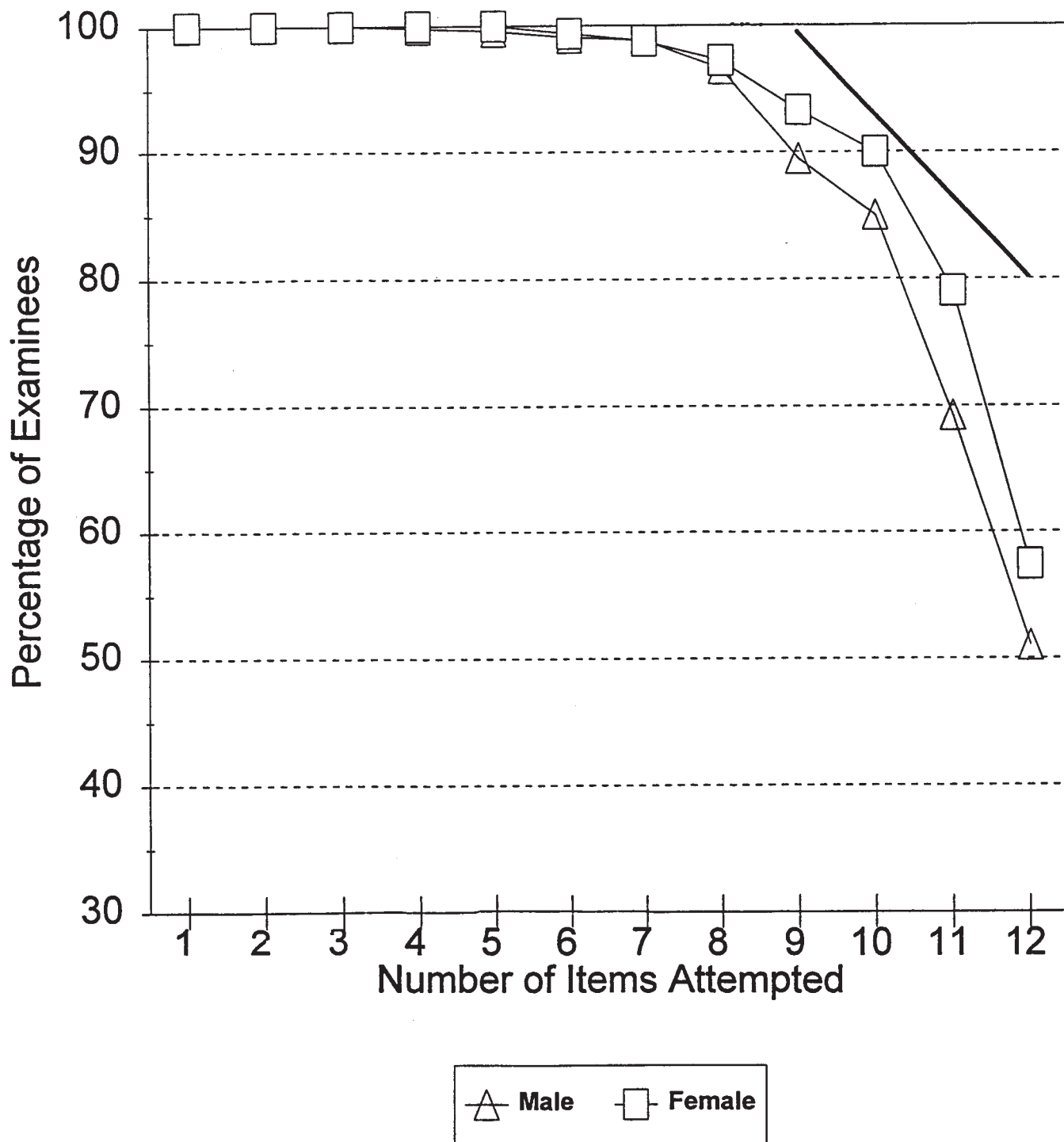
Figure B-9. Percentage of Male and Female Examinees Attempting Nonspeeded
Computation Items in the Old Instructions/Format Condition.
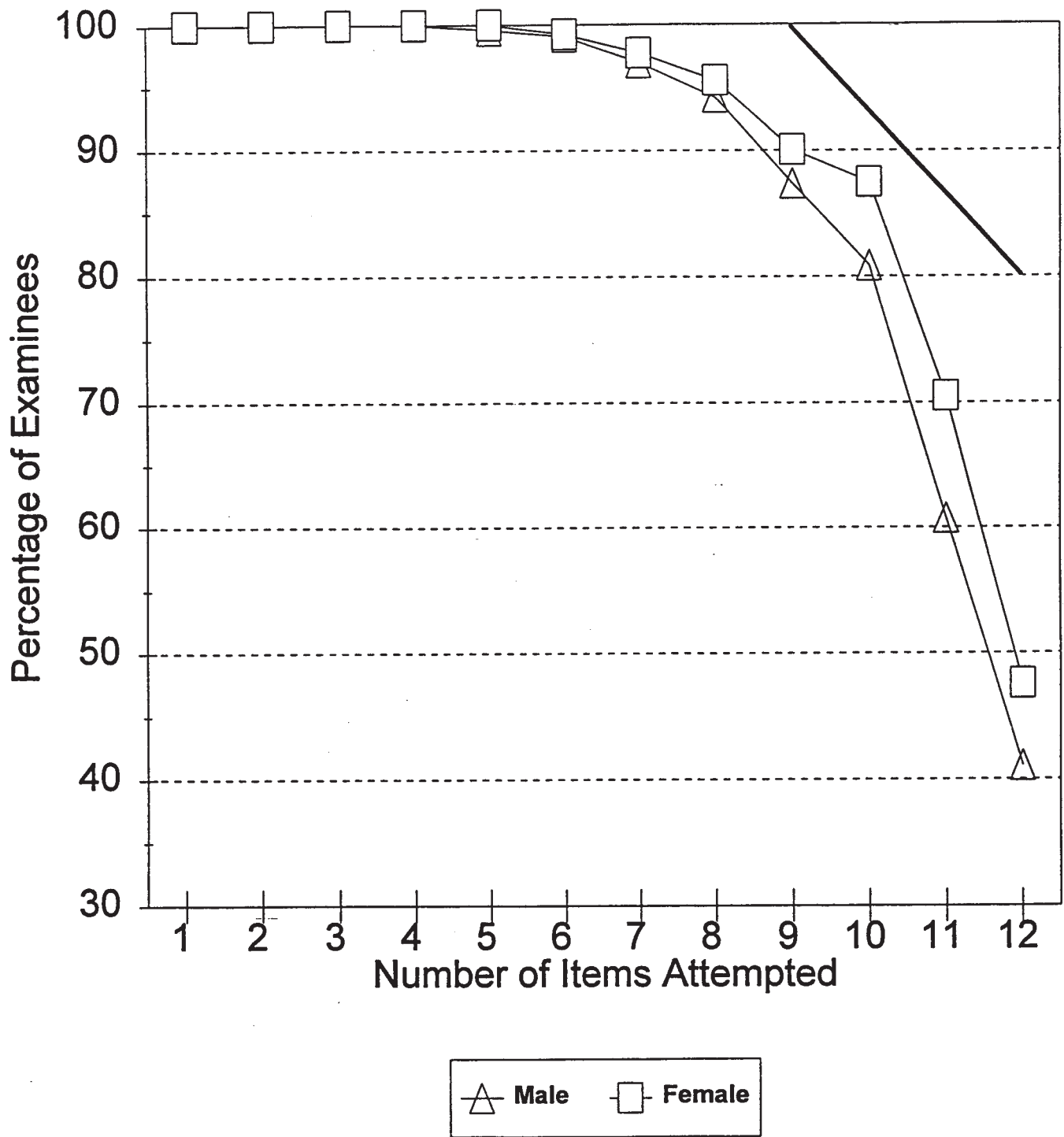
Figure B-10.   Percentage of Male and Female Examinees Attempting Nonspeeded
Computation Items in the New Instructions/Format Condition.
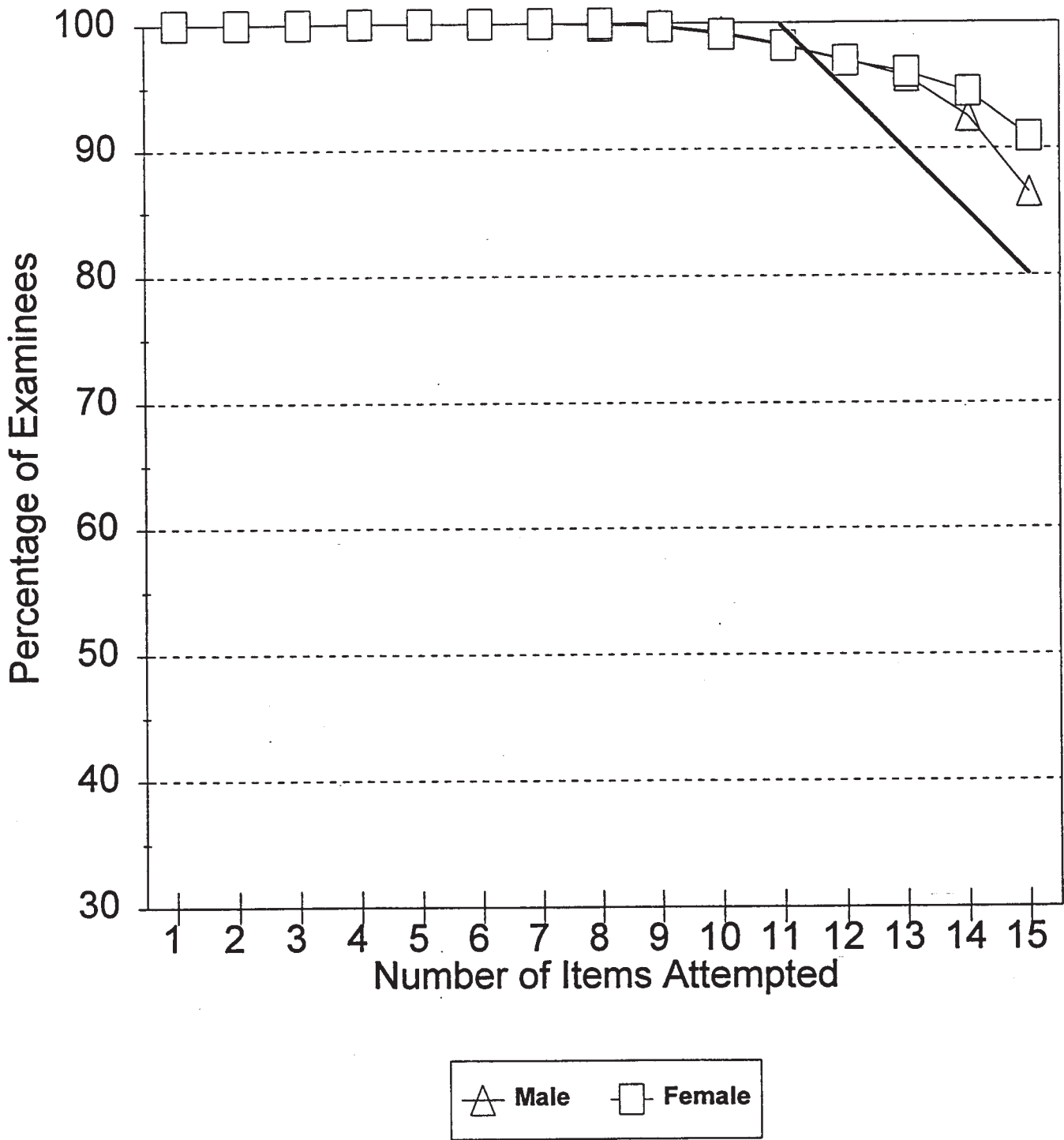
Figure B-11. Percentage of Male and Female Examinees Attempting Nonspeeded Three-Dimensional Space Items in the Old Instructions/Format Condition.
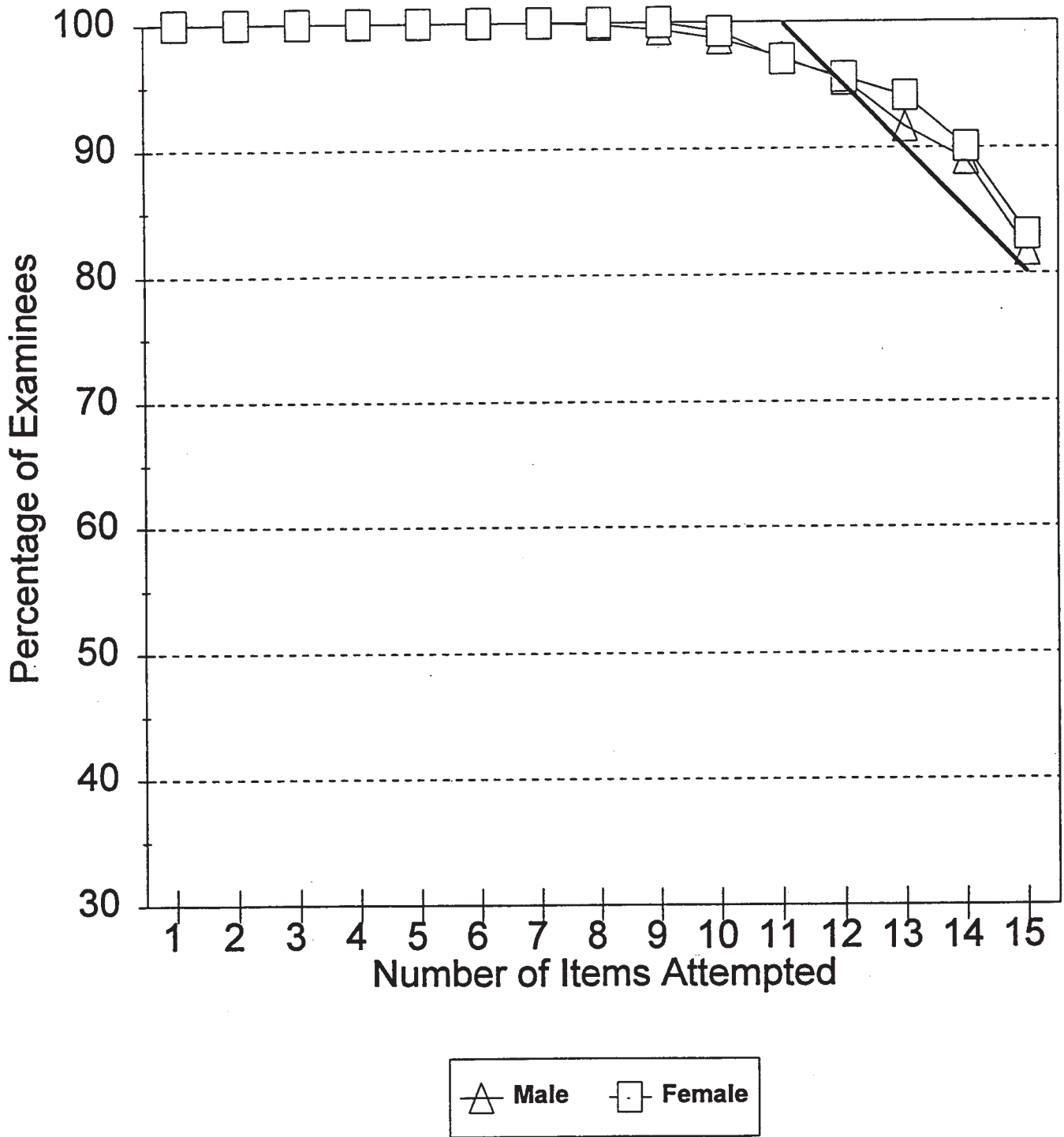
Figure B-12. Percentage of Male and Female Examinees Attempting Nonspeeded Three-Dimensional Space Items in the New Instructions/Format Condition.
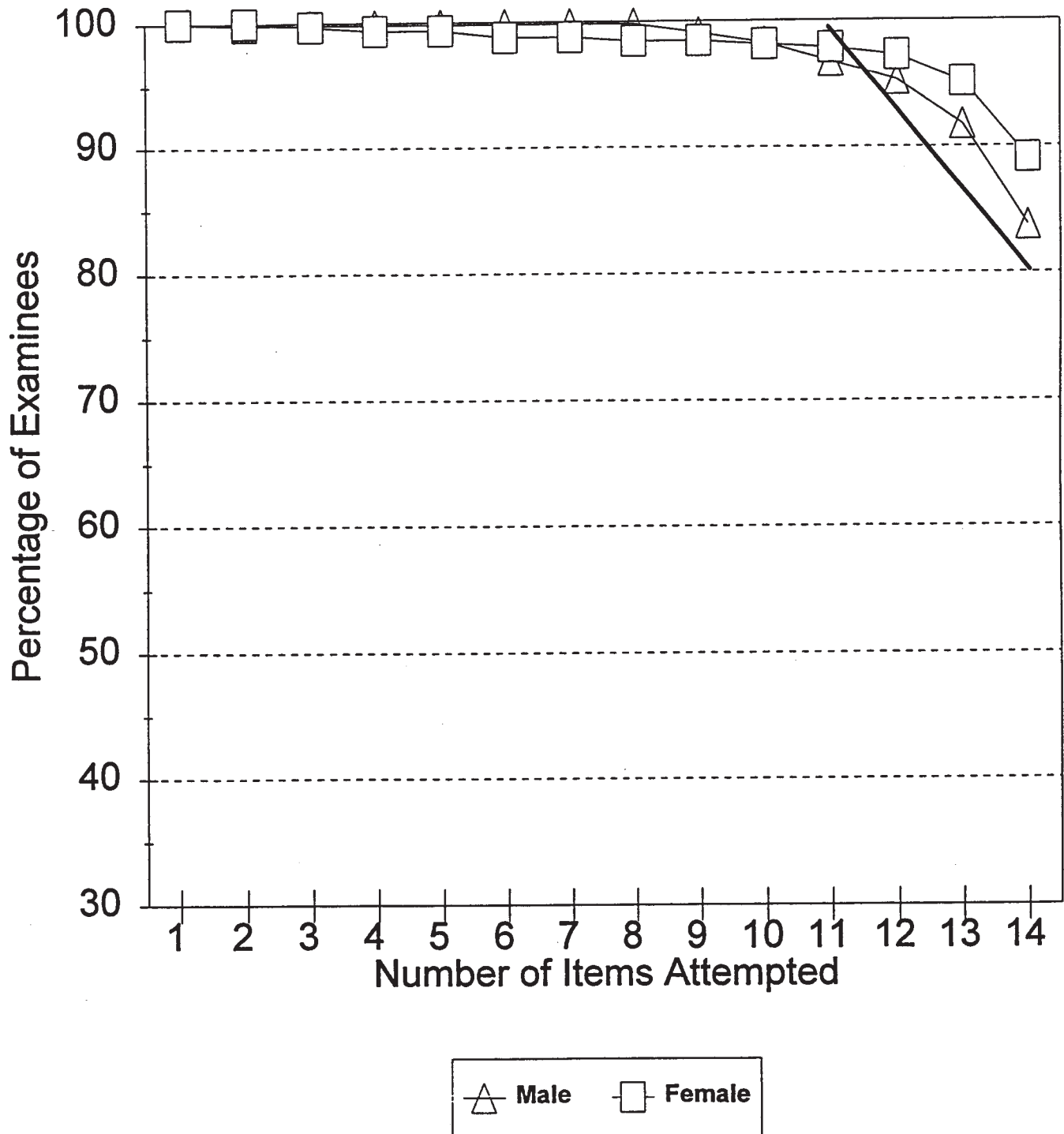
Figure B-13. Percentage of Male and Female Examinees Attempting Nonspeeded Vocabulary Items in the Old Instructions/Format Condition.
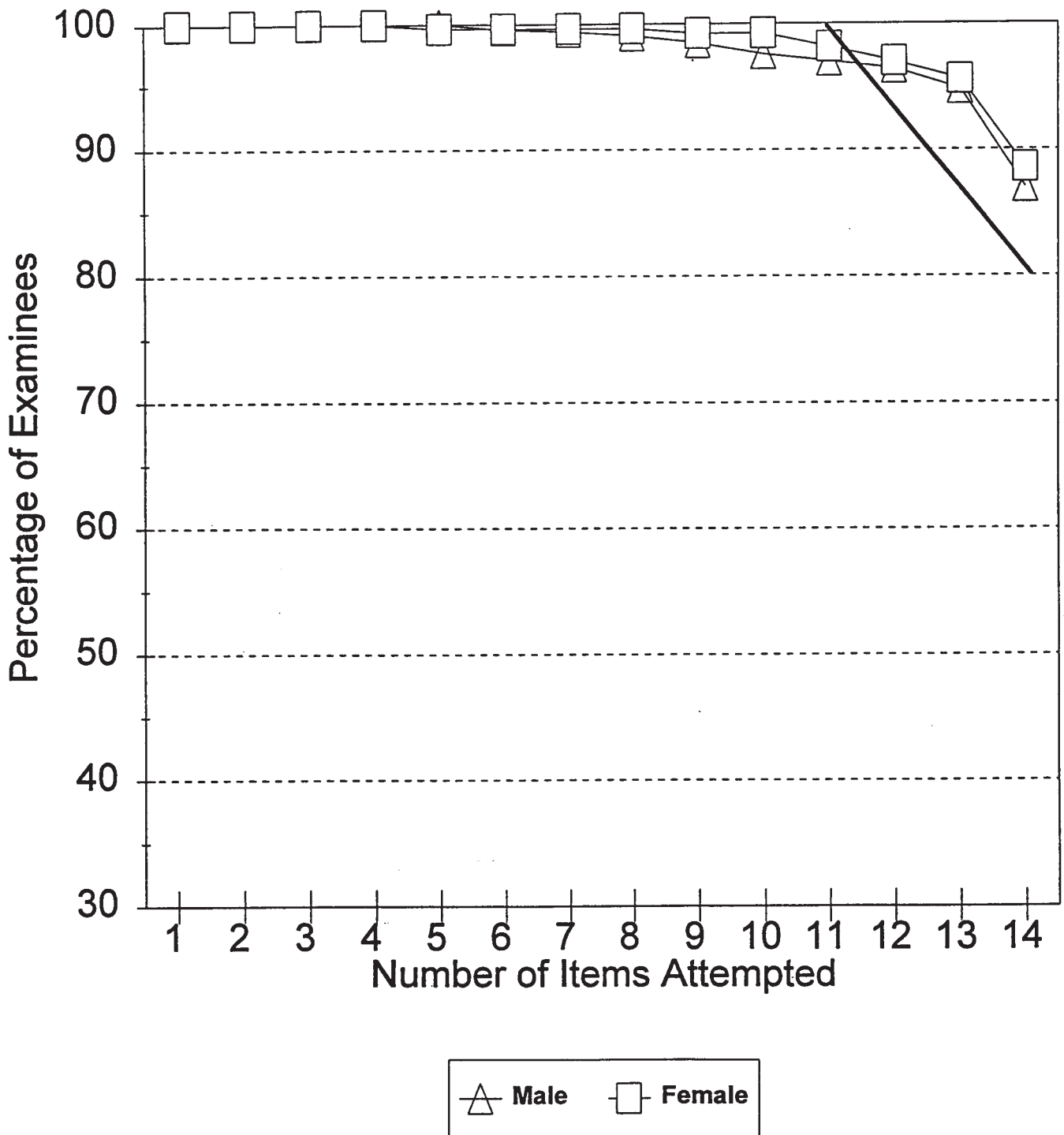
Figure B-14. Percentage of Male and Female Examinees Attempting Nonspeeded Vocabulary Items in the New Instructions/Format Condition.
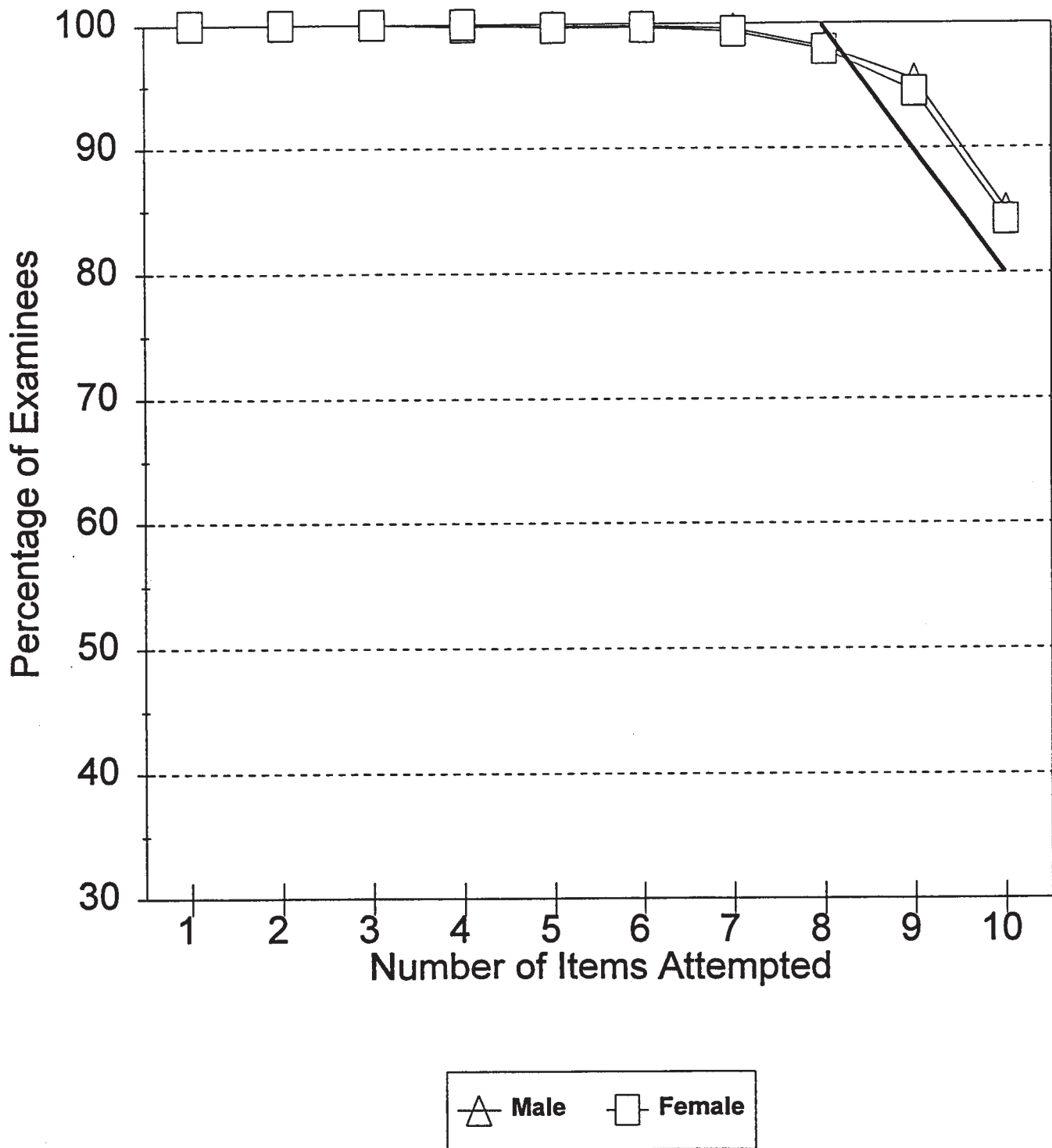
Figure B-15. Percentage of Male and Female Examinees Attempting Nonspeeded Arithmetic Reasoning Items in the Old Instructions/Format Condition.
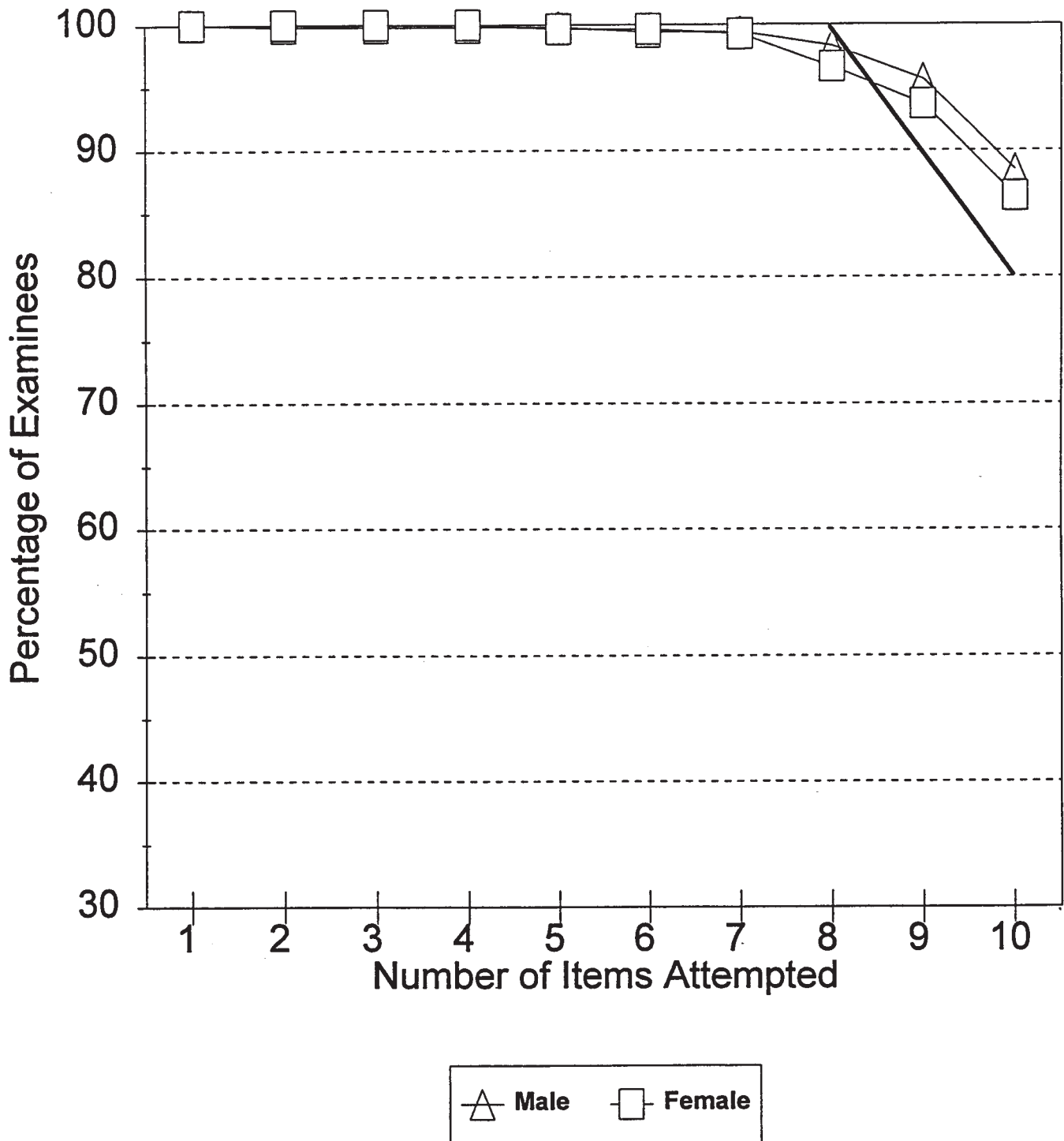
Figure B-16. Percentage of Male and Female Examinees Attempting Nonspeeded Arithmetic Reasoning Items in the New Instructions/Format Condition.
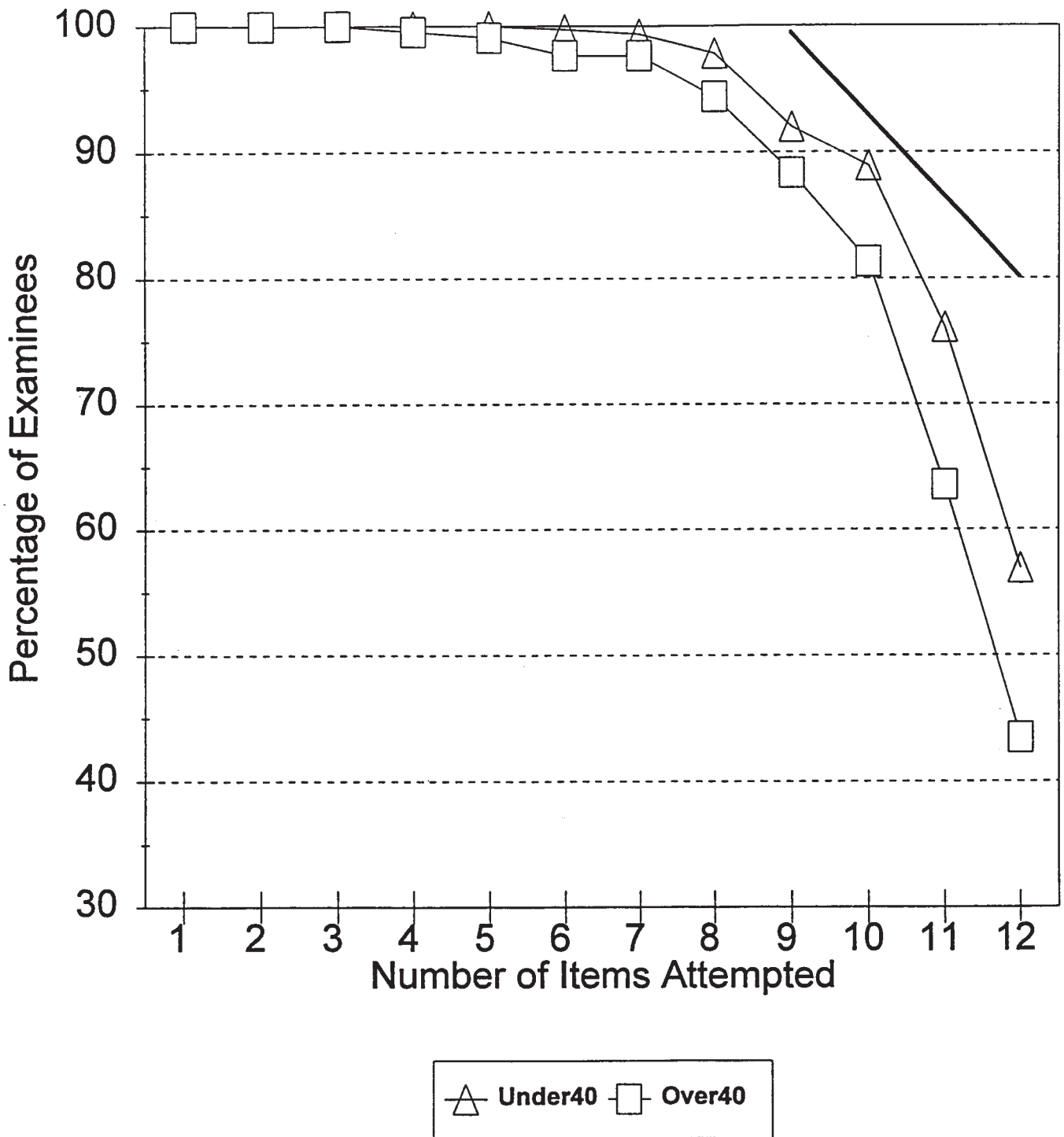
Figure B-17. Percentage of Examinees Under 40 Years of Age and 40 Years of Age and Over Attempting Nonspeeded Computation Items in the Old Instructions/Format Condition.
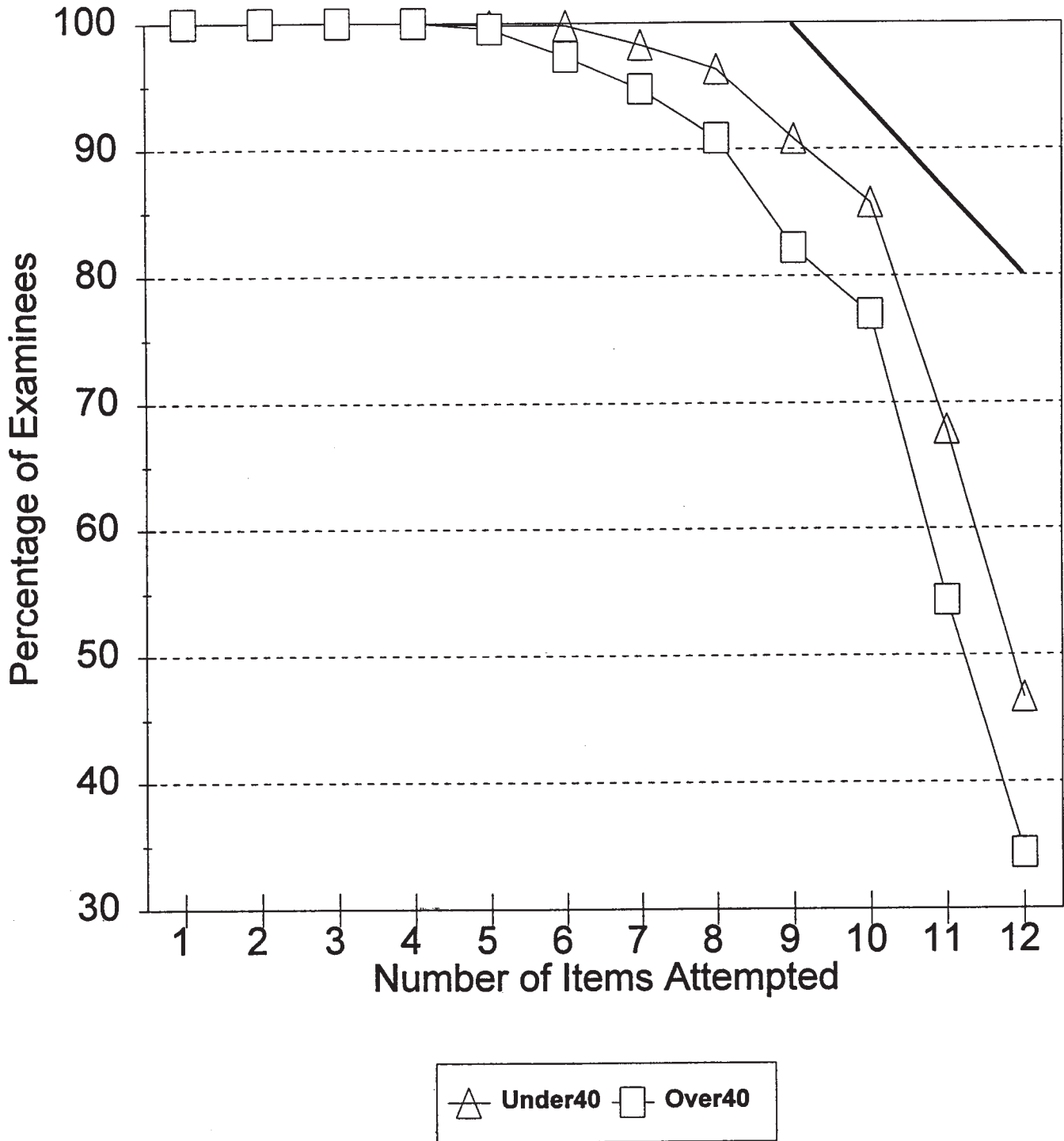
Figure B-18. Percentage of Examinees Under 40 Years of Age and 40 Years of Age and Over Attempting Nonspeeded Computation Items in the New Instructions/Format Condition.
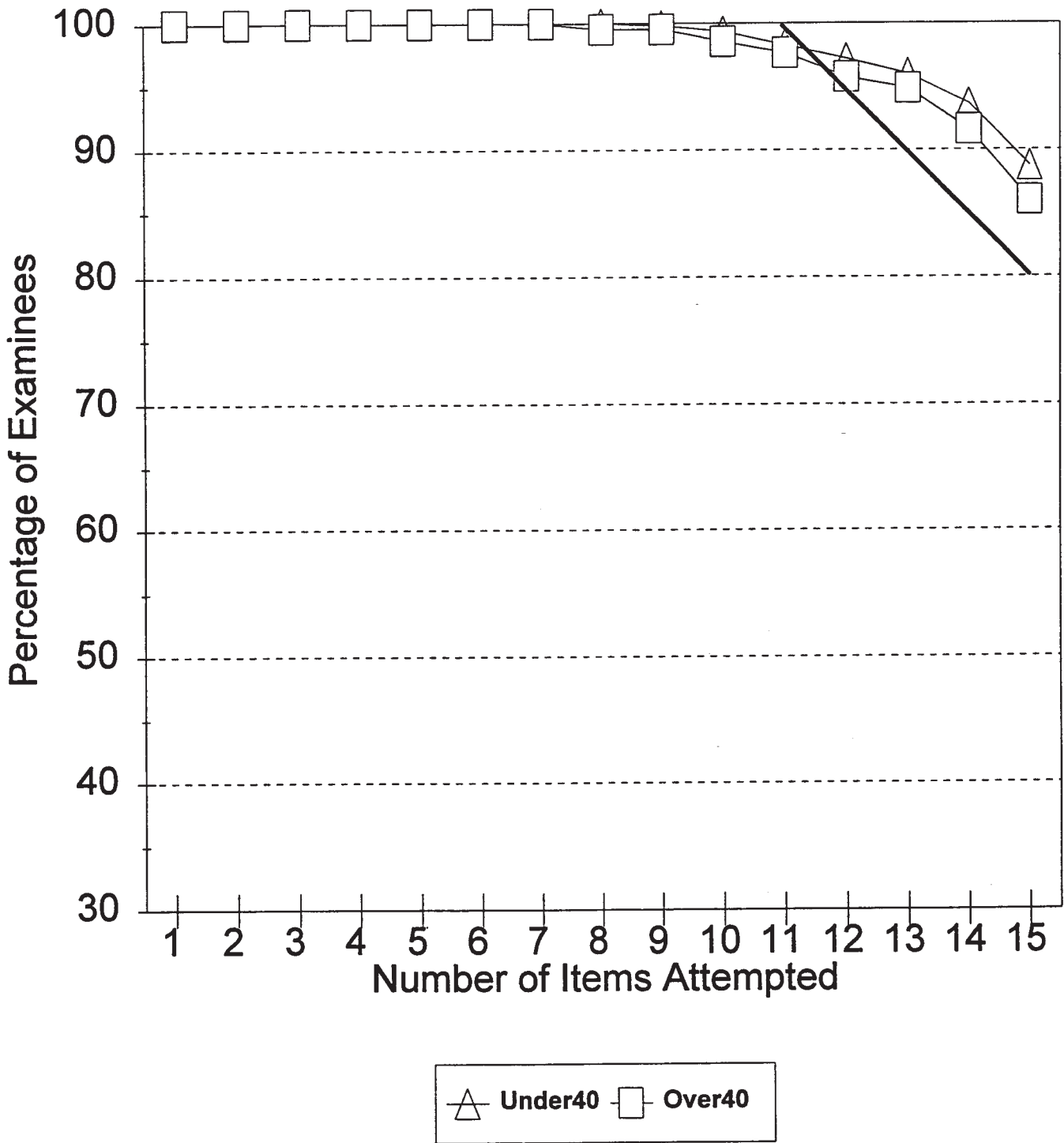
Figure B-19. Percentage of Examinees Under 40 Years of Age and 40 Years of Age and Over Attempting Nonspeeded Three-Dimensional Space Items in the Old Instructions/Format Condition.
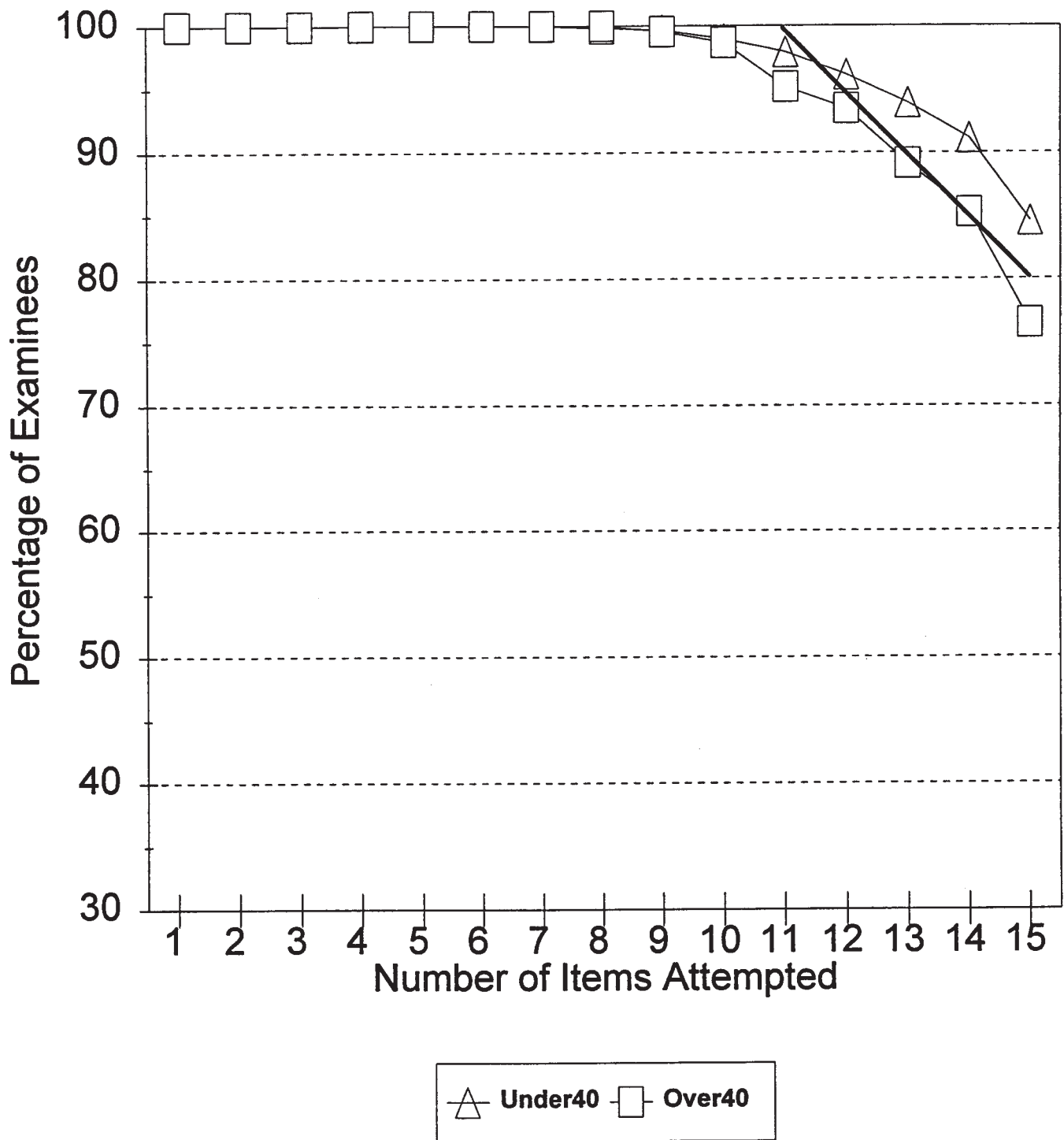
Figure B-20. Percentage of Examinees Under 40 Years of Age and 40 Years of Age and Over Attempting Nonspeeded Three-Dimensional Space Items in the New Instructions/Format Condition.
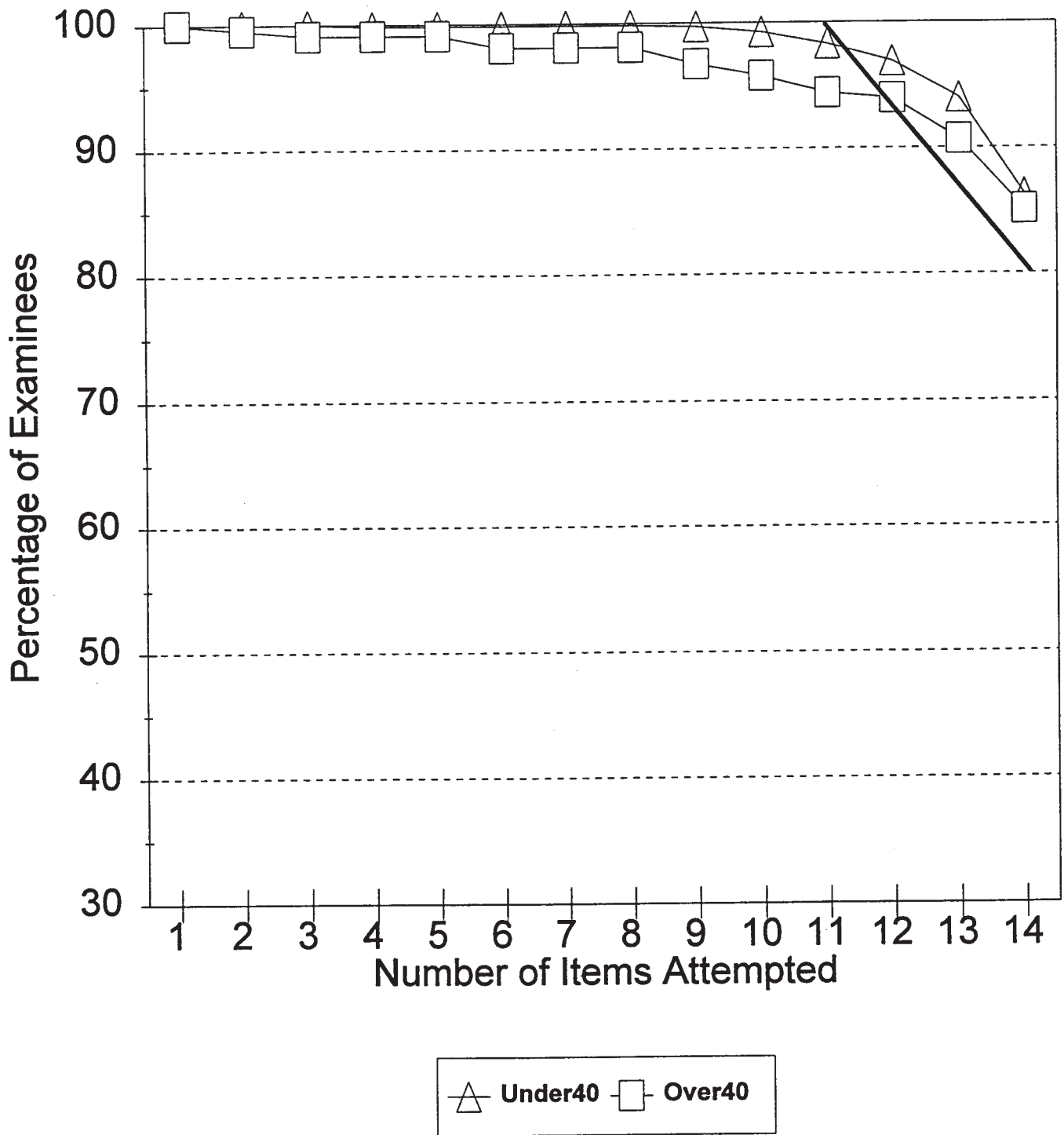
Figure B-21. Percentage of Examinees Under 40 Years of Age and 40 Years of Age and Over Attempting Nonspeeded Vocabulary Items in the Old Instructions/Format Condition.
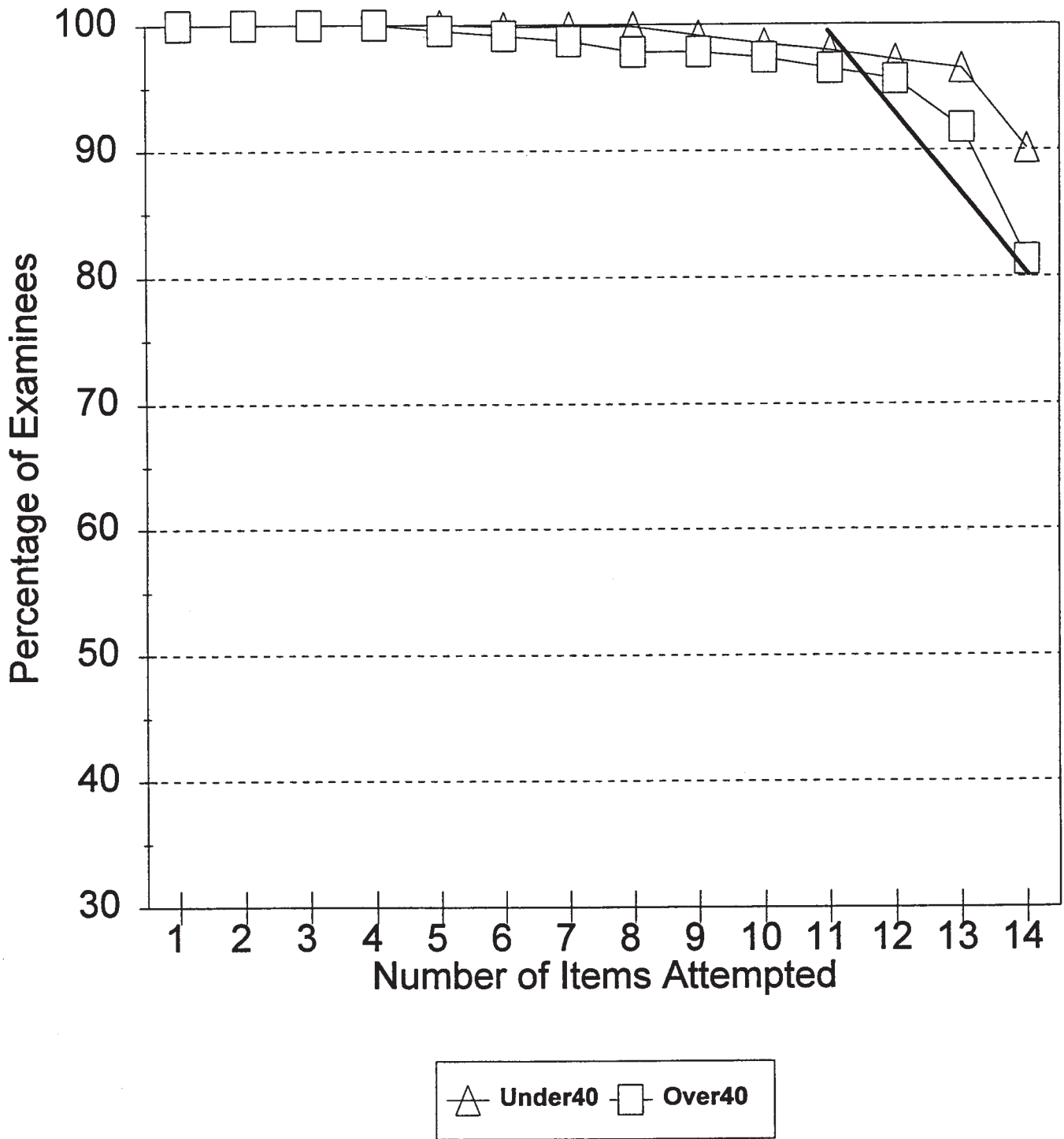
Figure B-22. Percentage of Examinees Under 40 Years of Age and 40 Years of Age and Over Attempting Nonspeeded Vocabulary Items in the New Instructions/Format Condition.
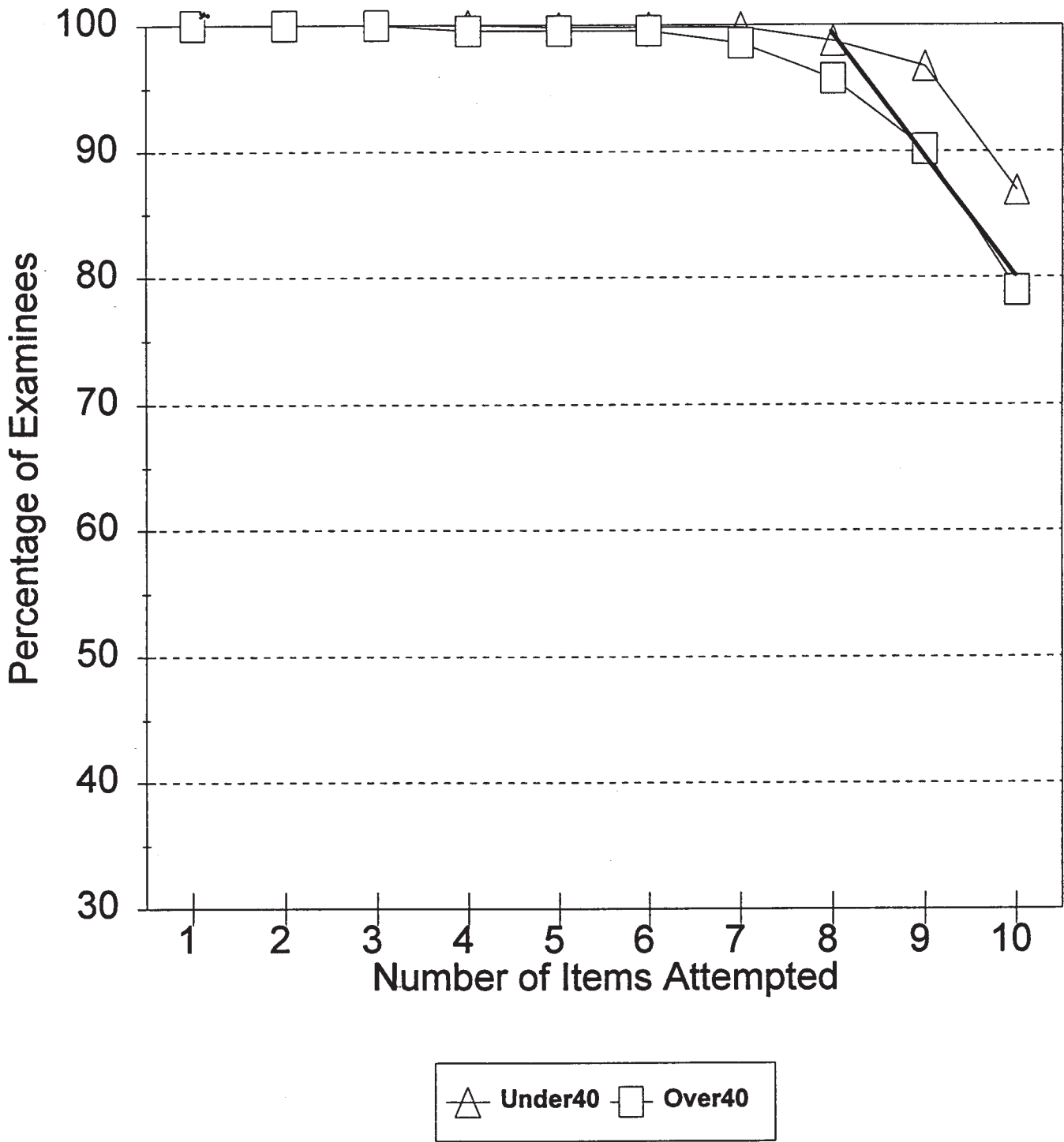
Figure B-23.   Percentage of Examinees Under 40 Years of Age and 40 Years of Age and
Over Attempting Nonspeeded Arithmetic Reasoning Items in the Old
Instructions/Format Condition.
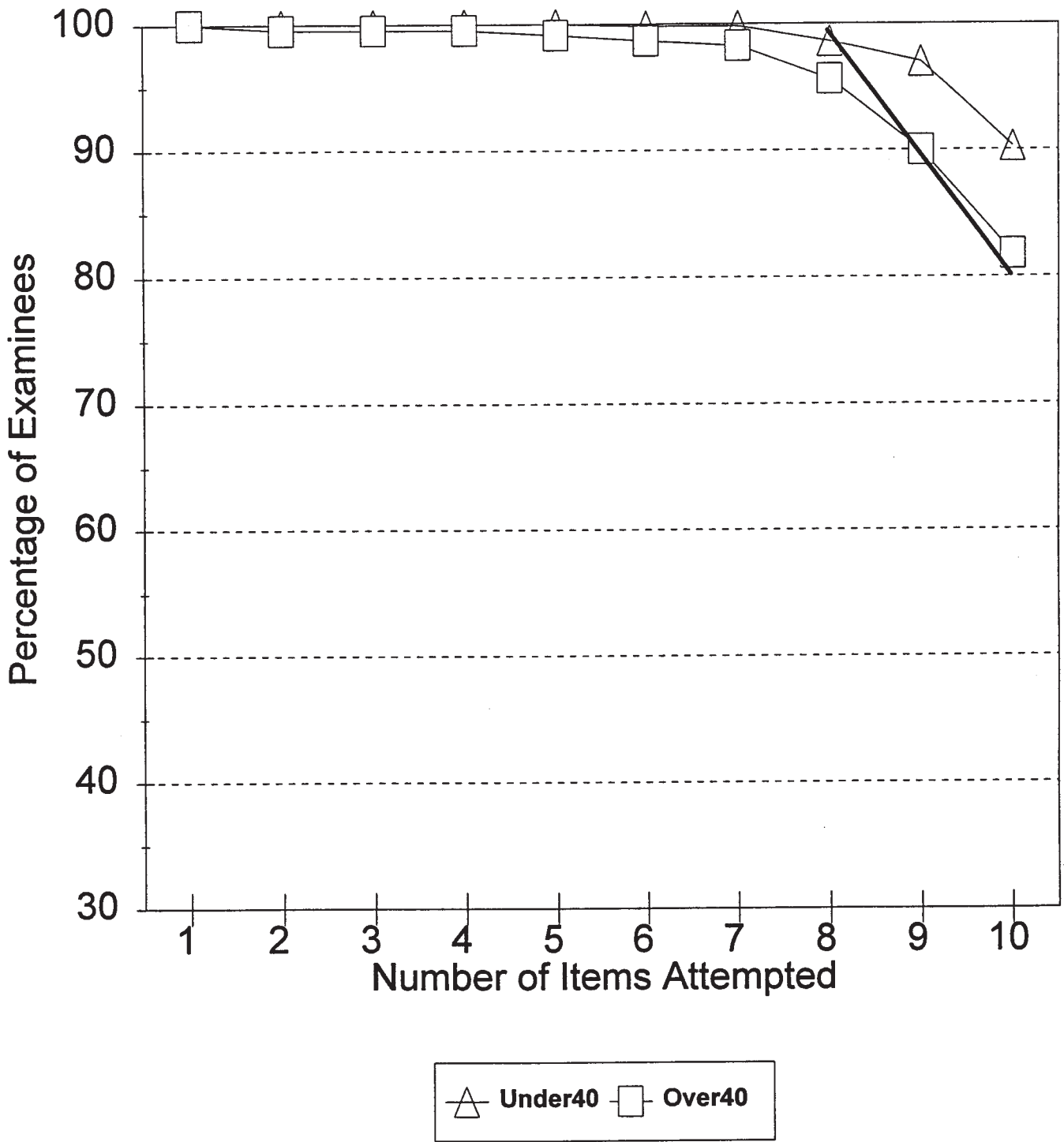
Figure B-24. Percentage of Examinees Under 40 Years of Age and 40 Years of Age and Over Attempting Nonspeeded Arithmetic Reasoning Items in the New Instructions/Format Condition.

APPENDIX C

Five Alternative Scores
by Group
Within the Old Instructions/Format and New Instructions/Format Conditions

Note: In each of the following tables Correct refers to the raw score; Wrong refers to the number of items that were responded to incorrectly; Answered refers to the number of items to which there was a response; Proportion Attempted equals the number of items answered divided by the total number of items; and Accuracy equals the number of items correct divided by the number of items answered. The columns showing mean Proportion Attempted and Accuracy are in bold. Tables C-1 through C-7 are the five alternative scores tables for Whites, Blacks, Hispanics, Males, Females, examinees under 40 years of age, and examinees 40 years of age and over in the old instructions/format condition. Tables C-8 through C-14 are the same tables for examinees in the new instructions/format condition. The sample sizes in these tables are slightly smaller than the sample sizes in Tables 7, 8, 9, and 10 because a small number of individuals who did not answer any items on a test were eliminated. This was done so that calculation of the Accuracy score would not result in a value of 0 in the denominator.

Table C-1.   Five Alternative Scores on Old Instr./Format: Whites   (N=234).

Test	Correct		Wrong		Answered		Proportion Attempted		Accuracy	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Speeded Name Comparison	44.14	12.59	4.65	7.61	48.79	14.75	0.33	0.10	0.91	0.10
Speeded Computation	21.63	6.18	3.43	4.15	25.06	5.61	0.50	0.11	0.86	0.15
Nonspeeded Computation	8.77	2.24	2.42	2.07	11.19	1.19	0.93	0.10	0.78	0.18
Speeded Three Dimensional	16.76	6.49	10.38	8.37	27.14	7.19	0.68	0.18	0.64	0.24
Nonspeeded Three Dimensional	7.96	3.10	6.73	3.28	14.68	0.99	0.98	0.07	0.55	0.22
Speeded Vocabulary	18.91	7.67	8.58	7.49	27.50	7.48	0.46	0.12	0.69	0.22
Nonspeeded Vocabulary	8.47	3.36	5.26	3.25	13.73	0.99	0.98	0.07	0.61	0.24
Speeded Arithmetic Reasoning	10.31	3.65	4.32	4.13	14.62	3.83	0.58	0.15	0.72	0.22
Nonspeeded Arithmetic Reasoning	6.05	2.21	3.79	2.27	9.84	0.49	0.98	0.05	0.62	0.23

Note: Proportion Attempted = Number Answered / Number of Items.
         Accuracy = Number Correct / Number Answered.

C-2

Table C-2. Five Alternative Scores on Old Instr./Format: Blacks (N=312).

Test	Correct		Wrong		Answered		Proportion Attempted		Accuracy	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Speeded Name Comparison	36.86	11.66	5.37	5.17	42.23	11.82	0.28	0.08	0.87	0.13
Speeded Computation	17.72	6.14	3.76	4.03	21.48	5.38	0.43	0.11	0.82	0.18
Nonspeeded Computation	7.65	2.59	2.98	2.34	10.63	1.61	0.89	0.13	0.72	0.21
Speeded Three Dimensional	12.30	5.32	13.85	8.73	26.15	7.42	0.65	0.19	0.50	0.24
Nonspeeded Three Dimensional	6.12	2.66	8.54	2.84	14.66	1.05	0.98	0.07	0.42	0.19
Speeded Vocabulary	13.03	6.61	10.31	7.70	23.34	7.72	0.39	0.13	0.56	0.24
Nonspeeded Vocabulary	6.70	3.26	6.80	3.18	13.50	1.38	0.96	0.10	0.50	0.23
Speeded Arithmetic Reasoning	7.23	3.15	5.38	4.55	12.62	4.38	0.50	0.18	0.60	0.25
Nonspeeded Arithmetic Reasoning	4.59	1.87	5.15	1.93	9.74	0.61	0.97	0.06	0.47	0.19

Note: Proportion Attempted = Number Answered / Number of Items.
Accuracy = Number Correct / Number Answered.

C-3

Table C-3. Five Alternative Scores on Old Instr./Format: Hispanics (N=301).

Test	Correct		Wrong		Answered		Proportion Attempted		Accuracy	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Speeded Name Comparison	40.97	12.79	5.70	7.53	46.67	14.82	0.31	0.10	0.88	0.13
Speeded Computation	19.79	5.47	3.63	4.31	23.43	5.34	0.47	0.11	0.85	0.15
Nonspeeded Computation	8.34	2.40	2.87	2.18	11.22	1.38	0.93	0.11	0.74	0.19
Speeded Three Dimensional	15.97	5.75	11.56	8.33	27.53	7.27	0.69	0.18	0.61	0.22
Nonspeeded Three Dimensional	7.23	2.82	7.54	2.94	14.77	0.94	0.98	0.06	0.49	0.19
Speeded Vocabulary	14.76	6.33	11.79	7.94	26.55	7.66	0.44	0.13	0.57	0.21
Nonspeeded Vocabulary	6.65	2.97	7.15	2.97	13.80	1.04	0.99	0.07	0.48	0.21
Speeded Arithmetic Reasoning	8.39	2.93	5.18	4.34	13.57	3.91	0.54	0.16	0.65	0.22
Nonspeeded Arithmetic Reasoning	4.96	1.89	4.79	2.01	9.75	0.76	0.98	0.08	0.51	0.20

Note: Proportion Attempted = Number Answered / Number of Items.
       Accuracy = Number Correct / Number Answered.

Table C-4. Five Alternative Scores on Old Instr./Format: Males (N=533).

Test	Correct		Wrong		Answered		Proportion Attempted		Accuracy	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Speeded Name Comparison	38.55	12.27	5.30	6.03	43.85	12.93	**0.29**	0.09	**0.88**	0.13
Speeded Computation	19.20	6.14	3.63	3.97	22.83	5.48	**0.46**	0.11	**0.83**	0.17
Nonspeeded Computation	8.06	2.51	2.83	2.23	10.89	1.52	**0.91**	0.13	**0.74**	0.20
Speeded Three Dimensional	15.14	6.41	11.11	8.17	26.25	7.08	**0.66**	0.18	**0.60**	0.24
Nonspeeded Three Dimensional	7.16	3.06	7.53	3.18	14.69	0.99	**0.98**	0.07	**0.49**	0.21
Speeded Vocabulary	15.25	7.46	9.83	7.56	25.08	7.67	**0.42**	0.13	**0.61**	0.23
Nonspeeded Vocabulary	7.32	3.42	6.31	3.33	13.64	1.11	**0.97**	0.08	**0.53**	0.24
Speeded Arithmetic Reasoning	8.63	3.65	4.82	4.33	13.45	4.00	**0.54**	0.16	**0.66**	0.25
Nonspeeded Arithmetic Reasoning	5.33	2.11	4.45	2.16	9.78	0.65	**0.98**	0.06	**0.55**	0.22

Note: Proportion Attempted = Number Answered / Number of Items.
Accuracy = Number Correct / Number Answered.

Table C-5. Five Alternative Scores on Old Instr./Format: Females (N=327).

Test	Correct		Wrong		Answered		Proportion Attempted		Accuracy	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Speeded Name Comparison	43.13	12.80	5.31	8.00	48.44	15.15	0.32	0.10	0.89	0.12
Speeded Computation	20.09	5.99	3.56	4.39	23.65	5.73	0.47	0.11	0.85	0.15
Nonspeeded Computation	8.47	2.38	2.68	2.19	11.15	1.33	0.93	0.11	0.76	0.19
Speeded Three Dimensional	14.29	5.69	13.68	9.03	27.97	7.54	0.70	0.19	0.54	0.24
Nonspeeded Three Dimensional	6.84	2.73	7.91	2.94	14.75	0.96	0.98	0.06	0.47	0.19
Speeded Vocabulary	15.31	6.90	11.21	8.17	26.53	8.00	0.44	0.13	0.59	0.22
Nonspeeded Vocabulary	6.93	3.06	6.78	3.03	13.70	1.30	0.98	0.09	0.51	0.22
Speeded Arithmetic Reasoning	8.24	3.05	5.34	4.38	13.58	4.36	0.54	0.17	0.63	0.22
Nonspeeded Arithmetic Reasoning	4.76	1.94	4.99	2.03	9.75	0.69	0.98	0.07	0.49	0.20

Note: Proportion Attempted = Number Answered / Number of Items.
Accuracy = Number Correct / Number Answered.

Table C-6.   Five Alternative Scores on Old Instr./Format: Under 40 Years of Age   (N=647).

Test	Correct		Wrong		Answered		Proportion Attempted		Accuracy	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Speeded Name Comparison	40.97	12.34	5.21	6.06	46.18	12.94	0.31	0.09	0.89	0.12
Speeded Computation	19.59	5.99	3.70	4.02	12.29	5.30	0.47	0.11	0.84	0.16
Nonspeeded Computation	8.28	2.45	2.83	2.26	11.11	1.33	0.93	0.11	0.74	0.20
Speeded Three Dimensional	15.07	6.31	12.50	8.79	27.58	7.32	0.69	0.18	0.57	0.24
Nonspeeded Three Dimensional	7.14	2.96	7.60	3.13	14.73	0.93	0.98	0.06	0.49	0.21
Speeded Vocabulary	14.83	6.96	11.24	7.99	26.07	7.76	0.43	0.13	0.58	0.22
Nonspeeded Vocabulary	6.88	3.12	6.86	3.09	13.74	0.87	0.98	0.06	0.50	0.22
Speeded Arithmetic Reasoning	8.41	3.39	5.35	4.42	13.77	4.19	0.55	0.17	0.64	0.24
Nonspeeded Arithmetic Reasoning	5.00	1.97	4.82	2.03	9.82	0.55	0.98	0.05	0.51	0.20

Note: Proportion Attempted = Number Answered / Number of Items.
       Accuracy = Number Correct / Number Answered.

Table C-7.   Five Alternative Scores on Old Instr./Format: 40 Years of Age and Over   (N=213).

Test	Correct		Wrong		Answered		Proportion Attempted		Accuracy	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Speeded Name Comparison	38.32	13.41	5.62	8.82	43.94	16.72	0.29	0.11	0.87	0.14
Speeded Computation	19.35	6.41	3.43	4.56	22.78	6.34	0.46	0.13	0.85	0.16
Nonspeeded Computation	7.99	2.52	2.66	2.13	10.64	1.70	0.89	0.14	0.75	0.19
Speeded Three Dimensional	14.16	5.62	10.88	7.88	25.04	7.00	0.63	0.18	0.59	0.24
Nonspeeded Three Dimensional	6.76	2.87	7.88	3.00	14.63	1.15	0.98	0.08	0.46	0.20
Speeded Vocabulary	16.74	7.86	7.77	6.78	24.51	8.02	0.41	0.13	0.67	0.23
Nonspeeded Vocabulary	8.12	3.64	5.36	3.41	13.47	1.78	0.96	0.13	0.60	0.25
Speeded Arithmetic Reasoning	8.72	3.55	4.05	4.11	12.77	3.95	0.51	0.16	0.70	0.24
Nonspeeded Arithmetic Reasoning	5.47	2.29	4.15	2.33	9.62	0.91	0.96	0.09	0.57	0.23

Note: Proportion Attempted = Number Answered / Number of Items.
          Accuracy = Number Correct / Number Answered.

Table C-8. Five Alternative Scores on New Instr./Format: Whites (N=230).

Test	Correct		Wrong		Answered		Proportion Attempted		Accuracy	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Speeded Name Comparison	48.53	13.24	4.97	5.49	53.49	12.76	0.59	0.14	0.90	0.12
Speeded Computation	19.50	4.66	3.14	3.80	22.64	4.38	0.58	0.11	0.86	0.13
Nonspeeded Computation	8.96	2.17	2.07	1.89	11.03	1.33	0.92	0.11	0.81	0.17
Speeded Three Dimensional	16.10	6.06	11.90	7.73	28.00	6.23	0.80	0.18	0.59	0.23
Nonspeeded Three Dimensional	8.32	3.06	6.10	3.26	14.41	1.33	0.96	0.09	0.58	0.22
Speeded Vocabulary	18.83	6.75	11.04	8.17	29.87	7.57	0.60	0.15	0.64	0.20
Nonspeeded Vocabulary	8.42	3.25	5.39	3.15	13.81	0.85	0.99	0.06	0.61	0.23
Speeded Arithmetic Reasoning	10.69	3.36	4.66	4.10	15.34	3.89	0.64	0.16	0.72	0.20
Nonspeeded Arithmetic Reasoning	6.10	2.25	3.76	2.19	9.86	0.50	0.99	0.05	0.62	0.22

Note: Proportion Attempted = Number Answered / Number of Items.
Accuracy = Number Correct / Number Answered.

Table C-9.　Five Alternative Scores on New Instr./Format:　Blacks　(N=283).

Test	Correct		Wrong		Answered		Proportion Attempted		Accuracy	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Speeded Name Comparison	40.85	13.19	5.95	6.05	46.80	13.58	0.52	0.15	0.87	0.12
Speeded Computation	16.07	5.03	3.36	3.17	19.43	4.65	0.50	0.12	0.82	0.15
Nonspeeded Computation	7.54	2.54	2.53	2.01	10.08	1.95	0.84	0.16	0.74	0.20
Speeded Three Dimensional	11.99	5.13	14.63	7.58	26.62	6.79	0.76	0.19	0.47	0.21
Nonspeeded Three Dimensional	6.40	2.76	8.05	3.06	14.45	1.37	0.96	0.09	0.45	0.20
Speeded Vocabulary	13.00	5.98	12.26	7.74	25.26	8.07	0.51	0.16	0.52	0.20
Nonspeeded Vocabulary	6.20	2.92	7.32	2.83	13.53	1.36	0.97	0.10	0.46	0.21
Speeded Arithmetic Reasoning	7.42	3.22	5.60	4.21	13.01	4.31	0.54	0.18	0.59	0.23
Nonspeeded Arithmetic Reasoning	4.82	1.86	4.91	1.90	9.73	0.86	0.97	0.09	0.50	0.19

Note:　Proportion Attempted = Number Answered / Number of Items.
　　　　Accuracy = Number Correct / Number Answered.

Table C-10. Five Alternative Scores on New Instr./Format: Hispanics (N=278).

Test	Correct		Wrong		Answered		Proportion Attempted		Accuracy	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Speeded Name Comparison	46.03	12.87	6.67	6.78	52.70	13.61	0.59	0.15	0.87	0.12
Speeded Computation	17.87	4.37	4.09	4.47	21.96	4.48	0.56	0.11	0.82	0.15
Nonspeeded Computation	8.53	2.22	2.51	2.06	11.03	1.16	0.92	0.10	0.77	0.18
Speeded Three Dimensional	15.50	5.29	14.00	7.64	29.50	5.44	0.84	0.16	0.55	0.21
Nonspeeded Three Dimensional	7.45	2.88	7.35	2.99	14.79	0.74	0.99	0.05	0.51	0.20
Speeded Vocabulary	14.84	6.02	15.18	8.39	30.02	7.51	0.60	0.15	0.51	0.20
Nonspeeded Vocabulary	6.72	2.94	7.12	2.90	13.84	0.80	0.99	0.06	0.49	0.21
Speeded Arithmetic Reasoning	8.65	3.18	6.15	4.83	14.80	4.33	0.62	0.18	0.61	0.23
Nonspeeded Arithmetic Reasoning	5.06	1.84	4.74	1.92	9.81	0.67	0.98	0.07	0.52	0.19

Note: Proportion Attempted = Number Answered / Number of Items.
        Accuracy = Number Correct / Number Answered.

Table C-11. Five Alternative Scores on New Instr./Format: Males (N=538).

Test	Correct		Wrong		Answered		Proportion Attempted		Accuracy	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Speeded Name Comparison	43.33	12.65	6.01	6.20	49.34	13.29	0.55	0.15	0.88	0.12
Speeded Computation	17.57	5.01	3.61	3.94	21.17	4.91	0.54	0.13	0.83	0.15
Nonspeeded Computation	8.16	2.43	2.44	2.02	10.60	1.64	0.88	0.14	0.77	0.19
Speeded Three Dimensional	14.70	5.77	12.95	7.74	27.65	6.29	0.79	0.18	0.55	0.22
Nonspeeded Three Dimensional	7.52	3.05	7.01	3.23	14.53	1.24	0.97	0.08	0.52	0.21
Speeded Vocabulary	15.38	6.74	12.52	8.21	27.91	8.02	0.56	0.16	0.56	0.21
Nonspeeded Vocabulary	7.04	3.25	6.65	3.13	13.70	1.11	0.98	0.08	0.51	0.23
Speeded Arithmetic Reasoning	8.88	3.65	5.25	4.23	14.13	4.18	0.59	0.17	0.64	0.22
Nonspeeded Arithmetic Reasoning	5.45	2.08	4.36	2.08	9.81	0.70	0.98	0.07	0.56	0.21

Note: Proportion Attempted = Number Answered / Number of Items.
       Accuracy = Number Correct / Number Answered.

Table C-12.   Five Alternative Scores on New Instr./Format: Females (N=271).

Test	Correct		Wrong		Answered		Proportion Attempted		Accuracy	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Speeded Name Comparison	48.31	14.42	5.63	6.09	53.94	13.87	0.60	0.15	0.89	0.12
Speeded Computation	18.00	4.65	3.48	3.69	21.48	4.34	0.55	0.11	0.84	0.14
Nonspeeded Computation	8.59	2.28	2.29	1.93	10.88	1.49	0.91	0.12	0.79	0.18
Speeded Three Dimensional	13.91	5.71	14.89	7.56	28.80	6.31	0.82	0.18	0.50	0.21
Nonspeeded Three Dimensional	6.98	2.81	7.61	3.03	14.59	1.12	0.97	0.07	0.48	0.20
Speeded Vocabulary	15.49	6.62	13.66	8.47	29.15	8.00	0.58	0.16	0.54	0.21
Nonspeeded Vocabulary	7.16	3.07	6.60	2.99	13.76	0.92	0.98	0.07	0.52	0.22
Speeded Arithmetic Reasoning	8.71	3.25	6.02	4.78	14.73	4.55	0.61	0.19	0.62	0.23
Nonspeeded Arithmetic Reasoning	4.97	1.95	4.78	2.00	9.75	0.75	0.98	0.07	0.51	0.22

Note: Proportion Attempted = Number Answered / Number of Items.
Accuracy = Number Correct / Number Answered.

Table C-13. Five Alternative Scores on New Instr./Format: Under 40 Years of Age (N=576).

Test	Correct		Wrong		Answered		Proportion Attempted		Accuracy	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Speeded Name Comparison	46.59	13.25	5.91	6.29	52.50	13.70	0.58	0.15	0.89	0.12
Speeded Computation	17.94	4.66	3.68	3.95	21.62	4.57	0.55	0.12	0.83	0.15
Nonspeeded Computation	8.41	2.30	2.44	1.98	10.85	1.45	0.90	0.12	0.77	0.18
Speeded Three Dimensional	15.02	5.67	13.43	7.64	28.45	6.06	0.81	0.17	0.55	0.22
Nonspeeded Three Dimensional	7.60	2.97	7.02	3.15	14.62	1.11	0.97	0.07	0.52	0.21
Speeded Vocabulary	15.02	6.35	13.86	8.28	28.89	7.87	0.58	0.16	0.53	0.20
Nonspeeded Vocabulary	6.88	2.95	6.91	2.89	13.79	0.87	0.98	0.06	0.50	0.21
Speeded Arithmetic Reasoning	8.89	3.42	5.77	4.46	14.66	4.30	0.61	0.18	0.63	0.22
Nonspeeded Arithmetic Reasoning	5.24	2.03	4.62	2.06	9.86	0.51	0.99	0.05	0.53	0.21

Note: Proportion Attempted = Number Answered / Number of Items.
Accuracy = Number Correct / Number Answered.

C-14

Table C-14. Five Alternative Scores on New Instr./Format: 40 Years of Age and Over (N=230).

Test	Correct		Wrong		Answered		Proportion Attempted		Accuracy	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Speeded Name Comparison	41.04	13.15	5.77	5.84	46.81	12.72	0.52	.14	0.87	0.12
Speeded Computation	17.19	5.38	3.31	3.61	20.50	4.98	0.53	0.13	0.84	0.15
Nonspeeded Computation	8.04	2.58	2.26	2.04	10.30	1.83	0.86	0.15	0.78	0.20
Speeded Three Dimensional	13.01	5.78	13.96	7.95	26.97	6.75	0.77	0.19	0.50	0.23
Nonspeeded Three Dimensional	6.68	2.94	7.70	3.23	14.38	1.39	0.96	0.09	0.47	0.21
Speeded Vocabulary	16.50	7.36	10.54	7.94	27.04	8.23	0.54	0.16	0.62	0.22
Nonspeeded Vocabulary	7.65	3.65	5.91	3.43	13.56	1.37	0.97	0.10	0.56	0.26
Speeded Arithmetic Reasoning	8.72	3.76	4.84	4.34	13.57	4.23	0.57	0.18	0.66	0.23
Nonspeeded Arithmetic Reasoning	5.45	2.10	4.17	2.04	9.62	1.05	0.96	0.10	0.56	0.21

Note: Proportion Attempted = Number Answered / Number of Items.
Accuracy = Number Correct / Number Answered.

APPENDIX D

Estimated Internal Consistency Reliabilities
for Nonspeeded GATB Power Tests of Various Lengths

Note: The first row of each table contains the internal consistency reliability and the time limit for the version of the test used in this investigation. Each consecutive row contains an estimate of the internal consistency reliability (based on the Spearman-Brown prophecy formula) and administration time (based on the mean time required per item in the actual test) for a nonspeeded version of the same test that is one item longer than the test referred to in the previous row. Tables D-1a through D-4a provide these estimates for the old instructions/format power tests; Tables D-1b through D-4b provide these estimates for the new instructions/format power tests.

D-1

Table D-1a.
Estimates of Alpha for Various Test Lengths Based on Nonspeeded Computation Old Instructions/Format (12 items, 6 minutes, Alpha = 0.71)

Number of Items	Estimated Alpha	Estimated Time
12	.71	6:00
13	.73	6:30
14	.74	7:00
15	.75	7:30
16	.77	8:00
17	.78	8:30
18	.79	9:00
19	.79	9:30
20	.80	10:00
21	.81	10:30
22	.82	11:00
23	.82	11:30
24	.83	12:00
25	.84	12:30
26	.84	13:00
27	.85	13:30
28	.85	14:00
29	.86	14:30
30	.86	15:00
31	.86	15:30
32	.87	16:00
33	.87	16:30
34	.87	17:00
35	.88	17:30
36	.88	18:00
37	.88	18:30
38	.89	19:00
39	.89	19:30
40	.89	20:00
41	.89	20:30
42	.90	21:00
43	.90	21:30
44	.90	22:00
45	.90	22:30
46	.90	23:00
47	.91	23:30
48	.91	24:00
49	.91	24:30
50	.91	25:00

Table D-1b.
Estimates of Alpha for Various Test Lengths Based on Nonspeeded Computation New Instructions/Format (12 items, 6 minutes, Alpha = 0.71)

Number of Items	Estimated Alpha	Estimated Time
12	.71	6:00
13	.73	6:30
14	.74	7:00
15	.75	7:30
16	.77	8:00
17	.78	8:30
18	.79	9:00
19	.79	9:30
20	.80	10:00
21	.81	10:30
22	.82	11:00
23	.82	11:30
24	.83	12:00
25	.84	12:30
26	.84	13:00
27	.85	13:30
28	.85	14:00
29	.86	14:30
30	.86	15:00
31	.86	15:30
32	.87	16:00
33	.87	16:30
34	.87	17:00
35	.88	17:30
36	.88	18:00
37	.88	18:30
38	.89	19:00
39	.89	19:30
40	.89	20:00
41	.89	20:30
42	.90	21:00
43	.90	21:30
44	.90	22:00
45	.90	22:30
46	.90	23:00
47	.91	23:30
48	.91	24:00
49	.91	24:30
50	.91	25:00

Table D-2a.
Estimates of Alpha for Various Test Lengths Based on Nonspeeded 3-D Space Old Instructions/Format (15 items, 6 minutes, Alpha = 0.69)

Number of Items	Estimated Alpha	Estimated Time
15	.69	6:00
16	.70	6:24
17	.72	6:48
18	.73	7:12
19	.74	7:36
20	.75	8:00
21	.76	8:24
22	.77	8:48
23	.77	9:12
24	.78	9:36
25	.79	10:00
26	.79	10:24
27	.80	10:48
28	.81	11:12
29	.81	11:36
30	.82	12:00
31	.82	12:24
32	.83	12:48
33	.83	13:12
34	.83	13:36
35	.84	14:00
36	.84	14:24
37	.85	14:48
38	.85	15:12
39	.85	15:36
40	.86	16:00

Table D-2b.
Estimates of Alpha for Various Test Lengths Based on Nonspeeded 3-D Space New Instructions/Format (12 items, 6 minutes, Alpha = 0.70)

Number of Items	Estimated Alpha	Estimated Time
15	.70	6:00
16	.71	6:24
17	.73	6:48
18	.74	7:12
19	.75	7:36
20	.76	8:00
21	.77	8:24
22	.77	8:48
23	.78	9:12
24	.79	9:36
25	.80	10:00
26	.80	10:24
27	.81	10:48
28	.81	11:12
29	.82	11:36
30	.82	12:00
31	.83	12:24
32	.83	12:48
33	.84	13:12
34	.84	13:36
35	.84	14:00
36	.85	14:24
37	.85	14:48
38	.86	15:12
39	.86	15:36
40	.86	16:00

Table D-3a.
Estimates of Alpha for Various Test Lengths
Based on Nonspeeded Vocabulary Old
Instructions/Format (14 items, 6 minutes,
Alpha = 0.78)

Number of Items	Estimated Alpha	Estimated Time
14	.78	6:00
15	.79	6:26
16	.80	6:52
17	.81	7:17
18	.82	7:43
19	.83	8:08
20	.84	8:34
21	.84	9:00
22	.85	9:26
23	.85	9:52
24	.86	10:17
25	.86	10:43
26	.87	11:08
27	.87	11:34
28	.88	12:00
29	.88	12:26
30	.88	12:52
31	.89	13:17
32	.89	13:43
33	.89	14:08
34	.90	14:34
35	.90	15:00
36	.90	15:26
37	.90	15:52
38	.91	16:17
39	.91	16:43
40	.91	17:08
41	.91	17:34
42	.91	18:00
43	.92	18:26
44	.92	18:52
45	.92	19:17
46	.92	19:43
47	.92	20:08
48	.92	20:34
49	.93	21:00
50	.93	21:26
51	.93	21:52
52	.93	22:17

Table D-3a (cont).
Estimates of Alpha for Various Test Lengths
Based on Nonspeeded Vocabulary Old
Instructions/Format (14 items, 6 minutes,
Alpha = 0.78)

Number of Items	Estimated Alpha	Estimated Time
53	.92	22:43
54	.92	23:08
55	.93	23:34
56	.93	24:00
57	.93	24:26
58	.93	24:52
59	.93	25:17
60	.93	25:43

Table D-3b.
Estimates of Alpha for Various Test Lengths Based on Nonspeeded Vocabulary New Instructions/Format (14 items, 6 minutes, Alpha = 0.76)

Number of Items	Estimated Alpha	Estimated Time
14	.76	6:00
15	.77	6:26
16	.78	6:52
17	.79	7:17
18	.80	7:43
19	.81	8:08
20	.82	8:34
21	.83	9:00
22	.83	9:26
23	.84	9:52
24	.84	10:17
25	.85	10:43
26	.85	11:08
27	.86	11:34
28	.86	12:00
29	.87	12:26
30	.87	12:52
31	.88	13:17
32	.88	13:43
33	.88	14:08
34	.88	14:34
35	.89	15:00
36	.89	15:26
37	.89	15:52
38	.90	16:17
39	.90	16:43
40	.90	17:08
41	.90	17:34
42	.91	18:00
43	.91	18:26
44	.91	18:52
45	.91	19:17
46	.91	19:43
47	.92	20:08
48	.92	20:34
49	.92	21:00
50	.92	21:26
51	.92	21:52
52	.92	22:17

Table D-3b (cont).
Estimates of Alpha for Various Test Lengths Based on Nonspeeded Vocabulary New Instructions/Format (14 items, 6 minutes, Alpha = 0.76)

Number of Items	Estimated Alpha	Estimated Time
53	.93	22:43
54	.93	23:08
55	.93	23:34
56	.93	24:00
57	.94	24:26
58	.94	24:52
59	.94	25:17
60	.94	25:43

Table D-4a.
Estimates of Alpha for Various Test Lengths Based on Nonspeeded Arith. Reas. Old Instructions/Format (10 items, 11 minutes, Alpha = 0.56)

Number of Items	Estimated Alpha	Estimated Time
10	.56	11:00
11	.58	12:06
12	.60	13:12
13	.62	14:18
14	.64	15:24
15	.66	16:30
16	.67	17:36
17	.68	18:42
18	.70	19:48
19	.71	20:54
20	.72	22:00
21	.73	23:06
22	.74	24:12
23	.75	25:18
24	.75	26:24
25	.76	27:30

Table D-4b.
Estimates of Alpha for Various Test Lengths Based on Nonspeeded Arith. Reas. New Instructions/Format (10 items, 11 minutes, Alpha = 0.57)

Number of Items	Estimated Alpha	Estimated Time
10	.57	11:00
11	.59	12:06
12	.61	13:12
13	.63	14:18
14	.65	15:24
15	.67	16:30
16	.68	17:36
17	.69	18:42
18	.70	19:48
19	.72	20:54
20	.73	22:00
21	.74	23:06
22	.74	24:12
23	.75	25:18
24	.76	26:24
25	.77	27:30

APPENDIX E

**Source Tables for
the Univariate Condition by Group
Analyses of Variance
for the Speeded and Nonspeeded Tests**

Note: The following source tables are summarized in the text (Tables 15 and 18). Each analysis is a univariate 2 x 2 between Analysis of Variance. The levels of the first independent variable in all analyses are Old Instructions/Format and New Instructions/Format; the levels of the second independent variable in each analysis are two of the seven subgroups (i.e., Whites, Blacks, Hispanics, Males, Females, Under 40 Years of Age, 40 Years of Age and Over). Accuracy is a computed score (Number of Items Correct / Number of Items Attempted). Analyses of the speeded test scores are based on slightly smaller sample sizes than the analyses of the Nonspeeded test scores because the speeded test score analyses involve the Accuracy score; therefore, individuals who did not answer any items on a test were eliminated. This was done so that calculation of the Accuracy score would not result in a value of 0 in the denominator.

Table E-1. Instr./Format (i.e., Old versus New) by Subgroup (i.e., White versus Black) on Nonspeeded Computation Raw Scores.

Source	d.f.	SS	MS	F	p
Instr./Format	1	0.24	0.24	0.04	0.84
Subgroup	1	409.52	409.52	70.28	0.00
Interaction	1	5.68	5.68	0.97	0.32

Table E-2. Instr./Format (i.e., Old versus New) by Subgroup (i.e., White versus Hispanic) on Nonspeeded Computation Raw Scores.

Source	d.f.	SS	MS	F	p
Instr./Format	1	8.15	8.15	1.59	0.21
Subgroup	1	46.40	46.40	9.03	0.00
Interaction	1	0.00	0.00	0.00	1.00

Table E-3. Instr./Format (i.e., Old versus New) by Subgroup (i.e., Male versus Female) on Nonspeeded Computation Raw Scores.

Source	d.f.	SS	MS	F	p
Instr./Format	1	3.76	3.76	0.64	0.42
Subgroup	1	71.13	71.13	12.13	0.00
Interaction	1	0.05	0.05	0.01	0.93

Table E-4. Instr./Format (i.e., Old versus New) by Subgroup (i.e., Under 40 Years of Age versus 40 Years of Age and Over) on Nonspeeded Computation Raw Scores.

Source	d.f.	SS	MS	F	p
Instr./Format	1	2.10	2.10	0.36	0.55
Subgroup	1	34.47	34.47	5.86	0.02
Interaction	1	0.56	0.56	0.09	0.76

Table E-5.   Instr./Format (i.e., Old versus New) by Subgroup (i.e., White versus Black) on Nonspeeded 3-D Space Raw Scores.

Source	d.f.	SS	MS	F	p
Instr./Format	1	29.16	29.16	3.53	0.06
Subgroup	1	926.26	926.26	112.05	0.00
Interaction	1	0.66	0.66	0.08	0.78

Table E-6.   Instr./Format (i.e., Old versus New) by Subgroup (i.e., White versus Hispanic) on Nonspeeded 3-D Space Raw Scores.

Source	d.f.	SS	MS	F	p
Instr./Format	1	22.40	22.40	2.57	0.11
Subgroup	1	166.06	166.06	19.07	0.00
Interaction	1	2.08	2.08	0.24	0.63

Table E-7.   Instr./Format (i.e., Old versus New) by Subgroup (i.e., Male versus Female) on Nonspeeded 3-D Space Raw Scores.

Source	d.f.	SS	MS	F	p
Instr./Format	1	25.14	25.14	2.88	0.09
Subgroup	1	72.12	72.12	8.27	0.00
Interaction	1	4.49	4.49	0.51	0.47

Table E-8.   Instr./Format (i.e., Old versus New) by Subgroup (i.e., Under 40 Years of Age versus 40 Years of Age and Over) on Nonspeeded 3-D Space Raw Scores.

Source	d.f.	SS	MS	F	p
Instr./Format	1	13.13	13.13	1.51	0.22
Subgroup	1	141.12	141.12	16.22	0.00
Interaction	1	24.30	24.30	2.79	0.09

Table E-9.   Instr./Format (i.e., Old versus New) by Subgroup (i.e., White versus Black) on Nonspeeded Vocabulary Raw Scores.

Source	d.f.	SS	MS	F	p
Instr./Format	1	17.45	17.45	1.71	0.19
Subgroup	1	1036.43	1036.43	101.42	0.00
Interaction	1	16.20	16.20	1.59	0.21

Table E-10.   Instr./Format (i.e., Old versus New) by Subgroup (i.e., White versus Hispanic) on Nonspeeded Vocabulary Raw Scores.

Source	d.f.	SS	MS	F	p
Instr./Format	1	0.16	0.16	0.02	0.90
Subgroup	1	805.98	805.98	82.90	0.00
Interaction	1	0.30	0.30	0.03	0.86

Table E-11.   Instr./Format (i.e., Old versus New) by Subgroup (i.e., Male versus Female) on Nonspeeded Vocabulary Raw Scores.

Source	d.f.	SS	MS	F	p
Instr./Format	1	0.20	0.20	0.02	0.89
Subgroup	1	7.25	7.25	0.69	0.41
Interaction	1	24.27	24.27	2.31	0.13

Table E-12.   Instr./Format (i.e., Old versus New) by Subgroup (i.e., Under 40 Years of Age versus 40 Years of Age and Over) on Nonspeeded Vocabulary Raw Scores.

Source	d.f.	SS	MS	F	p
Instr./Format	1	18.74	18.74	1.82	0.18
Subgroup	1	321.74	321.74	31.17	0.00
Interaction	1	21.07	21.07	2.04	0.15

Table E-13. Instr./Format (i.e., Old versus New) by Subgroup (i.e., White versus Black) on Nonspeeded Arithmetic Reasoning Raw Scores.

Source	d.f.	SS	MS	F	p
Instr./Format	1	5.71	5.71	1.38	0.24
Subgroup	1	482.15	482.15	116.42	0.00
Interaction	1	1.59	1.59	0.38	0.54

Table E-14. Instr./Format (i.e., Old versus New) by Subgroup (i.e., White versus Hispanic) on Nonspeeded Arithmetic Reasoning Raw Scores.

Source	d.f.	SS	MS	F	p
Instr./Format	1	1.80	1.80	0.43	0.51
Subgroup	1	292.15	292.15	70.18	0.00
Interaction	1	0.05	0.05	0.01	0.92

Table E-15. Instr./Format (i.e., Old versus New) by Subgroup (i.e., Male versus Female) on Nonspeeded Arithmetic Reasoning Raw Scores.

Source	d.f.	SS	MS	F	p
Instr./Format	1	10.99	10.99	2.63	0.11
Subgroup	1	104.89	104.89	25.06	0.00
Interaction	1	0.88	0.88	0.21	0.65

Table E-16. Instr./Format (i.e., Old versus New) by Subgroup (i.e., Under 40 Years of Age versus 40 Years of Age and Over) on Nonspeeded Arithmetic Reasoning Raw Scores.

Source	d.f.	SS	MS	F	p
Instr./Format	1	3.79	3.79	0.90	0.34
Subgroup	1	36.67	36.67	8.67	0.00
Interaction	1	6.63	6.63	1.57	0.21

Table E-17. Instr./Format (i.e., Old versus New) by Subgroup (i.e., White versus Black) on Nonspeeded Computation Number of Items Attempted.

Source	d.f.	SS	MS	F	p
Instr./Format	1	33.90	33.90	13.63	0.00
Subgroup	1	144.98	144.98	58.28	0.00
Interaction	1	9.89	9.89	3.98	0.05

Table E-18. Instr./Format (i.e., Old versus New) by Subgroup (i.e., White versus Hispanic) on Nonspeeded Computation Number of Items Attempted.

Source	d.f.	SS	MS	F	p
Instr./Format	1	7.72	7.72	4.74	0.03
Subgroup	1	0.17	0.17	0.11	0.75
Interaction	1	0.01	0.01	0.01	0.93

Table E-19. Instr./Format (i.e., Old versus New) by Subgroup (i.e., Male versus Female) on Nonspeeded Computation Number of Items Attempted.

Source	d.f.	SS	MS	F	p
Instr./Format	1	30.12	30.12	13.02	0.00
Subgroup	1	28.94	28.94	12.52	0.00
Interaction	1	0.04	0.04	0.02	0.89

Table E-20. Instr./Format (i.e., Old versus New) by Subgroup (i.e., Under 40 Years of Age versus 40 Years of Age and Over) on Nonspeeded Computation Number of Items Attempted.

Source	d.f.	SS	MS	F	p
Instr./Format	1	29.16	29.16	12.90	0.00
Subgroup	1	80.52	80.52	35.63	0.00
Interaction	1	0.49	0.49	0.22	0.64

Table E-21.   Instr./Format (i.e., Old versus New) by Subgroup (i.e., White versus Black) on Nonspeeded 3-D Space Number of Items Attempted.

Source	d.f.	SS	MS	F	p
Instr./Format	1	14.98	14.98	10.54	0.00
Subgroup	1	0.00	0.00	0.00	0.97
Interaction	1	0.24	0.24	0.17	0.68

Table E-22.   Instr./Format (i.e., Old versus New) by Subgroup (i.e., White versus Hispanic) on Nonspeeded 3-D Space Number of Items Attempted.

Source	d.f.	SS	MS	F	p
Instr./Format	1	4.08	4.08	4.06	0.04
Subgroup	1	13.99	13.99	13.92	0.00
Interaction	1	5.34	5.34	5.31	0.02

Table E-23.   Instr./Format (i.e., Old versus New) by Subgroup (i.e., Male versus Female) on Nonspeeded 3-D Space Number of Items Attempted.

Source	d.f.	SS	MS	F	p
Instr./Format	1	9.82	9.82	8.27	0.00
Subgroup	1	1.47	1.47	1.24	0.27
Interaction	1	0.00	0.00	0.00	0.95

Table E-24.   Instr./Format (i.e., Old versus New) by Subgroup (i.e., Under 40 Years of Age versus 40 Years of Age and Over) on Nonspeeded 3-D Space Number of Items Attempted.

Source	d.f.	SS	MS	F	p
Instr./Format	1	10.78	10.78	9.04	0.00
Subgroup	1	9.26	9.26	7.76	0.01
Interaction	1	1.72	1.72	1.44	0.23

Table E-25.   Instr./Format (i.e., Old versus New) by Subgroup (i.e., White versus Black) on Nonspeeded Vocabulary Number of Items Attempted.

Source	d.f.	SS	MS	F	p
Instr./Format	1	0.76	0.76	0.53	0.47
Subgroup	1	17.00	17.00	11.88	0.00
Interaction	1	0.27	0.27	0.19	0.66

Table E-26.   Instr./Format (i.e., Old versus New) by Subgroup (i.e., White versus Hispanic) on Nonspeeded Vocabulary Number of Items Attempted.

Source	d.f.	SS	MS	F	p
Instr./Format	1	0.94	0.94	1.10	0.29
Subgroup	1	0.66	0.66	0.77	0.38
Interaction	1	0.17	0.17	0.20	0.66

Table E-27.   Instr./Format (i.e., Old versus New) by Subgroup (i.e., Male versus Female) on Nonspeeded Vocabulary Number of Items Attempted.

Source	d.f.	SS	MS	F	p
Instr./Format	1	1.35	1.35	1.07	0.30
Subgroup	1	1.79	1.79	1.42	0.23
Interaction	1	0.00	0.00	0.00	0.99

Table E-28.   Instr./Format (i.e., Old versus New) by Subgroup (i.e., Under 40 Years of Age versus 40 Years of Age and Over) on Nonspeeded Vocabulary Number of Items Attempted.

Source	d.f.	SS	MS	F	p
Instr./Format	1	1.47	1.47	1.20	0.27
Subgroup	1	19.52	19.52	16.02	0.00
Interaction	1	0.07	0.07	0.06	0.81

Table E-29. Instr./Format (i.e., Old versus New) by Subgroup (i.e., White versus Black) on Nonspeeded Arithmetic Reasoning Number of Items Attempted.

Source	d.f.	SS	MS	F	p
Instr./Format	1	0.00	0.00	0.00	0.95
Subgroup	1	3.37	3.37	8.15	0.00
Interaction	1	0.04	0.04	0.09	0.76

Table E-30. Instr./Format (i.e., Old versus New) by Subgroup (i.e., White versus Hispanic) on Nonspeeded Arithmetic Reasoning Number of Items Attempted.

Source	d.f.	SS	MS	F	p
Instr./Format	1	0.32	0.32	0.81	0.37
Subgroup	1	1.31	1.31	3.32	0.07
Interaction	1	0.11	0.11	0.28	0.60

Table E-31. Instr./Format (i.e., Old versus New) by Subgroup (i.e., Male versus Female) on Nonspeeded Arithmetic Reasoning Number of Items Attempted.

Source	d.f.	SS	MS	F	p
Instr./Format	1	0.08	0.08	0.18	0.67
Subgroup	1	0.58	0.58	1.22	0.27
Interaction	1	0.10	0.10	0.20	0.65

Table E-32. Instr./Format (i.e., Old versus New) by Subgroup (i.e., Under 40 Years of Age versus 40 Years of Age and Over) on Nonspeeded Arithmetic Reasoning Number of Items Attempted.

Source	d.f.	SS	MS	F	p
Instr./Format	1	0.12	0.12	0.26	0.61
Subgroup	1	15.11	15.11	32.63	0.00
Interaction	1	0.10	0.10	0.21	0.64

Table E-33.   Instr./Format (i.e., Old versus New) by Subgroup (i.e., White versus Black) on Speeded Name Comparison Accuracy.

Source	d.f.	SS	MS	F	p
Instr./Format	1	0.00	0.00	0.07	0.79
Subgroup	1	0.35	0.35	25.04	0.00
Interaction	1	0.00	0.00	0.31	0.58

Table E-34.   Instr./Format (i.e., Old versus New) by Subgroup (i.e., White versus Hispanic) on Speeded Name Comparison Accuracy.

Source	d.f.	SS	MS	F	p
Instr./Format	1	0.01	0.01	0.72	0.40
Subgroup	1	0.21	0.21	15.17	0.00
Interaction	1	0.00	0.00	0.00	0.98

Table E-35.   Instr./Format (i.e., Old versus New) by Subgroup (i.e., Male versus Female) on Speeded Name Comparison Accuracy.

Source	d.f.	SS	MS	F	p
Instr./Format	1	0.00	0.00	0.10	0.75
Subgroup	1	0.09	0.09	6.30	0.01
Interaction	1	0.00	0.00	0.27	0.61

Table E-36.   Instr./Format (i.e., Old versus New) by Subgroup (i.e., Under 40 Years of Age versus 40 Years of Age and Over) on Speeded Name Comparison Accuracy.

Source	d.f.	SS	MS	F	p
Instr./Format	1	0.00	0.00	0.05	0.83
Subgroup	1	0.05	0.05	3.40	0.07
Interaction	1	0.00	0.00	0.03	0.87

Table E-37. Instr./Format (i.e., Old versus New) by Subgroup (i.e., White versus Black) on Speeded Computation Accuracy.

Source	d.f.	SS	MS	F	p
Instr./Format	1	0.01	0.01	0.32	0.57
Subgroup	1	0.48	0.48	19.76	0.00
Interaction	1	0.00	0.00	0.04	0.85

Table E-38. Instr./Format (i.e., Old versus New) by Subgroup (i.e., White versus Hispanic) on Speeded Computation Accuracy.

Source	d.f.	SS	MS	F	p
Instr./Format	1	0.03	0.03	1.60	0.21
Subgroup	1	0.19	0.19	8.65	0.00
Interaction	1	0.06	0.06	2.73	0.10

Table E-39. Instr./Format (i.e., Old versus New) by Subgroup (i.e., Male versus Female) on Speeded Computation Accuracy.

Source	d.f.	SS	MS	F	p
Instr./Format	1	0.02	0.02	0.74	0.39
Subgroup	1	0.05	0.05	2.03	0.15
Interaction	1	0.00	0.00	0.19	0.66

Table E-40. Instr./Format (i.e., Old versus New) by Subgroup (i.e., Under 40 Years of Age versus 40 Years of Age and Over) on Speeded Computation Accuracy.

Source	d.f.	SS	MS	F	p
Instr./Format	1	0.02	0.02	0.71	0.40
Subgroup	1	0.01	0.01	0.50	0.48
Interaction	1	0.00	0.00	0.12	0.73

Table E-41. Instr./Format (i.e., Old versus New) by Subgroup (i.e., White versus Black) on Speeded 3-D Space Accuracy.

Source	d.f.	SS	MS	F	p
Instr./Format	1	0.39	0.39	7.57	0.00
Subgroup	1	4.56	4.56	88.15	0.00
Interaction	1	0.02	0.02	0.37	0.54

Table E-42. Instr./Format (i.e., Old versus New) by Subgroup (i.e., White versus Hispanic) on Speeded 3-D Space Accuracy.

Source	d.f.	SS	MS	F	p
Instr./Format	1	0.76	0.76	15.21	0.00
Subgroup	1	0.44	0.44	8.87	0.00
Interaction	1	0.01	0.01	0.26	0.61

Table E-43. Instr./Format (i.e., Old versus New) by Subgroup (i.e., Male versus Female) on Speeded 3-D Space Accuracy.

Source	d.f.	SS	MS	F	p
Instr./Format	1	0.78	0.78	14.77	0.00
Subgroup	1	1.11	1.11	21.13	0.00
Interaction	1	0.00	0.00	0.00	0.94

Table E-44. Instr./Format (i.e., Old versus New) by Subgroup (i.e., Under 40 Years of Age versus 40 Years of Age and Over) on Speeded 3-D Space Accuracy.

Source	d.f.	SS	MS	F	p
Instr./Format	1	1.05	1.05	19.76	0.00
Subgroup	1	0.05	0.05	0.95	0.33
Interaction	1	0.31	0.31	5.79	0.02

Table E-45. Instr./Format (i.e., Old versus New) by Subgroup (i.e., White versus Black) on Speeded Vocabulary Accuracy.

Source	d.f.	SS	MS	F	p
Instr./Format	1	0.48	0.48	10.14	0.00
Subgroup	1	3.90	3.90	82.28	0.00
Interaction	1	0.00	0.00	0.01	0.92

Table E-46. Instr./Format (i.e., Old versus New) by Subgroup (i.e., White versus Hispanic) on Speeded Vocabulary Accuracy.

Source	d.f.	SS	MS	F	p
Instr./Format	1	0.69	0.69	16.11	0.00
Subgroup	1	4.32	4.32	100.38	0.00
Interaction	1	0.02	0.02	0.35	0.55

Table E-47. Instr./Format (i.e., Old versus New) by Subgroup (i.e., Male versus Female) on Speeded Vocabulary Accuracy.

Source	d.f.	SS	MS	F	p
Instr./Format	1	0.79	0.79	16.34	0.00
Subgroup	1	0.11	0.11	2.30	0.13
Interaction	1	0.00	0.00	0.00	0.99

Table E-48. Instr./Format (i.e., Old versus New) by Subgroup (i.e., Under 40 Years of Age versus 40 Years of Age and Over) on Speeded Vocabulary Accuracy.

Source	d.f.	SS	MS	F	p
Instr./Format	1	0.83	0.83	17.72	0.00
Subgroup	1	2.67	2.67	57.17	0.00
Interaction	1	0.01	0.01	0.26	0.61

Table E-49. Instr./Format (i.e., Old versus New) by Subgroup (i.e., White versus Black) on Arithmetic Reasoning Accuracy.

Source	d.f.	SS	MS	F	p
Instr./Format	1	0.02	0.02	0.42	0.52
Subgroup	1	3.82	3.82	73.57	0.00
Interaction	1	0.00	0.00	0.02	0.87

Table E-50. Instr./Format (i.e., Old versus New) by Subgroup (i.e., White versus Hispanic) on Arithmetic Reasoning Accuracy.

Source	d.f.	SS	MS	F	p
Instr./Format	1	0.12	0.12	2.50	0.11
Subgroup	1	1.99	1.99	41.48	0.00
Interaction	1	0.06	0.06	1.15	0.28

Table E-51. Instr./Format (i.e., Old versus New) by Subgroup (i.e., Male versus Female) on Arithmetic Reasoning Accuracy.

Source	d.f.	SS	MS	F	p
Instr./Format	1	0.11	0.11	1.98	0.16
Subgroup	1	0.26	0.26	4.78	0.03
Interaction	1	0.00	0.00	0.01	0.93

Table E-52. Instr./Format (i.e., Old versus New) by Subgroup (i.e., Under 40 Years of Age versus 40 Years of Age and Over) on Arithmetic Reasoning Accuracy.

Source	d.f.	SS	MS	F	p
Instr./Format	1	0.19	0.19	3.59	0.06
Subgroup	1	0.82	0.82	15.46	0.00
Interaction	1	0.08	0.08	1.43	0.23

Table E-53. Instr./Format (i.e., Old versus New) by Subgroup (i.e., White versus Black) on Speeded Name Comparison Number of Items Attempted.

Source	d.f.	SS	MS	F	p
Instr./Format	1	5606.10	5606.10	32.25	0.00
Subgroup	1	11426.28	11426.28	65.74	0.00
Interaction	1	1.11	1.11	0.01	0.93

Table E-54. Instr./Format (i.e., Old versus New) by Subgroup (i.e., White versus Hispanic) on Speeded Name Comparison Number of Items Attempted.

Source	d.f.	SS	MS	F	p
Instr./Format	1	7415.82	7415.82	37.55	0.00
Subgroup	1	543.06	543.06	2.75	0.10
Interaction	1	113.03	113.03	0.57	0.45

Table E-55. Instr./Format (i.e., Old versus New) by Subgroup (i.e., Male versus Female) on Speeded Name Comparison Number of Items Attempted.

Source	d.f.	SS	MS	F	p
Instr./Format	1	11541.34	11541.34	61.86	0.00
Subgroup	1	8069.62	8069.62	43.25	0.00
Interaction	1	0.02	0.02	0.00	0.99

Table E-56. Instr./Format (i.e., Old versus New) by Subgroup (i.e., Under 40 Years of Age versus 40 Years of Age and Over) on Speeded Name Comparison Number of Items Attempted.

Source	d.f.	SS	MS	F	p
Instr./Format	1	6853.63	6853.63	36.45	0.00
Subgroup	1	5096.79	5096.79	27.11	0.00
Interaction	1	971.66	971.66	5.17	0.02

Table E-57. Instr./Format (i.e., Old versus New) by Subgroup (i.e., White versus Black) on Speeded Computation Number of Items Attempted.

Source	d.f.	SS	MS	F	p
Instr./Format	1	1301.05	1301.05	51.20	0.00
Subgroup	1	22999.00	22999.00	118.02	0.00
Interaction	1	8.96	8.96	0.35	0.55

Table E-58. Instr./Format (i.e., Old versus New) by Subgroup (i.e., White versus Hispanic) on Speeded Computation Number of Items Attempted.

Source	d.f.	SS	MS	F	p
Instr./Format	1	971.46	971.46	39.09	0.00
Subgroup	1	344.34	344.34	13.86	0.00
Interaction	1	58.79	58.79	2.37	0.12

Table E-59. Instr./Format (i.e., Old versus New) by Subgroup (i.e., Male versus Female) on Speeded Computation Number of Items Attempted.

Source	d.f.	SS	MS	F	p
Instr./Format	1	1393.82	1393.82	51.93	0.00
Subgroup	1	123.59	123.59	4.60	0.03
Interaction	1	25.15	25.15	0.94	0.33

Table E-60. Instr./Format (i.e., Old versus New) by Subgroup (i.e., Under 40 Years of Age versus 40 Years of Age and Over) on Speeded Computation Number of Items Attempted.

Source	d.f.	SS	MS	F	p
Instr./Format	1	1269.29	1269.29	47.61	0.00
Subgroup	1	215.34	215.34	8.08	0.00
Interaction	1	29.56	29.56	1.11	0.29

Table E-61. Instr./Format (i.e., Old versus New) by Subgroup (i.e., White versus Black) on Speeded 3-D Space Number of Items Attempted.

Source	d.f.	SS	MS	F	p
Instr./Format	1	116.65	116.65	2.41	0.12
Subgroup	1	367.32	367.32	7.60	0.01
Interaction	1	10.24	10.24	0.21	0.65

Table E-62. Instr./Format (i.e., Old versus New) by Subgroup (i.e., White versus Hispanic) on Speeded 3-D Space Number of Items Attempted.

Source	d.f.	SS	MS	F	p
Instr./Format	1	517.55	517.55	11.95	0.00
Subgroup	1	229.97	229.97	5.31	0.02
Interaction	1	77.98	77.98	1.80	0.18

Table E-63. Instr./Format (i.e., Old versus New) by Subgroup (i.e., Male versus Female) on Speeded 3-D Space Number of Items Attempted.

Source	d.f.	SS	MS	F	p
Instr./Format	1	473.77	473.77	10.22	0.00
Subgroup	1	787.78	787.78	16.99	0.00
Interaction	1	29.76	29.76	0.64	0.42

Table E-64. Instr./Format (i.e., Old versus New) by Subgroup (i.e., Under 40 Years of Age versus 40 Years of Age and Over) on Speeded 3-D Space Number of Items Attempted.

Source	d.f.	SS	MS	F	p
Instr./Format	1	637.14	637.14	13.82	0.00
Subgroup	1	1308.70	1308.70	28.39	0.00
Interaction	1	89.87	89.87	1.95	0.16

Table E-65.   Instr./Format (i.e., Old versus New) by Subgroup (i.e., White versus Black) on Speeded Vocabulary Number of Items Attempted.

Source	d.f.	SS	MS	F	p
Instr./Format	1	1201.76	1201.76	20.10	0.00
Subgroup	1	5009.56	5009.56	83.81	0.00
Interaction	1	13.77	13.77	0.23	0.63

Table E-66.   Instr./Format (i.e., Old versus New) by Subgroup (i.e., White versus Hispanic) on Speeded Vocabulary Number of Items Attempted.

Source	d.f.	SS	MS	F	p
Instr./Format	1	2198.11	2198.11	38.48	0.00
Subgroup	1	41.20	41.20	0.72	0.40
Interaction	1	76.21	76.21	1.33	0.25

Table E-67.   Instr./Format (i.e., Old versus New) by Subgroup (i.e., Male versus Female) on Speeded Vocabulary Number of Items Attempted.

Source	d.f.	SS	MS	F	p
Instr./Format	1	2840.98	2840.98	45.50	0.00
Subgroup	1	691.01	691.01	11.07	0.00
Interaction	1	4.08	4.08	0.07	0.80

Table E-68.   Instr./Format (i.e., Old versus New) by Subgroup (i.e., Under 40 Years of Age versus 40 Years of Age and Over) on Speeded Vocabulary Number of Items Attempted.

Source	d.f.	SS	MS	F	p
Instr./Format	1	2321.70	2321.70	37.23	0.00
Subgroup	1	936.99	936.99	15.03	0.00
Interaction	1	6.62	6.62	0.11	0.76

Table E-69. Instr./Format (i.e., Old versus New) by Subgroup (i.e., White versus Black) on Arithmetic Reasoning Number of Items Attempted.

Source	d.f.	SS	MS	F	p
Instr./Format	1	81.42	81.42	4.75	0.03
Subgroup	1	1225.08	1225.08	71.42	0.00
Interaction	1	6.70	6.70	0.39	0.53

Table E-70. Instr./Format (i.e., Old versus New) by Subgroup (i.e., White versus Hispanic) on Arithmetic Reasoning Number of Items Attempted.

Source	d.f.	SS	MS	F	p
Instr./Format	1	245.58	245.58	15.31	0.00
Subgroup	1	164.14	164.14	10.23	0.00
Interaction	1	17.03	17.03	1.06	0.30

Table E-71. Instr./Format (i.e., Old versus New) by Subgroup (i.e., Male versus Female) on Arithmetic Reasoning Number of Items Attempted.

Source	d.f.	SS	MS	F	p
Instr./Format	1	319.27	319.27	17.92	0.00
Subgroup	1	52.06	52.06	2.92	0.09
Interaction	1	21.05	21.05	1.18	0.28

Table E-72. Instr./Format (i.e., Old versus New) by Subgroup (i.e., Under 40 Years of Age versus 40 Years of Age and Over) on Arithmetic Reasoning Number of Items Attempted.

Source	d.f.	SS	MS	F	p
Instr./Format	1	230.54	230.54	13.05	0.00
Subgroup	1	354.93	354.93	20.09	0.00
Interaction	1	0.89	0.89	0.05	0.82

APPENDIX F

Correlations Among Test Scores Within Each Condition and Subgroup

Note: In the following correlation matrices NCSP = Name Comparison Speeded; CMSP = Computation Speeded; TDSP = Three-Dimensional Space Speeded; VOSP = Vocabulary Speeded; ARSP = Arithmetic Reasoning Speeded; CMNS = Computation Nonspeeded; TDNS = Three-Dimensional Space Nonspeeded; VONS = Vocabulary Nonspeeded; and ARNS = Arithmetic Reasoning Nonspeeded.

Table F-1. Correlations Among All Tests in the Old Instructions/Format Condition for All Examinees (N = 867).

	NCSP	CMSP	TDSP	VOSP	ARSP	CMNS	TDNS	VONS	ARNS
NCSP	1.000								
CMSP	.637	1.000							
TDSP	.424	.412	1.000						
VOSP	.551	.595	.476	1.000					
ARSP	.551	.734	.475	.655	1.000				
CMNS	.498	.716	.339	.482	.608	1.000			
TDNS	.336	.376	.683	.448	.436	.366	1.000		
VONS	.402	.499	.377	.744	.568	.467	.438	1.000	
ARNS	.353	.555	.378	.540	.635	.514	.436	.543	1.000

Table F-2. Correlations Among All Tests in the Old Instructions/Format Condition for White Examinees (N = 235).

	NCSP	CMSP	TDSP	VOSP	ARSP	CMNS	TDNS	VONS	ARNS
NCSP	1.000								
CMSP	.625	1.000							
TDSP	.411	.395	1.000						
VOSP	.549	.620	.472	1.000					
ARSP	.539	.776	.456	.638	1.000				
CMNS	.444	.720	.295	.524	.656	1.000			
TDNS	.383	.388	.719	.507	.452	.329	1.000		
VONS	.369	.525	.424	.763	.588	.503	.520	1.000	
ARNS	.415	.657	.382	.605	.708	.600	.456	.627	1.000

Table F-4.   Correlations Among All Tests in the Old Instructions/Format Condition for Black Examinees (N = 313).

	NCSP	CMSP	TDSP	VOSP	ARSP	CMNS	TDNS	VONS	ARNS
NCSP	1.000								
CMSP	.661	1.000							
TDSP	.408	.388	1.000						
VOSP	.539	.578	.457	1.000					
ARSP	.550	.719	.478	.594	1.000				
CMNS	.579	.777	.321	.516	.582	1.000			
TDNS	.320	.360	.624	.401	.396	.365	1.000		
VONS	.476	.539	.390	.761	.516	.484	.375	1.000	
ARNS	.337	.523	.365	.472	.580	.482	.386	.471	1.000

Table F-4.   Correlations Among All Tests in the Old Instructions/Format Condition for Hispanic Examinees (N = 301).

	NCSP	CMSP	TDSP	VOSP	ARSP	CMNS	TDNS	VONS	ARNS
NCSP	1.000								
CMSP	.599	1.000							
TDSP	.344	.336	1.000						
VOSP	.498	.498	.384	1.000					
ARSP	.501	.639	.362	.603	1.000				
CMNS	.435	.619	.314	.369	.573	1.000			
TDNS	.235	.267	.643	.318	.325	.309	1.000		
VONS	.309	.359	.290	.660	.512	.377	.365	1.000	
ARNS	.241	.392	.274	.396	.494	.433	.348	.446	1.000

Table F-5.   Correlations Among All Tests in the Old Instructions/Format Condition for Male Examinees (N = 533).

	NCSP	CMSP	TDSP	VOSP	ARSP	CMNS	TDNS	VONS	ARNS
NCSP	1.000								
CMSP	.675	1.000							
TDSP	.478	.414	1.000						
VOSP	.608	.632	.508	1.000					
ARSP	.602	.771	.449	.688	1.000				
CMNS	.546	.743	.310	.524	.648	1.000			
TDNS	.400	.388	.721	.469	.426	.369	1.000		
VONS	.459	.537	.387	.763	.607	.506	.440	1.000	
ARNS	.429	.608	.358	.564	.662	.570	.436	.580	1.000

Table F-6.   Correlations Among All Tests in the Old Instructions/Format Condition for Female Examinees (N = 329).

	NCSP	CMSP	TDSP	VOSP	ARSP	CMNS	TDNS	VONS	ARNS
NCSP	1.000								
CMSP	.575	1.000							
TDSP	.385	.431	1.000						
VOSP	.485	.536	.424	1.000					
ARSP	.524	.685	.531	.592	1.000				
CMNS	.416	.666	.427	.413	.556	1.000			
TDNS	.267	.369	.601	.409	.451	.382	1.000		
VONS	.366	.453	.356	.716	.483	.414	.431	1.000	
ARNS	.319	.511	.402	.515	.581	.463	.429	.467	1.000

Table F-7. Correlations Among All Tests in the Old Instructions/Format Condition for Examinees Under 40 Years of Age (N = 648).

	NCSP	CMSP	TDSP	VOSP	ARSP	CMNS	TDNS	VONS	ARNS
NCSP	1.000								
CMSP	.623	1.000							
TDSP	.419	.393	1.000						
VOSP	.563	.578	.480	1.000					
ARSP	.555	.721	.479	.643	1.000				
CMNS	.491	.698	.316	.466	.593	1.000			
TDNS	.302	.332	.685	.441	.407	.339	1.000		
VONS	.415	.478	.389	.733	.547	.443	.435	1.000	
ARNS	.343	.521	.394	.509	.597	.498	.412	.517	1.000

Table F-8. Correlations Among All Tests in the Old Instructions/Format Condition for Examinees 40 Years of Age and Over (N = 214).

	NCSP	CMSP	TDSP	VOSP	ARSP	CMNS	TDNS	VONS	ARNS
NCSP	1.000								
CMSP	.671	1.000							
TDSP	.424	.469	1.000						
VOSP	.573	.657	.519	1.000					
ARSP	.556	.772	.472	.685	1.000				
CMNS	.502	.764	.402	.558	.663	1.000			
TDNS	.420	.505	.677	.504	.535	.437	1.000		
VONS	.434	.581	.408	.760	.624	.575	.512	1.000	
ARNS	.416	.652	.369	.595	.730	.586	.540	.583	1.000

Table F-9. Correlations Among All Tests in the New Instructions/Format Condition for All Examinees (N = 814).

	NCSP	CMSP	TDSP	VOSP	ARSP	CMNS	TDNS	VONS	ARNS
NCSP	1.000								
CMSP	.637	1.000							
TDSP	.473	.457	1.000						
VOSP	.561	.591	.499	1.000					
ARSP	.569	.741	.467	.687	1.000				
CMNS	.532	.746	.386	.492	.627	1.000			
TDNS	.308	.360	.670	.421	.382	.329	1.000		
VONS	.440	.527	.389	.760	.578	.457	.379	1.000	
ARNS	..361	.541	.398	.547	.570	.502	.398	.552	1.000

Table F-10. Correlations Among All Tests in the New Instructions/Format Condition for White Examinees (N = 231).

	NCSP	CMSP	TDSP	VOSP	ARSP	CMNS	TDNS	VONS	ARNS
NCSP	1.000								
CMSP	.583	1.000							
TDSP	.462	.460	1.000						
VOSP	.519	.592	.504	1.000					
ARSP	.567	.741	.546	.658	1.000				
CMNS	.468	.719	.414	.529	.664	1.000			
TDNS	.287	.370	.687	.401	.433	.313	1.000		
VONS	.377	.512	.404	.730	.534	.423	.343	1.000	
ARNS	.354	.578	.424	.593	.625	.557	.435	.559	1.000

Table F-11. Correlations Among All Tests in the New Instructions/Format Condition for Black Examinees (N = 283).

	NCSP	CMSP	TDSP	VOSP	ARSP	CMNS	TDNS	VONS	ARNS
NCSP	1.000								
CMSP	.659	1.000							
TDSP	.473	.477	1.000						
VOSP	.630	.605	.500	1.000					
ARSP	.569	.783	.377	.631	1.000				
CMNS	.561	.763	.361	.486	.666	1.000			
TDNS	.287	.327	.656	.361	.258	.235	1.000		
VONS	.473	.500	.369	.748	.520	.435	.334	1.000	
ARNS	.384	.508	.367	.447	.530	.465	.255	.427	1.000

Table F-12. Correlations Among All Tests in the New Instructions/Format Condition for Hispanic Examinees (N = 279).

	NCSP	CMSP	TDSP	VOSP	ARSP	CMNS	TDNS	VONS	ARNS
NCSP	1.000								
CMSP	.583	1.000							
TDSP	.375	.321	1.000						
VOSP	.466	.466	.389	1.000					
ARSP	.498	.634	.357	.619	1.000				
CMNS	.460	.685	.257	.369	.491	1.000			
TDNS	.219	.255	.604	.341	.291	.325	1.000		
VONS	.377	.453	.304	.720	.532	.428	.338	1.000	
ARNS	.237	.442	.328	.423	.408	.424	.391	.514	1.000

Table F-13. Correlations Among All Tests in the New Instructions/Format Condition for Male Examinees (N = 540).

	NCSP	CMSP	TDSP	VOSP	ARSP	CMNS	TDNS	VONS	ARNS
NCSP	1.000								
CMSP	.627	1.000							
TDSP	.486	.431	1.000						
VOSP	.603	.615	.493	1.000					
ARSP	.607	.755	.453	.725	1.000				
CMNS	.496	.751	.378	.529	.657	1.000			
TDNS	.322	.335	.687	.406	.364	.326	1.000		
VONS	.455	.537	.389	.768	.592	.488	.396	1.000	
ARNS	.441	.583	.432	.487	.618	.567	.433	.578	1.000

Table F-14. Correlations Among All Tests in the Old Instructions/Format Condition for Female Examinees (N = 271).

	NCSP	CMSP	TDSP	VOSP	ARSP	CMNS	TDNS	VONS	ARNS
NCSP	1.000								
CMSP	.676	1.000							
TDSP	.501	.525	1.000						
VOSP	.505	.540	.512	1.000					
ARSP	.549	.709	.500	.609	1.000				
CMNS	.595	.733	.528	.420	.567	1.000			
TDNS	.333	.437	.629	.453	.430	.377	1.000		
VONS	.435	.507	.396	.748	.553	.395	.354	1.000	
ARNS	.300	.472	.316	.484	.456	.397	.314	.518	1.000

Table F-15. Correlations Among All Tests in the New Instructions/Format Condition for Examinees Under 40 Years of Age (N = 577).

	NCSP	CMSP	TDSP	VOSP	ARSP	CMNS	TDNS	VONS	ARNS
NCSP	1.000								
CMSP	.624	1.000							
TDSP	.438	.404	1.000						
VOSP	.564	.571	.507	1.000					
ARSP	.549	.708	.449	.676	1.000				
CMNS	.498	.707	.340	.474	.590	1.000			
TDNS	.267	.312	.662	.448	.363	.288	1.000		
VONS	.430	.492	.404	.736	.571	.438	.420	1.000	
ARNS	.341	.492	.394	.525	.552	.547	.396	.530	1.000

Table F-16. Correlations Among All Tests in the New Instructions/Format Condition for Examinees 40 Years of Age and Over (N = 231).

	NCSP	CMSP	TDSP	VOSP	ARSP	CMNS	TDNS	VONS	ARNS
NCSP	1.000								
CMSP	.658	1.000							
TDSP	.499	.551	1.000						
VOSP	.651	.658	.558	1.000					
ARSP	.631	.803	.507	.724	1.000				
CMNS	.600	.816	.467	.561	.702	1.000			
TDNS	.344	.441	.660	.427	.419	.401	1.000		
VONS	.561	.625	.434	.795	.607	.535	.365	1.000	
ARNS	.458	.661	.450	.595	.613	.613	.435	.605	1.000